Replicating Your Heart: Exploring Presentation Attacks on ECG Biometrics

Yang Gao, Wei Wang
University at Buffalo, SUNY
Buffalo, NY 14260
{ygao36, wwang49}@buffalo.edu

Borui Li, Omkar R. Patil
Binghamton University, SUNY
Binghamton, NY 13902
{bli28, opatil1}@binghamton.edu

Zhanpeng Jin
University at Buffalo, SUNY
Buffalo, NY 14260
zjin@buffalo.edu

***Abstract*—Electrocardiograph (ECG) has been well proved to contain adequately unique patterns for individual recognition and considered as a promising biometric which is hard to spoof and forge because of its intrinsic liveness detection and dynamic variance. However, unlike other conventional biometrics, the security vulnerabilities of ECG biometric systems have been largely under-explored. For example, given the multi-faceted roles of ECGs in both healthcare and biometrics, many large-scale ECG databases are publicly available online, which would provide the attackers a potential way to hack the authentication system. In this study, we present a new presentation attack method that can spoof the target authentication system by generating a sufficiently large number of high-quality fake ECG samples based upon the public datasets and limited attempt efforts. Our approach combines the cluster-based template searching, off-line substitute classifier training, data synthesis, and VAE-based counterfeits generation. Considering a robust ECG authentication system with effective replay detection mechanism that can effectively identify and reject common noise injection attacks, experimental results show that the fake ECG samples generated by our approach can achieve an average acceptance rate of 95%, compared with the best acceptance rate of 39% for noise-injected fake samples.**

I. Introduction

The past decade has seen an dramatically growing popularity of biometrics in security authentication applications. Although physiological (like fingerprint and iris) and behavioral (like keystrokes and gaits) biometrics have been widely used in people's daily life, they still suffer from either the constraints of ease of replication and non-cancellability, or the shortcomings of higher variance and less permanence. Recently another type of biometrics based on electrophysiological signals have gained increasing attention. In particular, electrocardiogram (ECG)-based biometrics become more prominent because ECG contains sufficiently detailed information pertaining to the highly individualized, functional and structural properties of the heart [1]. For instance, a smart wristband named "Nymi Band" that uses ECGs to authenticate user identity and any conceivable device has been used for wearable credit card payment. Prior research has also demonstrated ECG's superior advantages as a biometric: ECGs are present in all living individuals and ECG signals are hard to forge and counterfeit. More importantly, ECG signals exhibit a small level of intrinsic dynamic variance and are not constantly identical even for the same individual, which make it more resistant to conventional presentation or replay attacks [2].

The security of biometric systems has been extensively studied. As shown in Fig. 1, the authentication process could be compromised by the following three types of attacks [3]:

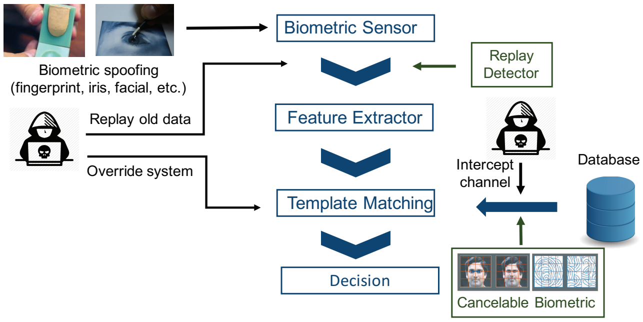


Fig. 1. Attacks and defenses in the biometric authentication system

- **Administration attack**, also known as the *insider attack*, which indicates the vulnerabilities of the administration of the biometric system, e.g., improper enrollment process, the collusive system administrator, or an authorized user coerced by the adversary.
- **Non-secure infrastructure**, referring to the vulnerabilities of hardware and software, includes intercepted communication channels, overridden authentication protocols, and stolen or modified biometric templates.
- **Biometric overtness**, originates from the limitations of the biometric itself and the biometric authentication system, which allows the adversary to spoof the biometric system by synthesizing fake samples.

From the attacker's perspective, compared with the high-cost administration attack, intercepting and stealing legitimate biometric templates are easier to implement. For example, the attacker could launch the replay attack once illegally obtaining the accepted samples from the authorized user (e.g., replicating the fingerprint left on the bottle [4]) or from the template database by eavesdropping the communication channel [5]. However, the legitimacy of input biometric samples can be verified by the replay detector [6]. Also, to avoid the information leakage during the communication process and increase the revocability of biometrics, the cancelable biometric system has been proposed [7], which means that the system will only transmitting the modified biometric patterns by revocable and non-invertible transformations.

Biometric authentication is normally done in an unsupervised manner, which means the attacker can spoof the system by feeding fake artificial signals into the biometric sensor [8]. However, those attacks often demand strict prerequisites. For example, the hill-climbing attack needs the detailed matching scores from the decision module for every fake sample attempt, and the brute-force attack requires a significant amount of computational resources and time, especially for the authentication system with a low false match rate (FMR).

Although some prior research on ECG biometrics have reached very high authentication accuracy over a large amount of subjects' data [1], [9], the security vulnerabilities of ECG biometric systems have been largely under-explored. Many existing studies have revealed the security threats of conventional biometric systems by using *a priori* knowledge of biometric characteristics, e.g., the attacker often first attempts a few high frequency PIN codes in a 4-digit PIN system [10], or uses some carefully designed "masterprints" [11] for spoofing the fingerprint authentication system. Thus we would like to ask the question: *Do ECGs hold any common characteristics or attributes so that there is a chance they can match with an arbitrary user's ECG template?* Given that ECG is the bioelectrical signal arising from the contraction of the heart muscles, ECGs of different individuals will hold similar cardiac patterns and some common morphological attributes.

To the best of our knowledge, our work is the first of its kind to explore an effective spoofing mechanism against the emerging ECG biometric authentication systems by generating a large amount of high-quality fake ECG samples that can be accepted by the authentication system. Our contributions are summarized as follows:

- We examined the vulnerability and resilience of ECG biometric systems and formulated the threat model.
- We proposed an attacking method capable of forging unlimited fake ECG samples with high acceptance rate and only a few attempt efforts, only based upon publicly accessible ECG databases and the returned authentication decisions (i.e., acceptance or rejection).
- We validated the effectiveness of the proposed spoofing mechanism on an ECG authentication system with strict replay detection and different FMRs.

