2012

Secure Speech Biometric Templates

Keerati -. Inthavisas
Lehigh University

Secure Speech Biometric Templates

by

Keerati Inthavisas

Presented to the Graduate Committee
of Lehigh University
in Candidacy for the Degree of
Doctor of Philosophy

in
Computer Engineering

Lehigh University
January 2012

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date

Dissertation Advisor/Committee Chair

Dr. Daniel P. Lopresti

Accepted Date

Committee Members:

Dr. Mooi Choo Chuah

Dr. Hector Munoz-Avila

Dr. Tiffany Jing Li

To my parents and my wife

Mr. On Inthavisas
Mrs. Rachada Inthavisas
Mrs. Panida Inthavisas

Acknowledgments

I have been working in biometric security for several years. I would like to thank my advisor, Professor Daniel Lopresti, for his research direction and guidance. I started working from scratch. Without him this work would not have been possible. Most importantly, I am now passionate about this area and eager to study further. I would also like to thank my committee members – Professors Mooi Choo Chuah, Hector Munoz-Avila and Tiffani Jing Li – for their helpful comments.

I would like to thank Jim Glass, who provided the MIT mobile device speaker verification corpus for this research. I gratefully acknowledge the anonymous subjects who devoted their time to the data collection. I would like to thank Candice Quinones, an adjunct of English as a Second Language (ESL) at Lehigh University, for her review of this manuscript.

I would like to thank my parents for their love and support. I would also like to thank my wife for her encouragement in finishing this work.

Finally, I would like to acknowledge the financial support of the Thai government.

Contents

Acknowledgments **v**

List of Tables **xi**

List of Figures **xv**

List of Algorithms **xvi**

Abstract **1**

1 Introduction **3**

1.1 Attacks Against Speech Biometric Systems 7

1.2 Dynamic Time Warping-based Biometric Key Binding (DBKB) . . . 10

1.3 Speech Cryptographic Key Regeneration based on Password 12

1.4 Dissertation Outline 14

2 Related Work **17**

2.1 Attack Against Biometric Systems 17

2.2 Biometric Cryptographic Systems 20

2.3 Template Protection Approaches 22

3 Background **26**

3.1 Speech Signal Processing 26

3.2 Discrete Fourier Transform (DFT) 29

3.3 Linear Predictive Coding (LPC) 30

3.4 Mel-Frequency Cepstrum Coefficients (MFCC) 32

3.5 Short-term Energy 34
3.6 Speaker Verification Models 35

4 Attack Against Speaker Verification 42
4.1 Introduction 42
4.2 Datasets 44
4.2.1 The MIT Mobile Device Speaker Verification Corpus 44
4.2.2 The Lehigh Quiet Environment Speaker Verification Dataset . 45
4.3 Speaker Verification Models 46
4.3.1 Dynamic Time Warping (DTW) 47
4.3.2 Vector Quantization (VQ) 47
4.3.3 Gaussian Mixture Models (GMM) 48
4.4 Attack Models 48
4.4.1 The Human Type with Assumption I (H-I) 48
4.4.2 The Human Type with Assumption II (H-II) 49
4.4.3 The Human Type with Assumption III (H-III) 49
4.4.4 The Algorithmic Type with Assumption I (A-I) 49
4.4.5 The Algorithmic Type with Assumption II (A-II) 50
4.5 Experiments and Results 53
4.5.1 Experimental Setup 53
4.5.2 Experimental Results 55
4.6 Summary 58

5 Dynamic Time Warping-based Biometric Key Binding 62
5.1 Introduction 63
5.2 Dynamic Time Warping-based Biometric Key Binding (DBKB) . . . 66
5.2.1 Hardening Template 68
5.2.2 Mapping the Biometric to a Binary String 70
5.2.3 Multi-thresholds Generation 71
5.2.4 Biometric Key Retrieval 74
5.3 Experiments and Results 74
5.3.1 Experiments Setup 74

5.3.2 Performance 75
5.3.3 Security Analysis 81
5.4 Summary 85

6 Performance and Security of the Hardened Template 87
6.1 Introduction 88
6.2 Transformation Approach 89
6.2.1 Experimental Setup 90
6.2.2 Experimental Results 92
6.3 Security of the Hardened Template 95
6.3.1 Experimental Setup 98
6.3.2 Experimental Results 100
6.4 Summary 100

7 Speech Cryptographic Key Regeneration based on Password 102
7.1 Introduction 103
7.2 Speech Cryptographic Key Regeneration based on Password (SCKRP) 106
7.2.1 Enrollment: Initialization 106
7.2.2 Enrollment: Regeneration 107
7.2.3 Verification 110
7.3 Experimental Setup 111
7.4 Experimental Results 112
7.4.1 One-layer Scheme 113
7.4.2 Two-layer Scheme 113
7.4.3 Security Analysis 115
7.5 Summary 117

8 Conclusion and Future Work 119
8.1 Conclusions 119
8.2 Future Work 121

Appendix A: Datasets 123
Appendix A.1: List of Pass-phrases in the MDS 123

Appendix A.2: List of Dedicated Users' Pass-phrases in the MDS 125
Appendix A.3: List of Pass-phrases in the LDS 126
Appendix A.4: Speech Corpus 127

Appendix B: List of Passwords in the Experiments **135**
Appendix B.1: Passwords in the MDS 135
Appendix B.2: Passwords in the LDS 137

List of Abbreviations **138**

List of Notations **140**

Bibliography **142**

Curriculum Vitae **152**

List of Tables

4.1 FARs (%) of speaker verification systems (DTW, VQ, and GMM) against various attacks using decision thresholds at operating points of imposters (H-II). 57

5.1 FARs (%) of speaker verification systems (DTW, VQ, GMM, and DBKB) against various attacks using decision thresholds at operating points of imposters (H-II). 82

5.2 The security of the multi-thresholds and the global-threshold scheme in the MDS. 85

6.1 Equal Error Rates (EERs) with the 95% confidence interval of the transformation approach (transformed template) and the unprotected approach (unprotected template) for the DTW-based systems against random attack, imposter, and generative. 93

6.2 Equal Error Rates (EERs) with the 95% confidence interval of the DBKB when the transformed template and hardened are applied. . . 94

6.3 Equal Error Rates of the modified pass-phrase attack, the original imposters' pass-phrases, the adversary using template information only. 100

7.1 EERs of speaker verification systems in the MDS and LDS against imposter attack for Dynamic Time Warping-based (DTW), Dynamic Time Warping-based Biometric key Binding (DBKB), and our approach (SCKRP) in the case that one of the applied password layer is excluded. Scenario I: genuine, Scenario II: compromised password, and Scenario III: compromised biometrics 115

7.2 EERs of two-layer SCKRP in the MDS and LDS against imposter attack (H-II). Scenario I: genuine, Scenario II: compromised password, and Scenario III: compromised biometrics 116

List of Figures

1.1 Points of attack in a generic biometric system [74]. 5

3.1 Block diagram of speech signal processing. 28

3.2 Block diagram of a speaker verification system. 35

3.3 An example of Dynamic Time Warping where the mapping between two signals (A and B) is given by the dot line. 36

3.4 An example of the VQ-based speaker model with 10 clusters. 39

3.5 An example of the GMM-based speaker model with 10 Gaussian Mixtures. 41

4.1 Block diagram of the VQ and GMM speech biometric user authentication. 51

4.2 The error rates of regenerated pass-phrases by varying κ in VQ and GMM system. 54

4.3 The EERs against various attacks and models with the 95% confidence interval for the same-gender experiment (a) the LDS and (b) the MDS 59

4.4 Comparisons of the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS (a) the H-I (b) the H-II 60

4.5 Comparisons of the A-II for the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS. 61

5.1 Categorization of template protection schemes. 64

5.2 Dynamic time warping-based biometric key binding in training phase. 65

5.3 Dynamic time warping-based biometric key retrieval in verification phase. 73

5.4 The error rates of regenerated pass-phrases by varying κ in VQ, GMM, and DBKB. 76

5.5 The EERs against various attacks and models with the 95% confidence interval for same-gender experiments on the DTW, VQ, GMM, and DBKB (a) the LDS and (b) the MDS 77

5.6 The performance of the DBKB against attackers using random pass-phrases (Random), true pass-phrases (Imposter), and the templates (Template): (a) the LDS (b) the MDS. 78

5.7 Comparisons of the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS (a) the H-I (b) the H-II 80

5.8 Comparisons of the A-II for the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS. 81

5.9 The comparison of a distribution of Normalized Hamming Distance and a binomial distribution (a) Distribution of Normalized Hamming Distances obtained from 4,512 comparisons of inter-speaker in the MDS and (b) A binomial distribution with $\mu = 0.5$ and $N = 120$ degrees-of-freedom. 84

6.1 The ROC curves of three approaches: Unprotected, transformation, and our approach (DBKB). (a) the imposter trial and (b) the random trial. 96

6.2 The EERs of attackers using true pass-phrases (Imposter), random pass-phrases (Random), and generative attack (Template): (a) the transformation approach is applied to the DBKB's DTW template (b) the hardened template is utilized. 97

7.1 Enrollment phase: Initialization. 106

7.2 Enrollment phase: Regeneration. 107

7.3 Biometric key retrieval in verification phase. 110

7.4 Distribution of users' passwords that are comprised of one word, combination of two or more words, unfamiliar numbers, familiar numbers, string of numbers and letters, or string of numbers, letters, and symbols. 112

7.5 ROC curves of two-layer scheme. Scenario I: genuine, Scenario II: compromised password, and Scenario III: compromised biometrics (a) the MDS (b) the LDS 114

List of Algorithms

1 Specification of the Hardening algorithm 69
2 Specification of the Mapping algorithm 72

Abstract

The security of biometrics against attacks is a a serious concern in biometric personal authentication systems. In particular, the security of biometric templates is a topic of rapidly growing importance in the area of user authentication.