II. Related Work

A. Brute-Force Attacks

In biometric domain, a crude brute-force attack means that the attacker keeps attempting the authentication system with an accessible corresponding large biometric database. Pagnin *et al.* [12] investigated brute-force attacks for different strategies on recovering a matching biometric. The results showed that these *centre search* attacks would imply the possibility of compromising the existing biometric authentication protocols based on simple distances (e.g., Hamming and Euclidean distances). Similar studies have investigated the brute-force attacks on Match-on-Card fingerprint [13], and palmprint [14] verification systems.

B. Hill-Climbing Attacks

Different from the brute-force attacks, when the large biometric database is not available or unfeasible, hill-climbing attacks can generate synthetic biometric data based on the feedback scores of the authentication matcher. Maiorana *et al.* [15] investigated the hill-climbing attacks on a multi-biometric recognition system including on-line signatures and EEGs. The results showed that the hill-climbing strategies may represent a potential threat even for the low False Acceptance Rate (FAR) biometric system, and require less efforts than a brute-force attack for successfully breaking the system. Specifically, many biometric systems, including fingerprint [13], iris [16], and signature [17], have been proved their potential risks by performing hill-climbing attacks.

C. Noise Injection Attacks

It is known that small random perturbations with suitable mathematical models that are introduced into a given signal can generate a very similar imposter dataset. Ghouzali *et.al.* described different attack modules including noise perturbations on biometric mobile applications [18]. Another study [19] evaluated the noise attacks on different biometrics including signature, fingerprint, and face. White noises were injected in given biometric templates to generate more synthetic data and a high false acceptance rate was observed by using the noise altered images. To address such type of noise perturbations, Gui *et al.* [6] proposed a defense mechanism based on the residual and statistical features of the authenticate samples.

D. Generative Models

Different from the static biometric traits (e.g., fingerprint, iris and face images), ECG is a continuous semi-periodical bio-signal and surely brings with slight variance among every single heartbeat of the same individual. Given this unique dynamic and continuous nature, in the real ECG attacking scenarios, it would be barely effective if only one or a few samples of the genuine user are falsified and forged. Therefore, it would be more aggressive and effective if a sufficiently large number of (ideally infinite) fake ECG samples that are slightly different from each other can be generated. To this end, generative models could be potential method to generate new samples based on the distribution over the observed data.

Currently, Generative Adversarial Network (GAN) and Variational Autoencoder (VAE) are the two most popular generative models. Although sharing the same rationale, they differ in their roots and training details. VAE [20] is rooted in probabilistic graphical models, which aims at learning the probability distribution of the observed data by modeling latent representation. GAN [21] alternatively trains a generator network and a discriminator network to find a Nash equilibrium where the generator can create fake samples that are real enough to fool the discriminator. Although GAN is believed to be effective for approximating complex distributions, it also has significant drawbacks and performance fluctuations. Compared with VAE, the training of GAN is unstable, because not only finding a Nash equilibrium cannot be guaranteed through gradient descent, but also its easy to fall into a "mode collapse" if the training of two adversarial networks are not balanced. In addition, training an accurate enough discriminator in GAN requires a large number of true samples, which is usually less feasible in biometric attacking scenarios.

III. Threat Model and Problem Formulation

A. Threat Model

For a privacy-preserving biometric authentication protocol, the major concerns can be concluded as follows [12], [22]:

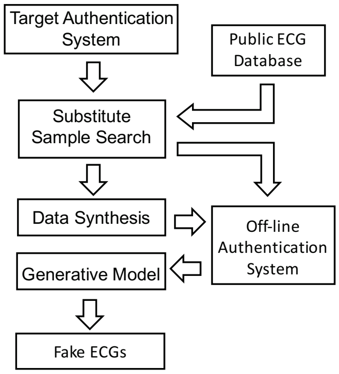


Fig. 2. Overview of the proposed attach scheme

- **Biometric template attack:** the adversary attempts to access the stored biometric reference templates.
- **Biometric sample attack:** the adversary intends to forge fake biometric samples that can be accepted by the authentication system.
- **Identity privacy:** the adversary tries to link the biometric of an authorized user based on references from different applications.
- **Transaction anonymity:** the adversary wants to trace the user in different authentication attempts based on learned templates and samples.

In this paper, we primarily focus on the first two security concerns, and formulate the following assumptions.

Assumption 1: The Target Authentication System (TAS) allows the user to have unlimited access attempts. Most of existing ECG biometric systems do not restrict the access times in order to achieve satisfactory user experience [23].

Assumption 2: TAS is able to effectively defend against the noise perturbation attacks, such as described in [6], [18].

Assumption 3: The attacker can freely access those publicly available ECG biometric databases online, such as MIT-BIH (48 subjects), PTBD (290 subjects), CYBHI (128 subjects) and UofTDB (1,012 subjects).

Assumption 4: The attacker has no access to the authentication system, which means that the attacker can only obtain the authentication decisions (Accept/Reject), instead of any detailed matching score or intermediate gradient information.

To emulate the real-world attacks, it is more reasonable and practical for attackers to acquire the authentication decisions for every malicious attempt, instead of the exact matching scores or other detailed metrics. Thus, the attack mechanisms which require the decision scores, such as black-box [24] and hill-climbing [15], are not considered in this paper.

B. Definitions

a) ***Definition 1: ECG Biometric****:* Let $\mathbb{B}$ represent the whole space of ECG biometric information for the entire human population. Specifically, b_k^i denotes the ECG of a single heartbeat k from an individual i and B^i indicates the complete ECG recordings from the individual i.

$$B^i = \{b_1^i, b_2^i, ..., b_k^i\}, \quad \forall b_k^i \in \mathbb{B}, \quad \forall B^i \subseteq \mathbb{B} \tag{1}$$

b) ***Definition 2: ECG Feature Extraction****:* Let $\mathcal{F}$ be the feature extraction method that transforms the raw ECG recordings to some certain feature dimensions (e.g., fiducial or non-fiducial features). u_k^i denotes the extracted features for every ECG recording b_k^i and U^i represents the complete feature set for the individual i.

$$U^i = \{u_1^i, u_2^i, ..., u_k^i\}, \quad u_k^i = \mathcal{F}(b_k^i) \tag{2}$$

c) ***Definition 3: ECG Template Matching****:* Let $\mathcal{H}$ represents the template extraction function, and O indicates the template matching function. Assume $\hat{b}$ is the unknown ECG sample, and $\Delta\varepsilon$ is the tolerance threshold.

$$O(\mathcal{H}(U^i), \hat{b}) = \begin{cases} > \Delta\ \varepsilon \Rightarrow \text{Reject} \\ \leq \Delta\ \varepsilon \Rightarrow \text{Accept} \end{cases} \tag{3}$$

C. Problem Formulation

a) ***Formulation 1: Adversarial Sample Search****:* The purpose is to find a few falsely accepted samples from the publicly available sources without intercepting or compromising the authentication system. Let B' denotes the public ECG datasets. $\mathcal{G}$ is the searching function that aims to find out the falsely accepted ECG sample b' in the database B' by the target template matching function O.

$$O(\mathcal{H}(U^i), b' \leftarrow \mathcal{G}(B')) \leq \Delta\varepsilon, \quad B' \subseteq \mathbb{B} \tag{4}$$

b) ***Formulation 2: Adversarial Sample Synthesis****:* Given the falsely accepted ECG sample b', to continuously break in the authentication system, let $\mathcal{C}$ represents the adversarial sample generation function that can keep generating fake ECG samples set $\mathbf{b}$ with a high acceptance rate.

$$O(\mathcal{H}(U^i), \mathbf{b} \leftarrow \mathcal{C}(b')) \leq \Delta\varepsilon \Rightarrow \text{Accept} \tag{5}$$

IV. Proposed Presentation Attack

A. Substitute Sample Searching

Different from other biometrics such as fingerprint and iris, ECGs originates from the bio-electric activity of the heart muscles. Thus, besides the uniqueness from different individuals, general biological structures and behavioral characteristics of human hearts enable ECGs to follow some common patterns. In addition, based on the FMRs in existing ECG biometric research (around 5%-10% [1]), given a database with a large population, it is possible that there are some substitute ECG samples in the database that can be falsely accepted by TAS.

1) Feature Extraction: To reduce the noise influences, we apply the One Dimension Multi-Resolution Local Binary Pattern (1DMRLBP) [25] to extract the features from each sample (Fig. 3). The coding mechanism considers different distances d and window sizes p. The LBP value is defined as:

$$\begin{aligned} BP(x(t)) = &\sum_{i=0}^{p-1} sign(x(t+i-p-d+1) - x(t))2^i \\ &+ sign(x(t+i+d) - x(t))2^{i+p} \end{aligned} \tag{6}$$

$$sign(x) = \begin{cases} 1 & \text{if } \mathrm{x} + \varepsilon \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where ε describes the quantization error.

Fig. 4(a) represents the raw 50 ECG samples from the same individual which are quite noisy, and Fig. 4(b) shows the corresponding LBP features which show very high consistency.

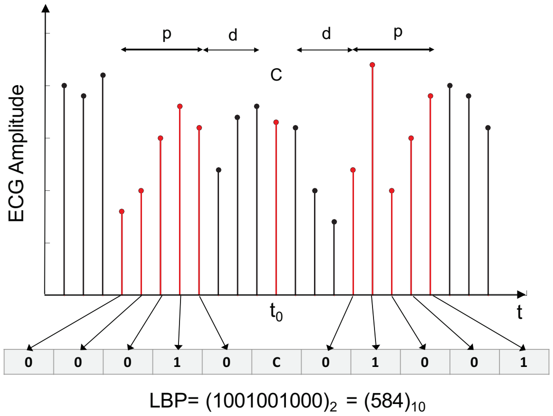
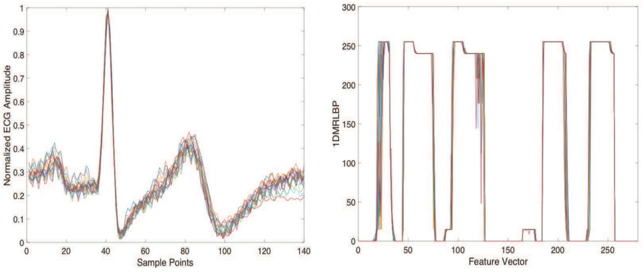


Fig. 3. Example of 1DMRLBP coding scheme at t_0 ($d = 4, p = 5$).



(a) ECG variance of a single user (b) 1DMRLBP extracted from (a)

Fig. 4. Comparison between the variance from the raw ECGs and the extracted 1DMRLBP features.

2) ECG Clustering: To reduce the searching space and attempt efforts, inspired by the stratified sampling, the Hierarchical Clustering Analysis (HCA) is used for dividing the given public database B' into k groups.

$$\{B_1, B_2, ..., B_k\} = \mathcal{C}(B') \tag{8}$$

where B_k denotes the kth ECG cluster, $\mathcal{C}$ is the Hierarchical clustering function. To better evaluate the similarity matrix for the ECG samples and eliminate the time-series variance caused by the changing heart rates, we use the Dynamic Time Warping (DTW) to measure the distance between each individual sample during the clustering. The optimization of k will be discussed in Section V-D.

3) Bayesian Decision Based Searching: Considering the parameter $\theta = \{B_1, B_2, ...B_k\}$ as the samples searched from the kth cluster and the evidence $X = \{0, 1\}$ as the return label from the TAS for each sample. We have *a prior* $p(\theta)$ which is the probability of choosing each cluster, and the observations x with the likelihood $p(x|\theta)$. Based on the Bayesian theory, the posterior probability which is the probability of acceptance samples coming from each cluster is defined as:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{9}$$

Similar with the dictionary guess attacks on passwords, we first cluster ECG samples into k groups, and rank initial searching probability $\{p_1, p_2, ...p_k\}$ for each cluster based on the corresponding sample density. After we find the first accepted sample b_i^m and its corresponding cluster $B_m (m \leq k)$, as the samples in the same cluster hold the relatively high similarities, we can assume that other samples in the cluster B_m will have a higher probability to be accepted by the TAS. As a traditional authentication system shown in Equation 3, the tolerance $\Delta\varepsilon$ is resulted from the self-variance for the enrolled user. Instead of only choosing the cluster with the highest probability, to ensure the diversity for the searched substitute samples and have a better estimation of the enrolled user's template and $\Delta\varepsilon$, we only increase the weight P_m for choosing the next sample from the cluster B_m by the its posterior probability $p(B_m|x)$.

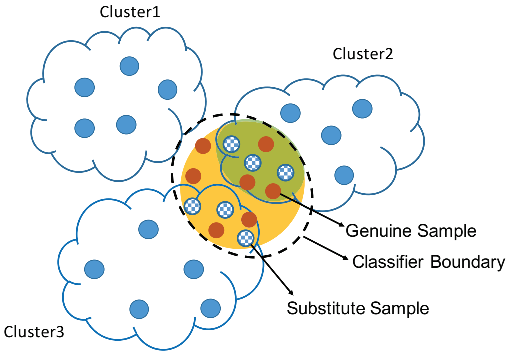


Fig. 5. An example of clustering based template search.