In this dissertation, we investigate the security of Dynamic Time Warping (DTW), Vector Quantization (VQ), and Gaussian Mixture Model (GMM) methods that have been used in speaker verification systems. We present attack models based on adversary knowledge. We start with naive adversaries without knowledge of an authentic speaker and develop them into highly knowledgeable adversaries who know the speaker's information, have the speaker's voice samples, acquire the speaker's template, and know the algorithm used by the speaker verification system. We propose an analysis-synthesis forgery in which the informed adversary can exploit information, such as feature vectors from the template and a statistical probability from the voice samples of the target speakers to regenerate a forgery that can be used in remote or on-line authentication. We show that the effectiveness of the regenerated forgery is better than the other attack models. In addition, we have demonstrated that the traditional approach to evaluate the security of speech biometric speaker verifications was insufficient. These results raise important issues for researchers when attempting

to demonstrate the security of speech biometric systems.

We then describe our approach to cryptographic-based speaker verification. We present a new scheme to transform speech biometric measurements (feature vector) to a binary string which can be combined with a pseudo-random key for cryptographic purposes. We utilize DTW in our scheme. The challenge of using DTW in a cryptosystem is that a template must be useful to create a warping function, while it must not be usable for an attacker to derive the cryptographic key. In this work, we propose a hardened template to address these problems. We evaluate our scheme with two speech datasets and compare with baseline DTW, VQ, and GMM speaker verifications. The experimental results show that the error rates of the proposed scheme against attackers utilizing the template information significantly outperform the DTW, VQ, and GMM speaker verifications. For the other attack models, the recognition performance of the proposed scheme outperforms the VQ and GMM. It is slightly degraded when compared to the DTW speaker verification.

Finally, we propose a way to strengthen a system by combining a password with a biometric cryptosystem. We show that attackers have to spend more time to search for the keys when we compare our scheme with a password approach. In addition, the security provided by the proposed framework remains unaffected even when the password is compromised, since the scheme only utilizes a password that is independent of the key used in the system. The experimental results show that the scheme enhances security and improves recognition performance of the system.

Chapter 1

Introduction

For more a decade, biometrics as cryptography has been an interesting area because of the inability of humans to remember strong passwords [4, 65]. The traditional approach uses a password to release a cryptographic key, but it is easy to guess using dictionary attacks [30, 53]. Hence, users have to select unusual keys for their passwords that are easy to forget. To address these problems, biometrics are used to combine or generate a cryptographic key to apply to applications, such as file encryption and user authentication for two reasons. First, it is hard to get past the biometrics compared to a common eight character password. Second, biometrics are human characteristics, so they cannot be forgotten.

Biometrics can be divided into two classes: physiological and behavioral biometrics. Physiological biometrics measures the shape of the body that experts can recognize, such as fingerprints, iris codes, and DNA or by humans, such as faces. Behavioral biometrics measures the action of the person, such as typing rhythm, handwriting or signature, gait, and voice. The applications that use physiological biometrics may

face two problems. First, an adversary can acquire physiological biometrics easily, for example, the ubiquitous fingerprint on the surface and the image of user by camera. As the users cannot change their physiological biometrics, their key may be compromised. The other problem is that it is inadvisable to use the same key in all applications. Users should be concerned about the security of their keys. If one of their applications is compromised, the other applications will be in danger. Although some researchers proposed the ways to re-issue a new key after the previous key has been compromised [36, 75], when the key was compromised, the biometrics might be derived from the old template [4]. For this reason, behavioral biometrics can be used to alleviate this concern. The users can change their key by changing their behavioral biometrics. However, the questionable security of the biometric system against adversary attacks is a concern.

There are eight points of attack in a generic biometric system as indicated in Figure 1.1 [74]. For the first type, the attacker presents fake biometrics to the sensor. Examples include a fake finger, a forgery signature, a face mask, and an imposter passphrase. For the second type, the attacker bypasses the sensor and resubmits an old recorded signal. Examples include an audio recording. For the third type, the feature extractor could be replaced with a Trojan horse program so that it would produce the feature sets desired by the attacker. For the fourth type, the extracted features may be replaced with synthesized or modified features which match the stored template. For the fifth type, the matcher is overridden by the attacker. Consequently, the system always produces a high or low match score. For the sixth type, the attacker can modify the template so that the attacker can submit feature vectors which match the modified template. For the seventh type, the transmitting template may be captured

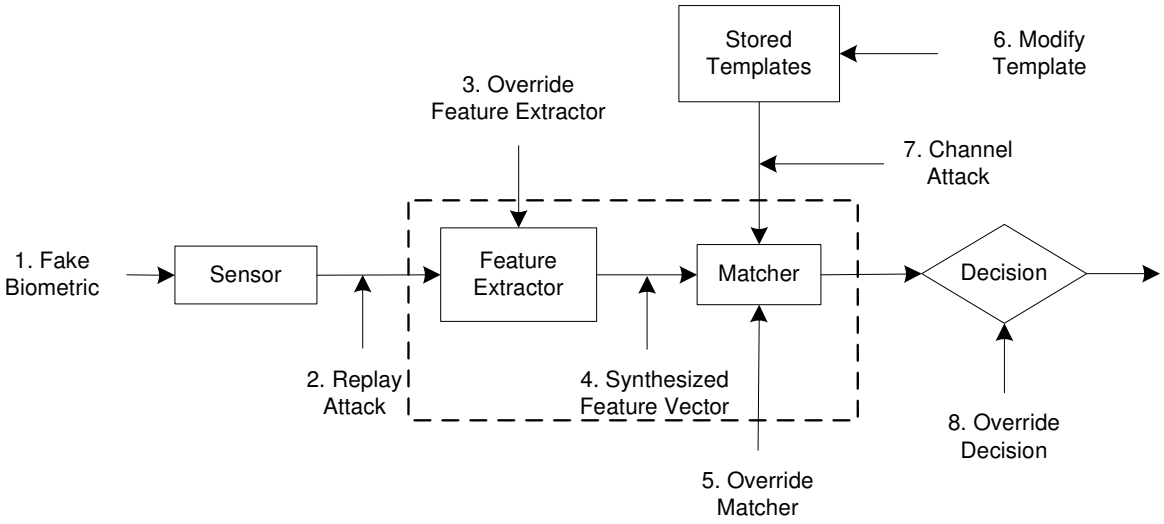


Figure 1.1: Points of attack in a generic biometric system [74].

or changed before arriving at the matcher. For the eighth type, the decision module is overridden.

In this dissertation, we focus on the first, second, fourth, and seventh types of attack. For the first, second, and fourth types of attack, two classes of adversaries [2] have to be considered: human or algorithmic. For both classes, the ability of the adversary to compromise the system depends on their knowledge of private information, public information and the motivation [5]. The knowledge of private information includes target biometrics and public information includes auxiliary information such as the construction of the authentication system and templates [2]. The motivation depends on how important gaining access to the system is to the adversary.

For the third, fifth, and eighth types of attack, we assume that the system uses secure code execution or specialized hardware which can protect the system from attacks, such as Trojan horse attacks.

For the sixth and seventh types of attack, we assume that the stored templates are located in a secure location, such as a central database. However, the templates have to be transmitted to the matcher for verification. Thus, attackers may capture the template information at this point. Then, they create a physical spoof from the template information to replay to the sensor or matcher. A common way to secure a channel is to encrypt the data with cryptographic keys. Nevertheless, the system we refer to is a cryptographic-based biometric system; it is unreasonable to assume that we can access the keys from other sources. Hence, it is necessary to protect the stored template.

We choose to study speech-based biometrics. Several reasons for investigating the security of speech-based biometrics are: 1) The system is inexpensive compared with the implementation of other kinds of biometrics (e.g., iris or fingerprints). Why is it inexpensive? Because of the ubiquity of cell phones and microphones embedded in computers. 2) Voice is behavioral biometrics; users can change their pass-phrases easily. 3) Attacks against speech biometric templates have not been studied as much as attacks against other kinds of biometrics (e.g., fingerprints or handwriting).