For example, as shown in Fig. 5, blue circles are the samples in the public database, and blue circles with lattice pattern are the substitute samples that can be accepted by the TAS. Given a database with a sufficiently large population, the samples of the enrolled user will have overlaps with our clusters. Once we find out the substitute samples (e.g., cluster 2 and 3), the corresponding selecting probabilities for cluster 2 and 3 will increase. Thus, the probability of searching the next substitute sample will increase compared with randomly searching among the entire database (i.e., brute force). Finally the enrolled user's template will be estimated within the yellow area. However, if we only search samples from cluster 2, the template will be restricted in the green region, which will affect the acceptance rate for further sample synthesis. The searching algorithm is shown in Algorithm 1.

4) Template Ranking Search: After identifying N substitute samples, to further reduce the searching effort, we asked one question: *Can we find out more substitute samples based on a few initial accepted samples without further accessing the authentication system?* According to the National Institute of Standards and Technology (NIST) [26], fingerprint biometric on average can achieve 1% FAR so far, which is believed to represent the best performance in all biometrics. On the other hand, most of existing research on ECG biometrics hasn't achieved an accuracy level like this [1]. Thus, we propose a hypothesis that, given the database with a sufficiently large population, there are at least 0.5% samples from other users that could be misclassified and falsely accepted. To validate this hypothesis, we evaluate the acceptance rates of the top 0.5% samples on our benchmark with 106 users, as shown in Fig. 6. As the EER (details in Sec V-B) increases, the average acceptance rate also increases to above 0.95.

B. Substitute Model Training

To reduce the attempt efforts and increase the acceptance rate of our forged ECG samples in Section IV-C and IV-D, based on the ranking samples, we train an off-line model as the substitute of the target authentication system. Instead of using regular neural network classifiers (e.g., Softmax), to mimic a more strict decision boundary for the accepted samples, the one-class SVM learning is used for the outlier detection. Different with the SVM for binary classification, the training

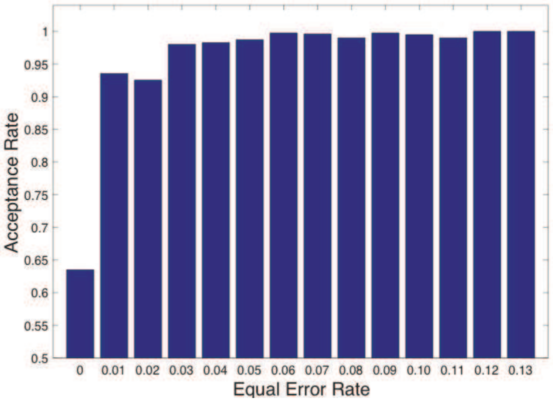


Fig. 6. Acceptance rate of the initial 100 samples from the template ranking search over 106 users with different EERs.

ALGORITHM 1: Substitute Sample Searching Algorithm

```
Input: $\{B_1, B_2, ..., B_k\}$: $k$ clusters for the given ECG database;
N: number of needed substitute samples;
S: O: target authentication system;
i: number of search attempts;
Output: $\{b'_1, b'_2, ...b'_N\}$: substitute ECG sample set
Set initial choosing probability $P_0 = \{p_1, p_2, ...p_k\}$ for $k$ clusters;
Start access the target authentication system;
Choose the initial sample $b_1^m$ from cluster $B_m$ under the
  probability $P_0$;
foreach $i$ do
    if $O(b_i^m) \rightarrow accept$ then
        calculate $P(\theta|x)$;;
        $W_m = W_m(1 + P(\theta|x))$; //update the $m$th cluster weight;
        $P_m = \frac{W_m}{\sum_{j=1}^k W_j}$; // update the choosing probability;
        $b_i^k \rightarrow b'_n$; // output the substitute sample;
    end
    $i = i + 1$;
    choose the new sample $b_{i+1}^k$ under updated cluster probability
      $P_i$;
end
```

data for one-class learning all come from the same label, and the objective function for one-class learning [27] is:

$$J = 0.5 \sum_{jk} a_j a_k G(x_j, x_k) \tag{10}$$

$$\sum a_j = n\nu, \quad 0 \leq a_j \leq 1, \quad j = 1, ..., n \tag{11}$$

where $G(x_j, x_k)$ is the element (j, k) of the Gram matrix (kernel function). ν is the parameter controlling the trade-off between the accuracy and weights. By setting an appropriate fraction of the observations as negative scores, we can get the optimized boundary for outlier detection. In our approach, all the substitute samples from the template ranking are used to train the one-class SVM.

C. Substitute Sample Synthesis

The above clustering-based template searching method can only find limited substitute samples from the public sources, which restricts the effectiveness and applicability of presentation attacks especially for the authentication system that continuously monitors the user's ECG signals. There, it is imperative to seek an alternative data synthesis approach that can augment our attacking sample dataset. Louis *et al.* [28] utilized the Multivariate Gaussian Distributions' deviation to generate more synthesized observations for the purpose of training data argumentation. Inspired by their work, we investigate a method to synthesize data based on Curve Fitting Model (CFM) and Autoencoder.

a) Curve Fitting Model: By defining the order of Gaussian distributions N (the number of peaks to fit), CFM can be represented as:

$$F = \sum_{i=1}^{N} f(x_i) = \sum_{i=1}^{N} a_i e^{[-(\frac{x-bi}{c_i})^2]} \tag{12}$$

where b_i and c_i represents the ith Gaussian distribution's mean and standard deviation among N Gaussian distributions. The parameter a_i can be interpreted as the weight of this i th Gaussian distribution. To find the optimal weight a_i in each Gaussian distribution. A cost function is defined as:

$$J = \sum_{i=1}^{N} |f(x_i) - y_i|^2 \tag{13}$$

where y_i denotes the target curve to be fitted, and $f(x_i)$ represents the result of fitting. Through calculating the first and second order derivatives of J, an optimized weight a_i can be generated for each Gaussian distribution. Our investigations show that, by properly choosing the CFM level (e.g., 5 or 6 Gaussian distributions), we can achieve a rather high fitting accuracy, while maintaining a certain level of randomness.

b) Autoencoder: Similar as the auto-regression model for learning the noise variance of ECG signals [29], we design an autoencoder to duplicate the substitute ECG samples (see Section IV-A4) while adding the unclonable and unobtrusive noises (error tolerance) during the reconstruction process that can still pass the TAS. However, due to the limited training samples, we apply CFM into the autoencoder's training phase to purposely randomize the reconstructed samples reflecting the variance nature of ECG signals.