The major motivation of the dissertation is based on the concern for the security of speaker verification systems. To demonstrate this issue, three popular systems are used in our experiments: the Dynamic Time Warping (DTW), Vector Quantization (VQ), and Gaussian Mixture Model (GMM). We employ attack models with various assumptions proposed in the literature to evaluate the security of the systems. Moreover, we propose a new attack model under assumptions we have created. The attack models and assumptions are briefly mentioned in Section 1.1 and fully discussed in Chapter 4. In the remaining chapters, we contribute to design schemes to protect a

speaker verification system against attacks and demonstrate the security of our system in a rigorous way. The details are briefly provided in Section 1.2 and 1.3. For full detail, see Chapter 5, 6, and 7.

1.1 Attacks Against Speech Biometric Systems

Biometric authentication systems are vulnerable to attacks. Two classes of adversaries have to be considered: human or algorithmic [2]. For both classes, the ability of the adversary to compromise the system depends on their knowledge of private information and public information as we introduced earlier.

Motivation

Three speaker verification systems based on a pattern matching technique are popularly used. According to [15], the pattern matching methods include a template, a codebook, and a statistical model. The DTW is used in the template model, the VQ in the codebook model, and the GMM in the statistical model. Even though a number of researchers [73, 62, 99, 61, 45] reported results which demonstrate the performance and security of these systems, the experiments were conducted using different methodologies (systems, datasets, assumptions, and attack models). Therefore, it is difficult to compare these systems.

Moreover, much of the work to demonstrate the performance and security has been done on the GMM-based speaker verification system. Thus far, the attacks on the DTW and VQ system reported in literature were based on human attacks that include random and imposter trial [71, 32, 37, 46, 78]. For ran-

dom trials, the researchers assumed that an adversary did not know the actual pass-phrase. For the imposter trial, they assumed that an adversary knew the pass-phrase and said the actual pass-phrase without listening to the authentic speaker pass-phrase. For the GMM, both human and algorithmic attacks were reported. Beyond the random and imposter trial, there was informed imposter trial where the researchers assumed that an adversary mimics the pass-phrase by listening to the authentic speaker pass-phrase. Another trial was an algorithmic attack. For this trial, the researchers assumed that an adversary knew the pass-phrases and acquired some voice samples of authentic speakers. Then, the adversary used this information to synthesize the pass-phrase. Thus, all attack models proposed in the literature should be used to investigate the security of all mentioned systems. We note that the mentioned trials are considered as the first type of attack.

One of the most serious attacks is against the stored template. Attacks on the template can lead to three vulnerabilities according to Jain et al. [44]. First, if the template is replaced by the attackers, they can use the input signal that corresponds to the replaced template to gain unauthorized access. Second, the attackers can utilize the template information to create a forgery. Lastly, the stolen template can be directly replayed to the matcher. One way to secure the biometric system is to put all the system modules and interfaces on a smartcard. However, the template can be gleaned from a stolen card [44]. Therefore, to demonstrate the security of the stored template, we should assume that an adversary knows the pass-phrase and acquires the stored template. Hence, the mentioned trial falls under the fourth and seventh types of attack.

While the template attack is one of the most serious concern, we have not seen any reports of this kind of attack for speaker verification systems elsewhere in the literature.

Contribution

To demonstrate these issues of concerns, we attack systems using human and algorithmic attacks. We study the state-of-the-art in speaker verification systems: DTW, VQ, and GMM. We investigate the security of those systems by doing a series of experiments that include both human and algorithmic attacks. We use the attack models proposed in literature to evaluate the security of the systems. In addition, we propose an analysis-synthesis forgery in which the highly informed adversaries can exploit information, such as feature vectors from the template and a statistical probability from the voice samples to regenerate a forgery that can be used in remote or on-line authentication. We conduct experiments in the same controlled environment (datasets, instruments, and assumptions). We show that the DTW yields the best recognition performance when we compare it with the VQ and GMM using the attack models proposed in the literature; the error rates of the DTW system are the best. Unfortunately, the error rate of the DTW against our attack model is significantly higher than the VQ and GMM error rates. These results suggest that if the DTW template is protected properly, it will be better than the other systems; our results also appear in [40].

1.2 Dynamic Time Warping-based Biometric Key Binding (DBKB)

When comparing two sequences in speech biometrics, the main problem is that the duration of the same biometrics provided by the same user at a different time changes with non-linear expansion and contraction. The solution to this problem is to use DTW to set up a non-linear mapping of one signal to another by minimizing the distance between two signals [79]. To utilize DTW, we need a template as a keying or reference signal *(reference template)* to set up a warping function for incoming inputs. Then, the result (warped signal) is compared with a *matching template* to decide whether to accept a user.

Motivation

In [40], we have shown that the reference and matching templates leak information to an adversary. Therefore, both templates must be protected. An ideal biometric template protection scheme should possess four properties [60]. 1) Diversity: Different templates must be used for different applications. 2) Revocability: A compromised template can be canceled and re-issued. 3) Security: It must be computationally hard to invert the secure template to the original template. 4) Performance: The system using the secure template should not degrade the recognition performance. Speech biometrics satisfies the first two properties as the users can easily change their biometric samples. The remaining properties are the critical issues that we will focus on. In particular, the proposed schemes in the literature typically apply a transformation function to

transform features. However the features in a transformed domain will degrade recognition performance.

There is a scheme in literature to protect a DTW template by using a non-invertible method for signature authentication [58]. Even though the authors proved that to recover the original templates was computationally as hard as random guessing [59], the system left the transformed templates which could be used in gaining access to the system.

Contribution

We present a scheme to protect the DTW templates. The scheme is used to create a hardened template which is useful in creating a warping function, while it is not usable for an attacker in gaining access to the system. As the hardened template is only used to create a warping function, an input signal is not transformed. Hence, the result remains unaffected by a transformation function which is utilized in the literature (e.g., [58], [75]). For the matching template, it is protected by cryptographic framework. To combine a secret with biometric information, we present a scheme to transform behavioral biometric measurements (feature vectors) to a binary string which can be combined with a pseudo-random key for cryptographic purposes. The binary string, as a requirement, should appear to be random in the context of cryptography. We propose a mapping algorithm using multi-thresholds that are determined by incorporation with pseudo-random bits. Hence, the algorithm can generate a binary string to meet the requirement.

We also present empirical results based on public dataset and our dataset. The

assumptions we made remain the same as the previous section (Section 1.1). Hence, we can directly compare results from this section with the previous. The experimental results show that the proposed scheme outperforms the VQ and GMM for all attack models. It is slightly degraded when compared to the DTW for the attack models proposed in the literature but, for the algorithmic attack we propose, our scheme significantly outperforms the other systems. Our results appear in [39].

Next, we compare the transformation approach proposed by Maiorana et al. [58] with ours. The experimental results show that our system outperforms the transformation approach. We also show that an adversary can exploit the transformed template to gain access to the system which does not differ from the unprotected approach [32]. These results appear in [41].

1.3 Speech Cryptographic Key Regeneration based on Password

Motivation

One of the promising ways to authenticate users is to combine a biometric cryptosystem with one or both of the other factors: knowledge or token. Therefore, the performance is improved in the case that the biometrics and the input factors are not compromised simultaneously. There are a number of works involving combination of biometrics with a password or a random key [70, 49, 90]. These works suffer from two main problems: 1) The error rate is still high. 2)

A hill-climbing attack is possible because a decision threshold is stored in the system.

In this section, we assume that the attacker acquires the biometric information (the second type of attack). This assumption is reasonable because, in our case (speech), this could happen with audio recording. However, it could also happen in other cases; the ubiquitous fingerprint left on a surface and the image of a face or an iris captured on camera are examples of these cases. Therefore, this scenario has to be investigated.

The other problem is that the security of the mentioned systems does not differ from a traditional password approach when the biometrics is compromised.

Contribution

We propose a way to combine a speech biometric cryptosystem with a password. The system consists of three layers. For the first layer, the biometrics is transformed using a password. Then, we map the transformed version to a binary string. For the second layer, the result from the second layer is permuted using a password in such a way that the attackers cannot discriminate the correct password from brute-force search if the biometrics is not compromised. For the third layer, a cryptographic key and the binary string are hidden using a fuzzy commitment framework so that it makes a hill-climbing attack more difficult as the attackers are not left with the match score to decide whether the attack is close to the original biometrics.

We show that the verification performance of the system meets the same level of a traditional password-based approach if biometrics and password are not

compromised simultaneously. Furthermore, the system increases the computational time for attackers to search for the key. Even if the attackers acquire the biometrics, they are forced to align query biometrics each time they guess the password.