c) Synthesis based on CFM and Autoencoder: Leveraging the CFM and autoencoder discussed above, we choose part of the data X_{ae} obtained by the substitute sample searching for training the autoencoder. The rest of data will be used as the templates T to generate synthesized data. For each template T_i, we contaminate the training data of the autoencoder X_ae gradually with the data generated by the CFM under different levels. According to different contamination ratios for the training data and the different levels of CFM, N autoencoders are trained for just one template T_i. It is worthy to note that, a higher N can lead to the risk that the synthesized samples are rejected by both TAS and replay detector, because of the decreasing variance and rather high similarity. Our results show that, only 53% synthesized data can pass the TAS when we set N as 20. For a higher N value, the success rate will keep decreasing. Although this approach can augment our attacking sample dataset, it is far enough to supply more high quality fake samples because of the limited capacity of the autoencoder in representing the dynamic variance of ECG signals. To this end, we further introduce the VAE-based generative model to generate a sufficiently larger set of fake ECG samples, based upon the substitute samples acquired above and the offline authentication model generated in Section IV-B to filter out low-quality synthetic data.

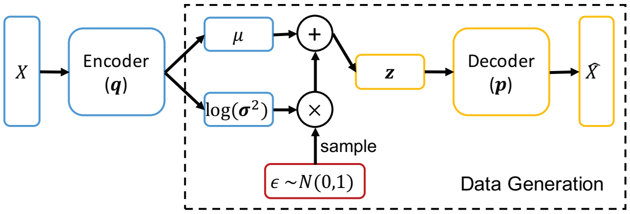


Fig. 7. Directed Graphic Model Representation of VAE

D. ECG Counterfeits Generation Using VAE

Because our goal is to generate a great number of fake ECG samples that are very similar but not exactly the same as the observed data (substitute and synthetic ECG samples), VAE addresses this issue with latent variables, which can be seen as the representation of the observed data. With a reasonable amount of latent variables, we can find a way to generate the desired samples. Figure 7 shows a directed graphical model, where z denotes latent variables that can be drawn from *a prior* distribution $p(z)$ and X indicates the observed data which have a likelihood $p(X|z)$.

Then, the first task of VAE is to infer the distribution of latent variables from the observed data, that is, to calculate the *posterior* $p(z|X)$. Since the *posterior* is intractable, it is approximated using variational inference in VAE, so that the true posterior $p(z|X)$ can be modeled with a family of simpler distributions (e.g., Gaussian) denoted as $q(z|X)$. This process is called recognition model, in terms of neural networks, it can be achieved by a "encoder" network with parameters ϕ. It brings in the observed data X and outputs the parameters to the distribution $q(z|X)$, namely the mean μ and variance σ^2 of the latent variables for each sample. Based on the representation of z parametrized by the recognition model, the likelihood of the data $p(X|z)$ can be parametrized with a generative model. A "decoder" network with parameters θ is adopted for data reconstruction and generation.

The "encoder" and "decoder" are connected with a reparameterization trick. It makes the network differentiable by diverting the non-differentiable sampling operation on μ and σ to a term ϵ out of the network, so that the "encoder" parameters ϕ can be trained with gradient descent. The reparameterization trick is implemented as follows:

$$z = \mu + \sigma\epsilon, \epsilon \sim \mathcal{N}(0, 1) \tag{14}$$

The neural network structure is trained by optimizing the parameters θ and ϕ with a loss function given in the form of the negative log-likelihood with a regularizer. Since we are considering the case with i.i.d. data samples, the loss function with regard to a single sample x_i can be represented as:

$$\mathcal{L}(\theta, \phi; x_i) = -E_{q_\phi(z|x_i)}[\log p_\theta(x_i|z_i)] + KL(q_\phi(z|x_i)||p_\theta(z)) \tag{15}$$

The first term is the reconstruction loss with the expectation taken over the latent variables z. Because the purpose is to maximize the likelihood $p_\theta(X|z)$, and the "decoder" is supposed to generate values between 0 and 1 (due to the normalization in our case), the $p_\theta(X|z)$ would be a multivariate Bernoulli [20]. The second term can be seen as a regularizer. It is the Kullback-Leibler divergence between the approximate posterior $q_\phi(z|x_i)$ and the prior $p_\theta(z)$. It tells how close are the two distributions when using q to represent p.

V. Experimental Setup

A. Dataset Selection

We adopted the UofTDB ECG biometric database [9], which has a large population size (1012 individuals), varying body postures, physical exercises and acquisition over a long period of time. The ECG signals in this database have a sampling frequency of 300 Hz with a 12-bit resolution. To guarantee the quality and stability of ECG signals, we implemented the strict outliers removal algorithm designed based on the Median Absolute Deviation [30] to filter out subjects whose samples have large variances and maintained the data of 606 subjects. Considering the *Assumption 3* in Section III, we randomly choose 400 subjects' data to constitute the Database for Attack (DfA), and the rest 206 subjects' data are used as Database for Defense (DfD).

B. Target Authentication System

1) Verification System: Our verification system is built based on DfD. Butterworth filter with cut-off frequencies 0.5 Hz to 40 Hz is implemented for removing baseline wander, electromyographic (EMG) signal noise, and power-line noise. R-peak detection is based on the Pan-Tompkins Algorithm [31]. The heartbeats are segmented into individual beats by a length of 200 msec before R peak and 500 msec after the R peak. Following the instructions of the existing work [32], we used discrete wavelet transform (5-level decomposition with db3 setting) as feature extraction and a customized Wavelet Distance measure (WDIST) method as classification. Each time, only one subject is chosen as the genuine user in DfD, all other 205 subjects are designated as the imposters. Each genuine user has his/her own verification system. As the target verification system, 66% of data of the genuine user is taken as the training data to generate template of each specific user. The data left are used as the test data for performance evaluation. Each decision (accept or reject) is made upon the performance of 5 beats. In this case, through calculation on these 206 verification systems for each genuine user, Equal Error Rate (EER) has its mean at 3.5% with standard deviation at 3.13%.

2) Residual-based Replay Detector: As most of popular biometric authentication systems based on bio-electricity signals have ability to defend against noise injection attacks, we utilized the residual distribution features described in [6] to design a replay detector, which showed a very high detection rate when the injection noise has a SNR lower than 30. For even lower level of noise injection, we also developed a defense mechanism using the Percent Residual Difference (PRD) metric shown below, which has been used to gauge the similarity of ECG biometrics [32].

$$PRD = \sqrt{\frac{\sum_{i=1}^{N}(s_0(i) - s_n(i))^2}{\sum_{j=1}^{N}(s_0(i) - \bar{s_0})^2}} \times 100\% \tag{16}$$

where s_0 and s_n is the unknown signal and the enrolled signal respectively.