1.4 Dissertation Outline

Chapter 2: We explore past research related to ours and explain how our work builds on or improves it.

Chapter 3: We describes the theory of speech signal processing and the speaker verification techniques related to this dissertation.

Chapter 4: We study the state-of-the-art in speaker verification systems: Dynamic Time Warping (DTW), Vector Quantization (VQ) and Gaussian Mixture Model (GMM). We investigate the security of those systems by doing a series of experiments that include both human and algorithmic attacks. We propose an analysis-synthesis forgery that can be used in remote or on-line authentication. We find that the DTW yielded the best recognition performance for the text-dependent speaker verification, but it was the most susceptible to attack.

Chapter 5: We propose a cryptosystem for a text-dependent speaker verification named DBKB (Dynamic time warping-based biometric key binding). We utilize DTW in our scheme. A DTW-based biometric user verification system needs a DTW template to set up a warping function for query biometrics. In addition, a matching template is required to examine similarity. The challenge

of using DTW in a cryptosystem is that a template must be useful to create a warping function, while it must not be usable for an attacker to derive the cryptographic key. In this chapter, we propose a hardened template to address these problems. For the matching template, it is protected by cryptographic framework. The experimental results show that the performance of the proposed scheme outperforms VQ and GMM. It is slightly degraded when compared to the DTW speaker verification.

Chapter 6: In this chapter, we investigate the performance and security of transformation approach applied in a DTW-based system. We compare the transformation and unprotected approach with the DBKB. The experimental results show that the DBKB outperforms the transformation approach. Moreover, it is slightly degraded when we compare it with the unprotected template. We also show that an adversary can exploit the transformed template to gain access to the system which does not differ from the unprotected approach.

Chapter 7: In this chapter, we propose a way to combine a password with a speech biometric cryptosystem. We present two schemes to enhance verification performance in a biometric cryptosystem using password. Both can resist a password brute-force search if biometric is not compromised. Even if the biometric is compromised, attackers have to spend many more attempts in searching for cryptographic keys when we compare ours with a traditional password-based approach. In addition, the experimental results show that the verification performance is significantly improved.

Chapter 8: This chapter is conclusion of this dissertation and suggested future work. We propose a new algorithmic attack based on template information to demonstrate that the traditional approach to evaluate the security of speech biometric user verification is insufficient. Then, we develop the cryptographic-based speaker verification to protect the biometric templates. Lastly, we use a password to protect stored templates, enhance security, and reduce error rates in biometric cryptosystems. Our techniques offer great potential to protect the speech biometric template. In further research, we will investigate security and performance of other behavioral biometrics, such as handwriting or a signature. Finally, the permutation technique we proposed is a general technique which we believe can be used in any biometric modalities for the key binding scheme. In further research, we will investigate security and performance of our technique for physiological biometrics, such as fingerprints, faces, and iris codes.

Chapter 2

Related Work

In this chapter, we first explore the evaluation of various attacks on the biometric systems. Next, we explore biometric cryptographic systems. Then, we review template protection approaches.

2.1 Attack Against Biometric Systems

Many biometric modalities are susceptible to attack because some information is leaked from the biometric template [4]. Moreover, the security against the attack is underestimated [5]. Lopresti and Raim proposed a generative model to attack a handwriting authentication system [57]. The basic units of user's handwriting samples were manually segmented and then the corresponding units were concatenated to form a user' pass-phrase. Next, a feature space search was performed within a predetermined time limit of 60 seconds. As a result, their attack succeeded 49% of the time. In later research, Ballard et al. expanded the work in [57]. In their work

[5], they conducted a series of attacks using human-based and generative models to attack a handwriting authentication system. For the human-based attack, they used trained imposters who were allowed to select and replay real time renderings of a target user's pass-phrase in the experiments. The authors showed that the FAR (False Acceptance Rate) of the trained imposters when compared to the untrained counterparts significantly increased. For the generative model, the results showed that the generative attack match or exceed the effectiveness of forgeries rendered by the trained imposters.

In chapter 4, we adopt the attack models mentioned above to investigate speech-biometric authentication systems: DTW, VQ, and GMM. The results are similar to the work in [57, 5]. Then, we further evaluate the systems when we assume that the attacker acquires template information. We propose algorithms to regenerate pass-phrases. The experimental results show that our algorithms outperform the other attack models.

There are a number of successful attacks against speaker verification [73, 62, 99, 61, 45]. Yee, Wagner and Tran [99] reported the attack on a GMM-based speaker verification system against human impostors. Two people, male and female, played the roles of imitators against 138 speakers in the YOHO database. At first, the imitator was required to speak the same utterance in the YOHO database to calculate the similarity score between the imitator and the same-gender subject in the YOHO database. Among 138 speakers, the closest, intermediate, and furthest speakers have been selected by similarity scores. Finally, each imitator tried to mimic three target speakers in the database. The best result of the female imitator was accepted 30% of the time by the system, while the male imitator achieved a 35% acceptance rate.

Pellom and Hansen [73] investigated the security of a GMM-based speaker verification system with an Equal Error Rate (EER) of 1.45% as the baseline for human impostors. They proposed a new trainable speech synthesis algorithm based on trajectory models of the speech Line Spectral Frequency (LSF) parameters to synthesize the target voice. After the algorithm has been applied, the result showed that the FAR was increased from 1.45% to 86%. Jin et al. [45] presented an attack on a classical GMM-based speaker identification system using a voice transformation technique. First, the models of each speaker were created from a set of training data, and then the test data from the unknown speaker was used to test the system. The baseline of the model was 100% accuracy. The experimental results showed that impostors using voice transformation were able to fool the GMM-based speaker identification system. In other words, the GMM system always hypothesizes the speaker that is used as the target speaker for voice transformation.

Even though these attacks have been done with GMM-based systems and the same database was used (in [99] and [73]), the experiments were conducted with different datasets. Moreover, these results are not necessarily applicable to other systems (DTW and VQ). To make these issues clear, we conduct experiments under the same controlled environment, such as assumptions, datasets, and instruments. These results are shown in Chapter 4.

Masuko et al. [61] presented an attack model using synthetic speech. An HMM-based speaker verification with an FAR of 0% for human impostors was used as a baseline. They used an HMM-based speech synthesizer to create synthetic speech. The results showed that the FAR against the HMM-based synthesized speech increased to 70% by using only 1 sentence as training data. However, De Leon et al. proposed

a technique to detect the synthesized speech [24]. The fact that the HMM-based synthesizer always produces the same optimal waveform in term of likelihood score was exploited to detect the synthetic speech. With this step, the speaker verification system was able to prevent some imposters using synthesized speech. However, to determine the likelihood score, the system needs a template. Therefore, the template attack we propose in Chapter 4 may work on this system.

2.2 Biometric Cryptographic Systems

A number of researchers have proposed biometric cryptosystems. Monrose et al. [64] proposed a behavioral biometric key generation based on keystroke biometrics. They use dynamic features (duration of keystrokes and latencies between keystrokes) to strengthen a user's password. This scheme makes the system more secure by adding 15 bits of entropy to the password for 15 dynamic features [95]. In their later works [65, 66, 67], they applied this scheme to voice data. The algorithm to generate cryptographic keys from voices was mainly based on the speaker verification and identification technologies, such as digital signal processing, feature extraction, and the vector quantization technique. Consequently, their system was eventually able to generate cryptographic keys up to 60 bits from voice features. However, the False Rejection Rate (FRR) was still high (20%). Moreover, since their approach was based on the VQ technique, the system left useful information which can be used in gaining access to the system. In Chapter 4, we show that we can exploit VQ template information to regenerate a pass-phrase. As a result, the error rate of the demonstrated system significantly increases. In Chapter 5, we show that the Dynamic

Time Warping-based cryptographic key regeneration, which we propose, yields better recognition performance when it is compared with the VQ technique.

Garcia-Perera et al. [33] proposed a way to generate cryptographic keys based on speech recognition. The phoneme of the user's pass-phrase was trained and mapped to binary by using a Support Vector Machine (SVM) classifier. However, their scheme could generate a short length of key; the bit length was equal to the number of phoneme in the pass-phrase. For example, in one of our datasets, the minimum number of phonemes is eight. Therefore, this scheme can generate only an eight-bit key which is very short when it is compared with a traditional password-based approach, which is insufficient for security applications.

Hao and Chan [35] proposed a way to generate biometric keys from hand-written signatures. The DTW template was protected by utilizing static features as the DTW template so that the template did not reveal the key that was generated from dynamic features. Their approach achieved 40 bits of entropy with 1.2% acceptance rate. However, Ballard et al. [4] demonstrated that the dynamic features can be reproduced given the static features by using temporal inference techniques that they proposed. The experimental results showed that the keys were accurately recreated 22% of the time on the first attempt, and approximately 50% of the keys were correctly guessed after making fewer than 2^{15} guesses.

In later research, Hao et al. [36] proposed the combining of biometrics and cryptography with a two factors scheme: biometrics and a token. They stored a lock data (encoded keys combine with biometrics) in a smartcard which can be unlocked and decrypted at later time by user's biometrics. The template was hidden by following the fuzzy commitment scheme [48]. They were able to generate 140 bits from iris

codes with 44 bits of entropy. The authors also reported that the benefits of their scheme include revocability. More precisely, if the key is compromised, the new key will be issued. However, the new key has to be hidden with biometric information (iris code) which cannot be changed. Therefore, if the key is compromised, the biometrics might be derived from the old template [4]. For our approach, we also utilize the fuzzy commitment scheme, but our construction is more flexible. If the key is compromised, the new key will be issued and the users will be required to provide a new biometric measurement to the system. Hence, the old biometrics will be canceled.

2.3 Template Protection Approaches

Ratha et al. [75] proposed cancelable fingerprint templates. The non-invertible transformation functions were used to transform fingerprint feature (minutiae) positions so that the matcher could still be applied in feature domain. The result showed that there was a trade-off between discriminability and non-invertibility. In this proposal, three transformation functions were proposed: cartesian, polar, and functional. The Cartesian transformation yielded the best security. However, the performance was relatively poor. In addition, Nagar et al. [69] have shown that Ratha et al.'s scheme was vulnerable to intrusion attack because it was relatively easy to obtain a pre-image of the transformed template.

Numerous researchers proposed schemes to protect biometric templates by incorporating with random keys or passwords [82, 90, 80, 3, 70, 49]. These systems satisfy these criteria 1) The attackers cannot discriminate the correct password from incorrect when they use a brute-force search to find the key without the knowledge of

biometrics. 2) When the password is compromised, it cannot be used to reveal the key.

Ballard et al. [3] used a password to encrypt selected biometric features and some helper data for their key generation scheme. Their construction follows the approach similar to [7] where a low-entropy password is used to encrypt a high-entropy string. The features were specified as indexes into a table, and then a subset of the features was randomly assigned to each user. The feature indexes of this subset were encrypted in the template with a password using a cipher with arbitrary finite domain [8]. In this way, any passwords that are used to decrypt the template yield a subset of features indexes that falls within the global table. The authors ensured the indistinguishable from decryption with the correct and incorrect password by assigning any given feature with the same probability across the population to a user. Therefore, in both cases, a decrypted template appears as a random permutation on a subset of feature indexes. They demonstrated that their scheme did better than the previous approaches against attacks even when the password was compromised. However, the error rate when the password was not compromised was not reported.

Nandakumar et al. [70] proposed a scheme to secure a fingerprint with a password. The password was used to select a transformation function to secure the fingerprint template. The transformed template was then secured using fuzzy vault framework. Finally, they used a key derived from a password to encrypt the vault. By using their scheme, the attackers are required to know the correct password before they can guess the key. Even if the correct password is selected, the security of the scheme is still at the same level as before using a password. Benefits of their scheme include template revocability, prevention of cross-matching, enhanced security and a reduction in the

False Accept Rate. However, their scheme noticeably affects the False Reject Rate.

By utilizing the idea from Hao et al. [36]., Kanade et al. proposed a three factors scheme (biometrics, smartcard, and password) to apply to iris codes where a password was used to permute the key [49]. They could generate the key of 198 bits (compared to 140 bits in [36]) with estimated entropy of 83 bits (compared to 44 bits in [36]). Unfortunately, their scheme creates a security loophole which allows the attacker to crack the helper data without any additional information [85].

Teoh and Chong [90] proposed secure speech template protection in speaker verification system. The speech template was hidden through the random subspace projection process. In this process, a speech feature matrix is integrated with a user-specific key to obtain a random-projected matrix which cannot be inverted to the original speech feature matrix. The random-projected matrix is used to form a speaker probabilistic model and a decision threshold. They showed that the verification performance was very high. However, it would make some attacks, such as hill-climbing easier, as the system left the decision threshold and random-projected vectors for matching process. In the case that the token is stolen, the attacker may make small changes in the input imposter's feature matrix and check to see how the match score changes. After a number of iterations, the attacker may be able to acquire a feature matrix that is close to the original.

To address the problems mentioned above, in Chapter 7, we propose schemes to combine a speech biometric cryptosystem with a password. We first transform the biometrics using a password. Then, the transformed version is mapped to a binary string. In this way, the transformation process forces the attackers to run dynamic programming every time they try another password. Next, the biometric information

is permuted using a password in such a way that the attackers cannot discriminate the correct password from brute-force search even when the biometrics is compromised. Lastly, a cryptographic key and the biometric information are hidden using a fuzzy commitment framework to protect the matching template so that it makes a hill-climbing attack more difficult as the attackers cannot discriminate whether the attack yields a positive result.

Chapter 3

Background

In this chapter, we describe the theory of speech signal processing and the speaker verification techniques which are utilized in this dissertation.

3.1 Speech Signal Processing

A speech signal is usually represented by a function of time, $s_a(t)$ in which t denotes time. The first step is the transformation of an analogue signal to digital. This process is called A/D conversion. The analogue signal is usually sampled at 8kHz. The reason is that most information in human speech is at frequencies below 10,000 Hz. For speech communication networks, only frequencies less than 4,000 Hz are transmitted. Thus an 8,000 Hz sampling rate is sufficient; frequencies less than or equal to 4k Hz. can be reconstructed according to the Nyquist theorem [25]. Hence, we can use a low-pass digital filter with a cut-off at 4 kHz to strip the higher frequencies from the signal. If we denote the sampling period as T_s, the digital signal will be represented

by the following equation.

$$s[n] = s_a(nT_s), n = 0, \ldots, N-1 \tag{3.1}$$

The next step is pre-emphasis, which is the process to raise the Signal to Noise Ratio. The signal is pre-emphasized by passing the signal to a first order digital filter represented by the following equation where γ ranges between 0.9 to 1 [31].

$$H[z] = 1 - \gamma z^{-1} \tag{3.2}$$

Framing is the next step. The signal is framed into the short time analysis interval. These frames have to be overlapped properly. The length of each frame is usually around 30 msecs; This length would yield good results for speech processing with 10 msecs overlap [31]. Each frame is multiplied by a window function to reduce abrupt changes at the start and the end of each frame. The result can be represented by equation 3.3.

$$x[n] = s[n] \cdot w[n] \tag{3.3}$$

The processes we mentioned above are illustrated in Figure 3.1

The next step is feature extraction where the features are extracted from the signal. The following features are popularly used for speech analysis.

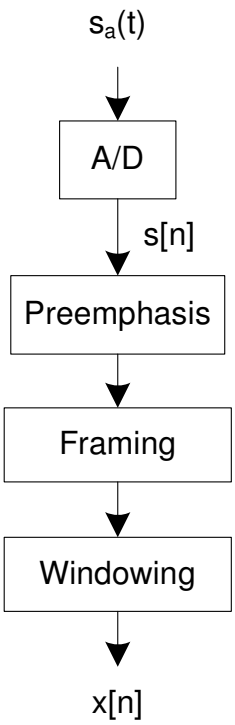


Figure 3.1: Block diagram of speech signal processing.

3.2 Discrete Fourier Transform (DFT)

A basic feature of voice is the DFT. The DFT of N points signal $s[n]$ for $k = 0, \ldots, N-1$ can be defined as:

$$X[k] = \sum_{n=0}^{N-1} s[n] \exp \frac{-j2\pi nk}{N} \tag{3.4}$$

The Inverse Discrete Fourier Transform (IDFT) for $n = 0, \ldots, N-1$ can be defined as:

$$s[n] = \sum_{k=0}^{N-1} X[k] \exp \frac{-j2\pi nk}{N} \tag{3.5}$$

According to the real function property [72], if $s[n]$ is real and $s[n]$ and $X[k]$ are transform pairs, then

$$X[-k] = X[N-k] \tag{3.6}$$

This symmetric property, equation (3.6), can be exploited to decrease the computation required to transform a real sequence. To derive DFT, there is no need to compute X for $N/2 < k < N$, since these values can be found from the first half of X.

The most efficient feature to identify a speaker is known as cepstral coefficients or cepstrum [52]. Cepstrum physically represents the movement of articulators (the teeth, alveolar ride, hard palate, and velum) of speakers [66]. Its use is popular because of low correlation [50]. Hence, it is appropriate to apply it for cryptographic purposes.