Figure 8 presents the performance (i.e., false acceptance rate, FAR) of white Gaussian noise injections with different SNRs under the different replay detection mechanisms. It is clearly seen that, our implemented TAS equipped with both two replay detectors can successfully identify and reject all fake samples, regardless of the level of injected noises.

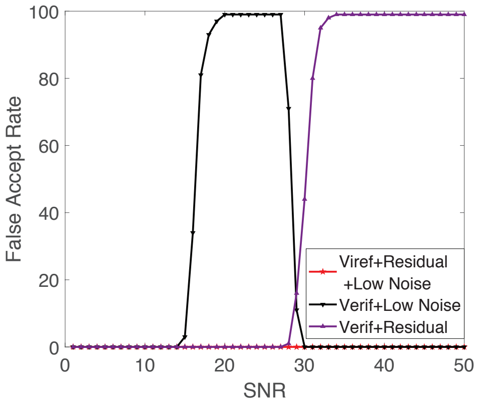


Fig. 8. Performance of Gaussian White Noise Injection on TAS with replay detectors

TABLE I. VAE Configuration

Model Type	Layer #	Layer Type	# of Neurons
Encoder	1	input	140
	2	hidden	60
	3(1)	mean	10
	3(2)	log var	10
Decoder	4	latent z	10
	5	hidden	60
	6	output	140

C. VAE Configuration

For the implementation of VAE, we adopted a simple fully-connected neural network structure — multi-layer perceptron (MLP). The "encoder" and "decoder" each contains a single hidden layer. The output layer of the "decoder" employs the sigmoid activation function while all other layers adopt rectified linear unit (ReLU). The network configuration of our VAE model is declared in Table I. The 140-dimensional synthetic ECG samples are normalized and fed into the "encoder" network. Two layers with the same dimension of 10 are drawn from the "encoder", indicating the variational means and log values of variance. The latent representation z is reparameterized based on the sampling from ϵ, which has the same dimension as z, μ and σ. The "decoder" takes in the latent variables and reconstructs some new data samples.

The whole network was trained for 100 epochs utilizing the random mini-batches with the size of 100. After training, the "encoder" can be disconnected and data generation would be done by the standalone generator model without any explicit input samples, except for the learned means and log variances. To evaluate the performance and capacity of VAE, 5000 ECG counterfeits were created for different genuine users.

D. Parameter Optimization

For ECG clustering, given a public ECG dataset with 400 individuals, as shown in the Algorithm 1 and Fig. 5, the number of clusters will largely influence the searching efficiency. For example, if the number of clusters k is too large, the clustering model will be over-fitted, and the searching effort for finding the first accepted samples will also be significant. On the contrary, if the k is too small, the knowledge about the enrolled user's templates will not be fully discovered, and the performance will be close to the brute-force searching. During the optimization, as the cluster numbers don't affect the acceptance rate for samples from template ranking, we only evaluate the attempt times for k ranging from 4 to 128 under different numbers of initial searched substitute samples. The result of the user with 1% EER is shown in Fig. 9, for different N, the optimal attempt times all locate at the convex points when k ranges from 60 to 80. By averaging the smallest cluster numbers k for different initial searched substitute sample numbers N, we obtain the optimal cluster number $k_{optimal}$ as 68.

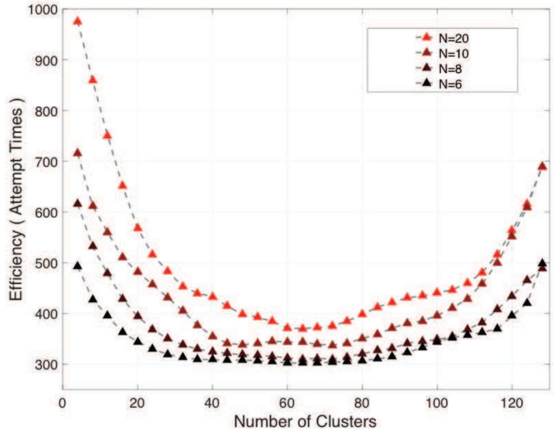


Fig. 9. Relationship between the number of clusters k and the averaged attempt times with different initial substitute sample numbers N for the enrolled user with 1% EER.

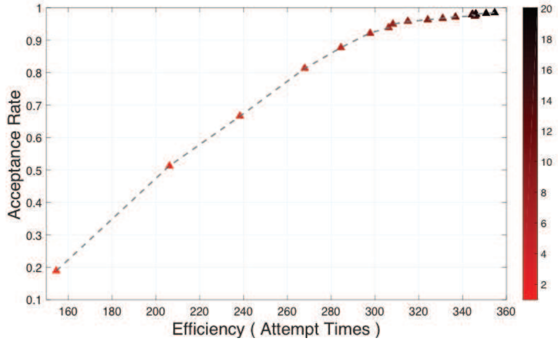


Fig. 10. Trade-off between the acceptance rate of 100 substitute samples from template ranking and the attempt times over different initial substitute samples N ranging from 1 (red) to 20 (black). Each triangular indicates the {accuracy,efficiency} pair for different N.

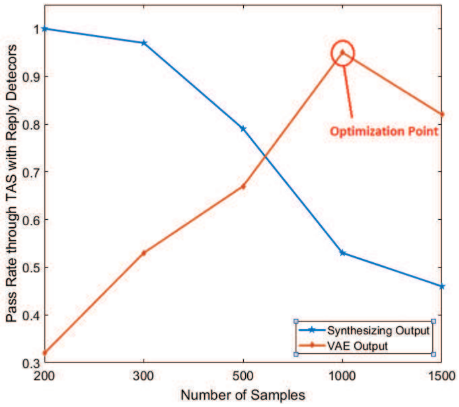


Fig. 11. System Pass Rate of Synthesized Data and VAE Output

In addition, after setting the cluster number k as 68, to determine the best number of initial substitute samples N, we evaluate the performance in terms of accuracy and efficiency over different values of N ranging from 1 to 20. As shown in Fig. 10, the triangular with darker color means the larger N. As the number of initial substitute samples increases,

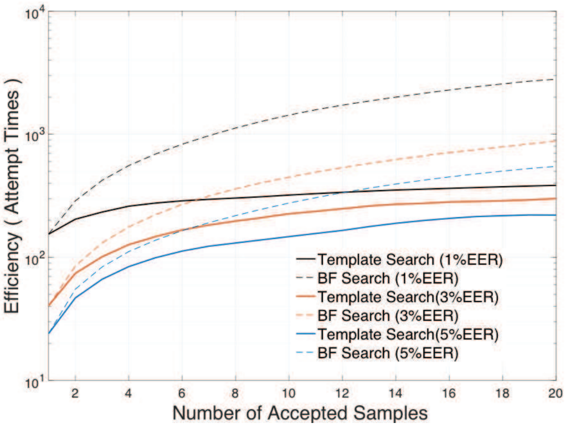


Fig. 12. Average attempts times using template search and brute force search over different number of accepted samples for users with 1%,3%,5% EER

the corresponding acceptance rate for 100 samples originated from the template ranking based on those N initial substitute samples also increases. Meanwhile, unsurprisingly it reduces the searching efficiency by requiring more attempt times. To achieve a balanced trade-off between the accuracy and efficiency, we select $N = 8$ as the optimal number of initial substitute samples.

To find the optimal number of training samples for VAE, we let the 5000 generated ECG samples by VAEs with different numbers of train ing samples (i.e., 200, 300, 500, 1000, 1500, respectively) go through TAS defined in Section V-B. It is observed from Figure 11 that, the optimal number of training samples for VAE (red line) is around 1000, which can lead an average pass rate at 95.37% (TAS with EER at 1%, 2%, 3%, 4% and 5%). Meantime, to provide more training samples for VAE, we need to optimize the parameter N (the number of autoencoders in Section IV-C) and thus let the synthesized samples go through TAS. However, the pass rate deceases dramatically with the increasing number of synthesized samples (blue line). This demonstrates that, the proposed substitute sample synthesis tends to generate more similar samples with little variance so that it can only work effectively in generating a set of synthesized samples necessary for the training of VAE, far away from sufficient to serve as the presentation attacks directly onto the ECG biometric TAS.

VI. Performance Evaluation

A. Substitute Sample Searching

We compared our proposed searching mechanism with the standard brute-force (BF) attacks based on the same public database. Five different enrolled users with EER of 1%, 2%, 3%, 4%, 5% respectively are randomly selected. To present the BF attack, we randomly and independently selected ECG samples from the public database to access the TAS. For our searching approach, based on the feedback from the TAS, we updated the weight associated with each cluster to better estimate the specific cluster that the substitute samples belong to. According to Fig. 12, as the increase of the number of required accepted samples, our method requires much less attempts than the BF search. The detailed comparison results for searching 8 accepted substitute samples are listed in Table II. It is shown that our proposed mechanism can effectively reduce the attempt efforts, especially for users with lower EERs. It is worthy to note that those numbers are based on acquiring 8 accepted samples, which aims at learning the full characteristics of authentic ECGs and thus generating unlimited artificial ECGs. That being said, even for the TAS with 1% EER, we can still successfully access the system within every 40 attempts on average, which is operationally feasible and much less than 140 attempts for brute-force attacks. Moreover, given the fact that many commercial authentication systems have the exponentially growing lockout time to wait for successive failed login attempts, our proposed method would be more feasible and efficient.

TABLE II. Comparison of Avg. Attempts for searching 8 accepted substitute samples under different EERs

Avg. Attempts	1% EER	2% EER	3% EER	4% EER	5% EER
Template Search	312	219	132	129	107
BF Search	1158	392	222	204	185

TABLE III. Acceptance Rates

Users	Acceptance Rate for Each Step/Module			
	Authentication	Residual Detector	Replay Detector	TAS
1% EER	96.34%	100%	99.82%	96.16%
2% EER	100%	100%	99.20%	99.20%
3% EER	94.82%	100%	99.94%	94.76%
4% EER	98.94%	100%	100%	98.94%
5% EER	97.20%	100%	90.50%	87.80%

B. Performance Comparison Against Noise Injections

We compared the performance of proposed fake sample generation against the conventional noise injection perturbations. As shown in Fig. 8, the TAS with low noise and residual-based replay detectors (red line) can successfully identify and reject all noise-injected fake samples (i.e., FAR = 0), regardless of the level of injected noises. Similar results were also observed for other types of noises, including white, blue, red, pink and violet noises.

Leveraging our proposed presentation attack methodology describe above, 5,000 counterfeit ECG samples were generated associated with each genuine user. In this paper, as targeting a commercial authentication system, we only consider the enrolled users with EER less than 5% (a high EER rate's authentication system could be considered unsafe and will be easily attacked by brute force). The attacking results of all 5 TASs are shown in Table III, which includes the respective acceptance rates of the authentication system, the residual detector, the replay detector, and the entire TAS. It can be seen that, our approach can not only achieve very high acceptance rates on the authentication system itself, but also pass through the strict verification of the residual-based replay detectors. Especially, the residual detector was completely cracked by our approach with all 100% acceptance rates. The results demonstrate that the fake ECG samples generated using our method can fully mimic and represent the intrinsic characteristics of authentic ECG samples, that is, a set of sufficiently similar counterfeits containing the dynamic variance by nature. Specifically, Figure 13 provides a more intuitive view of the comparison between 50 generated fake ECG samples and 50 authentic ones associated with the 1% EER user, which indicates that the generated samples can perfectly imitate the morphological shapes of the true data.

VII. Discussion and Conclusion

As the increasing popularity of bio-signal-based biometrics, it becomes more necessary to investigate the potential security risks and threats on emerging biometrics given

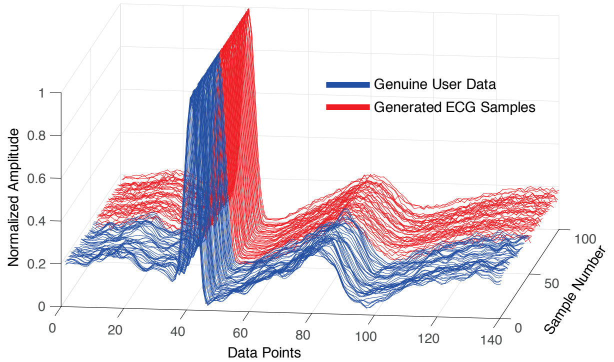


Fig. 13. Comparison Between Generated ECG Samples Using Our Scheme and the True Data With Regard to 1% EER User

their unique characteristics. In this paper, we explored the vulnerability of the ECG biometric authentication system and proposed a black-box presentation attack relying on the limited access efforts and feedback labels from the unknown verification classifier. Leveraging the public databases, we first obtained 8 accepted substitute ECG samples and selected 100 samples from the database based on the similarity ranking. Then we synthesized more fake samples using CFM and autoencoder to train a generative model, VAE. For each of the 5 users with different EER levels, 5,000 counterfeit ECG samples were generated from VAE (theoretically, infinite new samples can be generated). The results show that the fake samples can successfully pass through the verification of the target authentication system and the enhanced replay detectors, with very high acceptance rates cross all 5 genuine users. In summary, our approach possesses a much higher success rate than the noise injection attacks and demands much less attempt efforts than the brute-force attacks, under the same condition.