3.3 Linear Predictive Coding (LPC)

The coefficients of the Linear Predictive Coding play an important role in the speech signal processing. The concept of predicting the future signal from past samples was introduced in the late 1940's [72]. As the adjacent samples of the speech waveform are highly correlated, each sample can be approximated by a linear combination of the past samples. More precisely, an approximation of a speech signal $s[n]$ can be calculated by a linear combination of the LPC coefficients and P previous samples of the original signal (autoregressive model). Basically, the following equation represents this concept where $a[p], p = 1, \ldots, P$, are the LPC coefficients, P is the order of the linear predictor (number of the LPC coefficients).

$$s[n] \approx a[1]s[n-1] + a[2]s[n-2] + \ldots + a[p]s[n-P] \tag{3.7}$$

Given $Gu[n]$, an excitation term, equation 3.7 can be converted to equation 3.8 where $u[n]$ is a normalized excitation and G is the gain of the excitation.

$$s[n] = \sum_{p=1}^{P} a[p]s[n-p] + Gu[n] \tag{3.8}$$

We can express equation 3.8 in z-domain as indicated in equation 3.9.

$$S[z] = \sum_{p=1}^{P} a[p]z^{-p}S[z] + GU[z] \tag{3.9}$$

Considering equation 3.8 as an IIR filter, the transfer function of the filter is given by the following equation.

$$H[z] = \frac{S[z]}{U[z]} = \frac{G}{1 - \sum_{p=1}^{P} a[p] z^{-p}} \tag{3.10}$$

The LPC coefficients can be obtained by solving AR equations [72]; two methods can be used: covariance and autocorrelation method. The autocorrelation method is computationally more effective and stable since the positive definiteness of the correlation matrix is guaranteed by the definition of the correlation function, an inverse matrix exists for the correlation matrix [31]. For the autocorrelation method, the LPC coefficients can be obtained by solving the equation 6.3 where R is the autocorrelation sequence defined in equation 3.12.

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(P-1) \\ R(1) & R(0) & \cdots & R(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(P-1) & R(P-2) & \cdots & R(0) \end{bmatrix} \times \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(P) \end{bmatrix} = - \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(P) \end{bmatrix} \tag{3.11}$$

$$R[p] = \sum_{n=1}^{N-p} s[n] s[n-p] \tag{3.12}$$

However, the LPC coefficients are highly correlated [77]. A better feature set (less correlated) is Linear Predictive Cepstral Coefficient (LPCC) [52]. The LPCC, c_{LPCC}, can be obtained from equation 3.13 where $c_{LPCC}[0] = In(G)$.

$$c_{LPCC}[v] = \begin{cases} a[v] + \sum_{k=1}^{v-1} \frac{k}{v} c_{LPCC}[k] a[v-k] & 1 \leq v \leq P \\ \sum_{k=1}^{v-1} \frac{k}{v} c_{LPCC}[k] a[v-k] & v > P \end{cases} \tag{3.13}$$

3.4 Mel-Frequency Cepstrum Coefficients (MFCC)

Another feature that is also popular in speech processing is Mel-Frequency Cepstrum Coefficients. If we say that cepstral coefficients are derived from a speech production mechanism, MFCC will be derived from a speech recognition mechanism. For MFCC extraction, we use a set of non-linear scaled filters or a filterbank to filter the signal in a way similar to the human perceptual system. In [50, 38], the filterbank was defined as a set of band-pass filters whose frequency responses are triangular shape and whose center frequencies spread non-linearly across the frequency range of speech. Specifically, the scaled filterbank is approximately linear below 1kHz and logarithmic above. To derive MFCC with N samples, the magnitudes of Discrete Fourier Transform are passed to the filter, then the output of each filter is used to form MFCC parameters using discrete cosine transform.

From the explained above, let H_m for $m = 1, \ldots, M$ be a set of filters in the filterbank, the log-energy output of each filter $S[m]$ can be expressed as

$$S[m] = In\left[\sum_{k=1}^{N-1}\Big|X[k]\Big|^2 H_m[k]\right], \quad 0 < m \leq M$$

The MFCC, $c_{MFCC}[v]$, is then the discrete cosine transform of $S[m]$ that can be expressed as

$$c_{MFCC}[v] = \sum_{m=1}^{M-1} S[m]cos(\pi v \frac{(m-1)}{2M}) \tag{3.14}$$

The inversion of Mel-Frequency Cepstrum Coefficient

For this section, the objective is to find a signal that yields the desired MFCCs.

Consequently, when we pass the signal into the speaker verification system, the result is the desired MFCCs.

For L frames with M-order MFCC of the speech signal, let A and B represent the output of the filterbank and cosine transform matrix where the inner product of A and B is the MFCC, we rewrite equation (3.14) as

$$C = BA \tag{3.15}$$

where C is a MFCC matrix.

We multiply both sides of (3.15) with B^{-1} where B^{-1} is the inversion of B. Then equation (3.15) will be

$$B^{-1}C = A$$

Hence, the output of the m^{th} filter for l^{th} frame will be $S[m,l] = 10^{(B^{-1}C[m,l])}$ for l=*1, . . . , L*.

The next step is to interpolate the frequency response from the filterbank's outputs. For an overview of the filterbanks, each filter is characterized by the lowest and highest frequency represented by f_{low}, f_{high} in Hz where the center frequency $f_{center} = (f_{low} + f_{high})/2$. The f_{low} and f_{center} of the next filter are the f_{center} and f_{high} of the previous filter and so on. Hence, most frequency response corresponding DFT bin center frequency (f_{bin}) are affected by two connected filters. We interpolate the frequency response with the weight functions. Let w_1 and w_2 be the weight functions of the first and second filter which are defined as

$w_1 = [f_{bin} - f_{center}(m)/f_{center}(m+1) - f_{center}(m)]$ and $w_2 = 1 - w_1$. In case the DFT bin falls into one filter, one of the weight functions is set to zero and the other is set to one. More precisely, $w_1 = 1$, $w_2 = 0$ if $f_{bin} \in [f_{low}(1) \quad f_{center}(1)]$ and $w_1 = 0$, $w_2 = 1$ if $f_{bin} \in [f_{center}(M) \quad f_{high}(M)]$. The interpolated DFT ($F_{interp}$) can be expressed as

$$F_{interp} = w_1 S(m,l) + w_2 S(m+1,l) \quad , m = 1, \ldots, M \tag{3.16}$$

Therefore, the desired signal is the Inverse Fourier Transform of F_{interp}.

3.5 Short-term Energy

Energy of a signal expresses the strength of the signal. It is usually applied for voice activity detection. The energy of a voice is higher than the energy of the noise. Hence, it can use to segment the signal into speech and silence region. The short-term energy E is represented by the following equation.

$$E = \sum_{n=1}^{N} s^2[n] \tag{3.17}$$

Another measure that provides equivalent information is the short-term power $\mathcal{P}$ represented in the following equation.

$$\mathcal{P} = \frac{1}{N} \sum_{n=1}^{N} s^2[n] \tag{3.18}$$

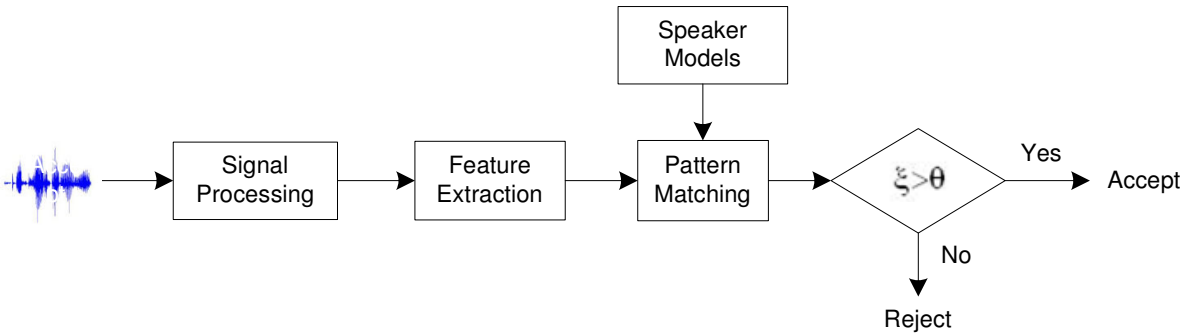


Figure 3.2: Block diagram of a speaker verification system.

3.6 Speaker Verification Models

The speaker verification model is represented in Figure 3.2. This figure shows five components of the speaker verification system. We have described the signal processing and feature extraction components. Now, we are going to explain the rest.

The following techniques based on a pattern matching method [15] are popularly used in the speaker verification system. All techniques are used to create a speaker model. The speaker model is used to examine the similarity between the model and a test utterance (pattern matching). Let ξ be the similarity between the speaker model and a test utterance. The decision is given as the following equation where θ is a decision threshold which is determined so that it minimizes an error rate.

$$\xi \begin{cases} > \theta & \text{Accept} \\ \leq \theta & \text{Reject} \end{cases} \tag{3.19}$$

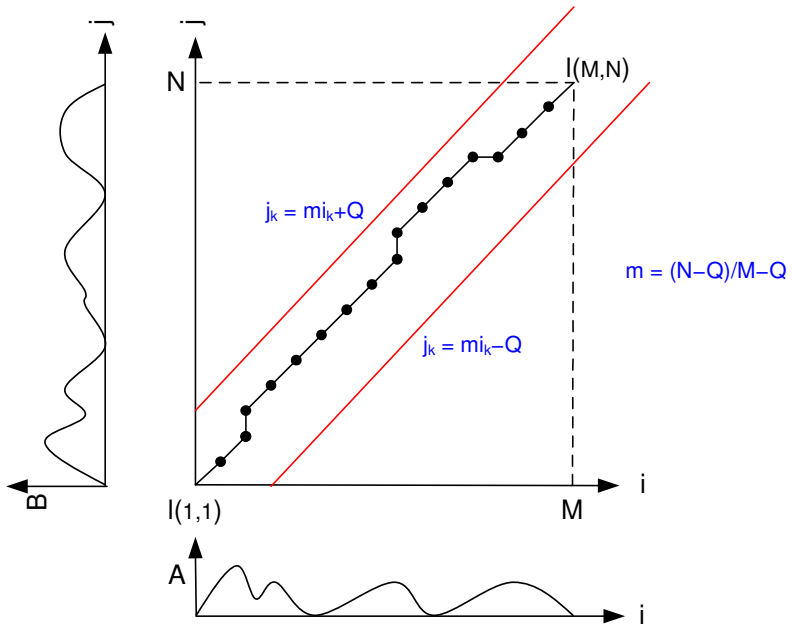


Figure 3.3: An example of Dynamic Time Warping where the mapping between two signals (A and B) is given by the dot line.

Dynamic Time Warping (DTW)

For DTW, given two time sequences of feature vectors, we have to find a warping function which minimizes the distance between the two feature vectors. Let the two sequences of feature vectors which should be compared be

$$A = a_1, a_2, \ldots, a_i, \ldots, a_M \quad and \quad B = b_1, b_2, \ldots, b_j, \ldots, b_N$$

The warping function can be represented by a sequence of lattice points on the plane (see Figure 3.3), $l = (i, j)$, as indicated in the following equation.

$$L = l_1, l_2, \ldots, l_k, \ldots, l_K, \quad l_k = (i_k, j_k)$$

Let $d(l_k)$ be a cost function which is defined as the distance between a_{i_k} and b_{j_k}. The overall cost function, $D(L)$, can be determined by the following equation [72].

$$D(L) = \sum_{k=1}^{K} d(l_k) \tag{3.20}$$

In addition, a warping function is required to minimize the overall cost function under the following constraints [72, 31].

1. The function must be monotonic:

$$i_k \geq i_{k-1} \quad and \quad j_k \geq j_{k-1}$$

2. The function must match the endpoints of A and B (Boundary condition).

$$i_1 = j_1 = 1, \quad i_K = M, \quad and \quad j_K = N$$

3. The function must be a continuity function.

$$i_k - i_{k-1} \leq 1 \quad and \quad j_k - j_{k-1} \leq 1$$

4. The function must be a global limit to prevent extreme expansion and contraction.

$$|i_k - j_k| < Q, \quad Q = \text{constant}$$

Hence, we have to find the minimum-cost path from point(1,1) to point (M, N). The minimum-cost warping path can be efficiently determined by using Dynamic Programming (DP). With the constraints described above, the cumulative distance $D(i, j)$ over a partial sequence of $l_1, l_2, \ldots, l_k, \ldots, l_K$ where $l_k = (i, j)$ can be expressed as the following equation.

$$D(i, j) = d(i, j) + min(D(i-1, j-1), D(i-1, j), D(i, j-1)) \tag{3.21}$$

An example of DTW is shown in Figure 3.3. The signals to be mapped are shown along the axes. The diagonal dot line shows the mapping between A and B where the i^{th} sample of A is aligned with the j^{th} sample of B.

Vector Quantization (VQ)

For VQ, the acoustic models of speakers are created by partitioning a collection of acoustic feature vectors to C clusters [83]. Each cluster is represented by a mean vector or centroid denoted by c_i for $i = 1, \ldots, C$. In literature, a set of controid $\mathcal{C} = \{c_1, \ldots, c_C\}$ are referred to as a codebook. An example of VQ-based speaker model is shown in Figure 3.4 with $C = 10$.

For verification, given an input vector $X = \{x_1, \ldots, x_m\}$, the quantization distortion for speaker j can be calculated by summing the nearest distance in

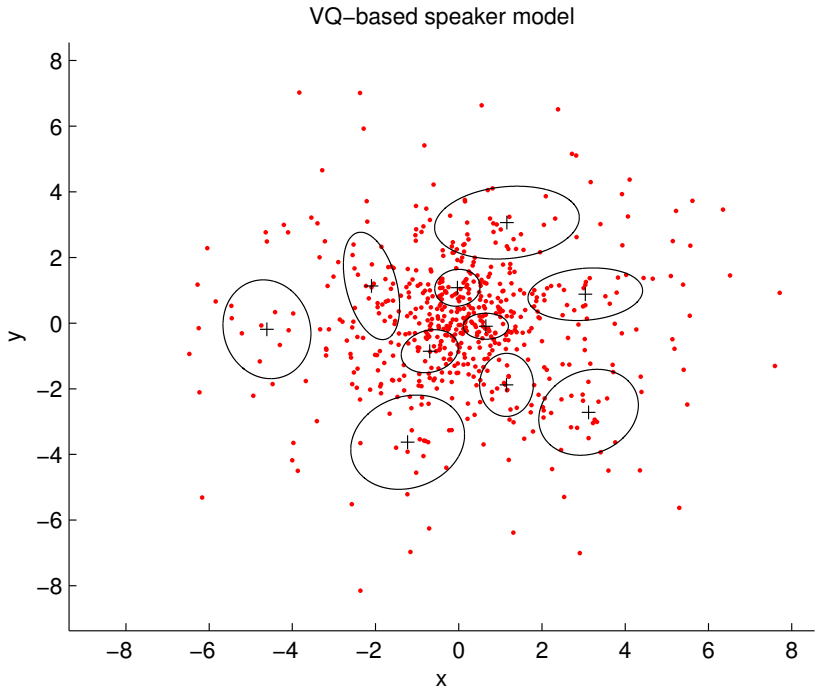


Figure 3.4: An example of the VQ-based speaker model with 10 clusters.

the codebook ($\mathcal{C}_j$). More precisely, the distortion d of the vector x_k from $\mathcal{C}_j$, $d(x_k, \mathcal{C}_j)$, is given by equation 3.22 where $d(x_k,c_i)$ is a distance between x_k and c_i.

$$d(x_k, \mathcal{C}_j) = \arg\ min_{c_i \epsilon \mathcal{C}_j}\ d(x_k, c_i) \tag{3.22}$$

Hence, the distortion of X from $\mathcal{C}_j$ is determined by the following equation.

$$D(X, \mathcal{C}_j) = \frac{1}{m} \sum_{k=1}^{m} d(x_k, \mathcal{C}_j) \tag{3.23}$$

Gaussian Mixture Models (GMM)

The GMM model consists of a finite number of Gaussian mixtures. Each mixture is parameterized by a priori probability π, mean vector μ, and covariance matrix Σ [76]. The GMM model can be represented by $\lambda = \{\lambda_1, \ldots, \lambda_K\}$ where $\lambda_k = (\pi_k, \mu_k, \Sigma_k)$. These parameters can be estimated by using the Expectation-Maximization (EM) algorithm [26]. An example of GMM-based speaker model is shown in Figure 3.5 with $K = 10$.

Given an input vector $X = \{x_1,\ \ldots,\ x_m\}$, the matching score for GMM is determined by the log-likelihood of the GMM, L_{GMM}, in the following equation where $\lambda_j = (\pi_j,\ \mu_j,\ \Sigma_j)$ and $\lambda_{j\prime} = (\pi_{j\prime},\ \mu_{j\prime},\ \Sigma_{j\prime})$ are the model of speaker j and the background model of speaker j.

$$L_{GMM} = log\ p(X|\lambda_j) - log\ p(X|\lambda_{j\prime}) \tag{3.24}$$

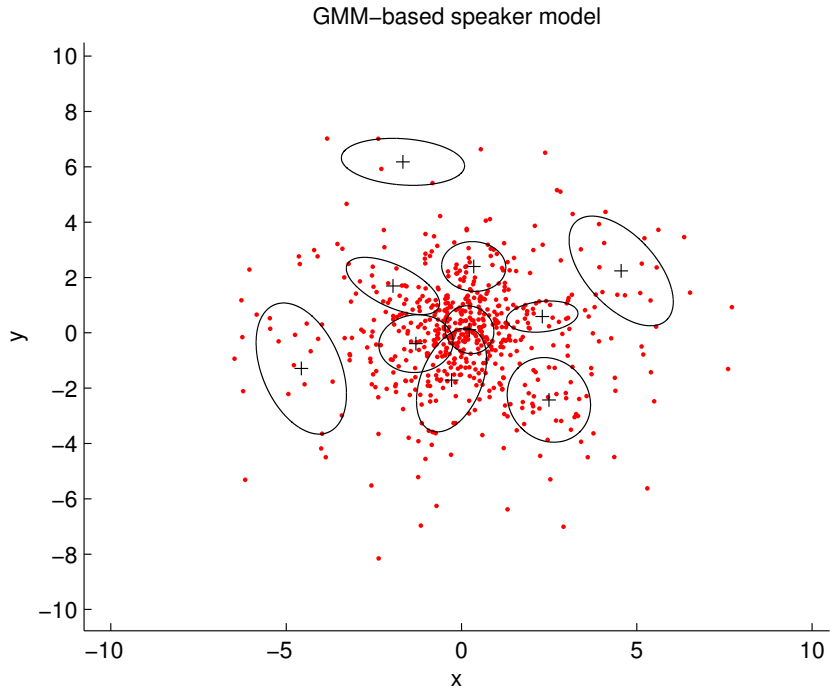


Figure 3.5: An example of the GMM-based speaker model with 10 Gaussian Mixtures.