VIII. Acknowledgment

This work is supported by the National Science Foundation under grants CNS-1422417 and CNS-1564046.

References

[1] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "ECG biometric recognition: A comparative analysis," *IEEE Trans. Inform. Forensics Secur.*, vol. 7, no. 6, pp. 1812–1824, 2012.

[2] B. Biggio, G. Fumera, G. L. Marcialis, and F. Roli, "Statistical meta-analysis of presentation attacks for secure multibiometric systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 561–575, 2017.

[3] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Adv. Sig. Proc.*, vol. 2008, p. 113, 2008.

[4] M. Ferrara, R. Cappelli, and D. Maltoni, "On the feasibility of creating double-identity fingerprints," *IEEE Trans. Inform. Forensics Secur.*, vol. 12, no. 4, pp. 892–900, 2016.

[5] L. R. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proc. IEEE*, vol. 91, no. 12, pp. 2021–2040, 2003.

[6] Q. Gui, W. Yang, Z. Jin, M. V. Ruiz-Blondet, and S. Laszlo, "A residual feature-based replay attack detection approach for brainprint biometric systems," in *Proc. Int'l Workshop on Inform. Forensics Secur.*, 2016, pp. 1–6.

[7] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 54–65, 2015.

[8] S. Eberz, N. Paoletti, M. Roeschlin, M. Kwiatkowska, I. Martinovic, and A. Patané, "Broken hearted: How to attack ECG biometrics," in *Proc. Network and Distributed System Security Symp.*, 2017, pp. 1–15.

[9] S. Pouryayevali, S. Wahabi, S. Hari, and D. Hatzinakos, "On establishing evaluation standards for ECG biometrics," in *Proc. Int'l Conf. Acoust., Speech and Signal Process.* IEEE, 2014, pp. 3774–3778.

[10] J. Bonneau, "The science of guessing: Analyzing an anonymized corpus of 70 million passwords," in *Proc. IEEE Symp. Security and Privacy*, 2012, pp. 538–552.

[11] A. Roy, N. Memon, and A. Ross, "MasterPrint: exploring the vulnerability of partial fingerprint-based authentication systems," *IEEE Trans. Inform. Forensics Secur.*, vol. 12, no. 9, pp. 2013–2025, 2017.

[12] E. Pagnin, C. Dimitrakakis, A. Abidin, and A. Mitrokotsa, "On the leakage of information in biometric authentication," in *Proc. Int'l Conf. in Cryptology in India*. Springer, 2014, pp. 265–280.

[13] M. Martinez-Diaz, J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia, and J. A. Siguenza, "Hill-climbing and brute-force attacks on biometric systems: A case study in match-on-card fingerprint verification," in *Proc. 40th Int'l Carnahan Conf. Security Technology*, 2017, pp. 151–159.

[14] A. Kong, D. Zhang, and M. Kamel, "Analysis of brute-force break-ins of a palmprint authentication system," *IEEE Trans. Syst., Man, Cyber. B*, vol. 36, no. 5, pp. 1201–1205, 2006.

[15] E. Maiorana, G. E. Hine, and P. Campisi, "Hill-climbing attacks on multibiometrics recognition systems," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 5, pp. 900–915, 2015.

[16] C. Rathgeb and A. Uhl, "Attacking iris recognition: An efficient hill-climbing technique," in *Proc. 20th Int'l Conf. Pattern Recognition*, 2010.

[17] D. Muramatsu, "Online signature verification algorithm using hill-climbing method," in *Proc. Int'l Conf. Embedded and Ubiquitous Computing*, 2008.

[18] S. Ghouzali, M. Lafkih, W. Abdul, M. Mikram, M. El Haziti, and D. Aboutajdine, "Trace attack against biometric mobile applications," *Mobile Information Systems*, vol. 2016, 2016.

[19] J. G. Herrero, "Vulnerabilities and attack protection in security systems based on biometric recognition," Ph.D. dissertation, Universidad Autónoma de Madrid, 11 2009.

[20] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[21] I. Goodfellow, J. Pouget-Abadie, B. Mirza, M.and Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Inform. Proc. Systems*, 2014, pp. 2672–2680.

[22] K. Simoens, J. Bringer, H. Chabanne, and S. Seys, "A framework for analyzing template security and privacy in biometric authentication systems," *IEEE Trans. Inf. Forensics .Security*, vol. 7, no. 2, pp. 833–841, 2012.

[23] F. P. Karegar, A. Fallah, and S. Rashidi, "Using recurrence quantification analysis and generalized hurst exponents of ECG for human authentication," in *Proc. 2nd Conf. Swarm Intelligence and Evolutionary Computation (CSIEC)*. IEEE, 2017, pp. 66–71.

[24] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against deep learning systems using adversarial examples," *arXiv preprint arXiv:1602.02697*, 2016.

[25] W. Louis, M. Komeili, and D. Hatzinakos, "Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ecg biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2818–2832, 2016.

[26] E.Tabassi, C. Watson, G. Fiumara, W. Salamon, P. Flanagan, and S. L. Cheng, "Performance evaluation of fingerprint open-set identification algorithms," in *Proc. Int'l Joint Conf. Biometrics*, 2014, pp. 1–8.

[27] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[28] W. Louis, S. Abdulnour, S. J. Haghighi, and D. Hatzinakos, "On biometric systems: electrocardiogram gaussianity and data synthesis," *EURASIP J. Bioinform. Syst. Biol.*, vol. 2017, no. 1, p. 5, 2017.

[29] D. F. N. Karimian, D. Woodard, "On the vulnerability of ecg verification to online presentation attacks," in *Proc. Int'l Joint Conf. Biometrics (IJCB)*. IEEE, 2017.

[30] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *J. Exp. Soc. Psychol.*, vol. 49, no. 4, pp. 764–766, 2013.

[31] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, no. 3, pp. 230–236, 1985.

[32] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 248–253, 2008.