Chapter 4

Attack Against Speaker Verification

4.1 Introduction

In this chapter, we investigate the security of user authentication based on adversary knowledge of private and public information and the motivation of the adversary as introduced in Chapter 1. Three systems based on a pattern matching technique are used in the study. According to [15], the pattern matching methods include a template, a codebook, and a statistical model. Dynamic Time Warping (DTW) is used in the template model, Vector Quantization (VQ) in the codebook model, and Hidden-Markov Model (HMM) in the statistical model. For statistical models, we used a single state HMM which is referred to as a Gaussian Mixture Model (GMM) [15].

We present attack models based on adversary knowledge. We start with naive

adversaries without knowledge of an authentic speaker and develop them into highly knowledgeable adversaries who know the speaker's information, have the speaker's voice samples, acquire the speaker's template, and know an algorithm of the speaker verification system. We propose an analysis-synthesis forgery in which the highly informed adversary can exploit information, such as feature vectors from the template and a statistical probability from the voice samples of the target speakers to regenerate a forgery that can be used in remote or on-line authentication.

We attack the systems using human and algorithmic attacks. For the first scenario (human), we ask a subject to say the pass-phrases of the target users for multiple rounds. For the first round, the impostors say the pass-phrases without listening to the target voice. In the second round, they are asked to imitate the pass-phrases of the target users by listening to the voice of the target users. In this round, we ensure that the subjects are well-motivated by providing an incentive reward for the best imitator. For the second scenario (algorithmic), we will use voice recordings from the target users to generate synthesized pass-phrases. The synthesized sound will be generated from state-of-the-art technologies; we use an HMM-based speech synthesizer. We carefully designed the collection of the voice data, so the voice would not overlap with the pass-phrases of the target users. In the last scenario (algorithmic), we regenerate users' pass-phrases based on the template information. Then, these pass-phrases will be used to attack the systems. These scenarios are detailed in Section 4.4.

4.2 Datasets

We evaluate the recognition performance using an Equal Error Rate (EER) which is the rate at which the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) are equal. The FAR is the percentage of the time that the system accepts the wrong speaker or one who is not authorized to access the system. In the same way, the FRR is the percentage of the time that the system rejects the authorized speaker. Two datasets are used in our experiments: The MIT mobile device speaker verification corpus dataset (MDS) [97] and The Lehigh quiet environment speaker verification dataset (LDS). The MDS is a public dataset available from MIT. The LDS is our own dataset collected over a one month period of time.

4.2.1 The MIT Mobile Device Speaker Verification Corpus

This dataset was collected from 48 speakers (22 females and 26 males). The utterances were recorded in three acoustic environments: office, lobby, and street intersection via two types of microphones: external earpiece headset and built-in mobile device. The dataset consists of two sets: a set of enrolled users and a set of dedicated imposters. For the enrolled set, speech data was collected over two sessions on separate days (20 minutes for each session). For the imposter set, users participated in a single 20 minutes session. There are six lists of pass-phrases that were varied by three environments and two types of microphones. We select the first list for our experiment because it provided pass-phrases that were said by the same speaker multiple times under the same environment (office). So, we can use this list in the training and the testing phase.

During each data collection session, the user recited a list of ice cream flavor phrases which were displayed on the hand-held device. The sets of enrolled users' pass-phrases and dedicated imposters' pass-phrases used in our experiments are provided in Appendix A.1 and A.2.

4.2.2 The Lehigh Quiet Environment Speaker Verification Dataset

This dataset contains 4,320 recordings collected on a laptop computer via an external earpiece headset microphone from six male speakers during several rounds. The data collection was taken in the graduate study room at Lehigh University's Library that can be referred to as a quiet environment.

In the first round, the subjects were asked to say their five pass-phrases which were chosen from idioms, famous phrases and everyday conversations (see Appendix A.3). Each pass-phrase was uttered 10 times. In addition, they were asked to say 270 short sentences (see Appendix A.4) to make a speech corpus. The set of short sentences in the speech corpus were chosen from "The CMU arctic speech databases" designed by Language Technologies Institute, Carnegie Mellon University, USA [54]. The databases consist of approximately 1200 phonetically balanced English utterances. We note that the 270 selected sentences are not overlapped with any user's pass-phrases. However, the selected sentences to cover all pass-phrases' diphones. Thus, we can synthesize reasonable quality sound to attack the systems.

Approximately two weeks later, in the second round, they were asked to say their same set of pass-phrases. Each was uttered five times. Furthermore, they were asked

to say other subjects' pass-phrases. Each was uttered five times. Lastly, they were asked to imitate the other subjects' pass-phrases by listening to the pass-phrases that we replayed to them. Each pass-phrase was uttered five times.

The third round began at the end of the fourth week. By listening to imitated pass-phrases, we selected the best imitator, who was then asked to mimic the target speaker's pass-phrases. Each pass-phrase was uttered five times.

4.3 Speaker Verification Models

For all constructions in the following subsections, we use a low-pass digital filter with a cut-off at 4 kHz to strip the higher frequencies from the signal. The next step is pre-emphasis. We set the pre-emphasis parameter γ to 0.98. Then the signal is framed into the short time analysis interval and multiplied with the Hamming window defined as follow.

$$w(n) = \begin{cases} 0.54 - 0.46cos\frac{2n\pi}{N} & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

We set the length of each frame to 30 msecs with 10 msecs overlap. For the sampling rate of 8 kHz, we use 240 samples per frame that are shifted every 80 samples.

A decision threshold is estimated based on the distribution of overall distances between each authentic speaker's and a set of imposters' features. For our setting, let mean and standard deviation of the inter-speaker score be μ and σ, we set the decision

threshold θ by the followings equation where c is some constant that minimize the error rate.

$$\theta = \mu - c\sigma \tag{4.2}$$

4.3.1 Dynamic Time Warping (DTW)

For DTW, we use the first utterance as the keying signal and perform DTW to the rest. The averaged result is stored as the matching template. The distance between an input and the template is determined by using the Euclidean distance. The system decides whether to accept or reject the speaker by comparing the Euclidean distance to the decision threshold.

4.3.2 Vector Quantization (VQ)

For our setting, K-means clustering is used to quantize the training vectors. We investigate the performance of VQ in our datasets by setting the number of clusters to 10, 20, 30, 40, and 50. The performance with 30 clusters yields the best results. Therefore, we set $C = 30$. The distance between an input vector and the nearest centroid is determined by using the Euclidean distance. The system decides whether to accept or reject the speaker by comparing the distance to the decision threshold.

4.3.3 Gaussian Mixture Models (GMM)

The GMM model consists of a finite number of Gaussian distributions parameterized by their priori probability π_j, mean vectors μ_j, and covariance matrices Σ_j [76]. In this experiment, we use nodal covariance matrices. We initialize the speaker models using the K-means clustering, then the parameters are estimated by using the EM algorithm [26].

The training utterances of all speakers except speaker j are used to create the background model and the rest is used to create the speaker model of speaker j. We use the GMM mixture order = 10 for the reason similar to the setting of the VQ. The system decides whether to accept or reject the speaker by comparing the log-likelihood to the decision threshold.

4.4 Attack Models

We investigate two types of attack: human and algorithmic. We vary the adversary knowledge by making three different assumptions in the human case and two different assumptions in the algorithmic case, for a total of five classes of attacks.

4.4.1 The Human Type with Assumption I (H-I)

We assume that the attackers do not know the authentic speakers and their pass-phrases. We evaluate the error rate of the authentic speakers compared with the adversary (*naive*) who say the random pass-phrase with different phonetic content than the actual pass-phrase.

4.4.2 The Human Type with Assumption II (H-II)

We assume that the attackers know the pass-phrase and say the actual pass-phrase. In this experiment, the adversaries (*imposter*) say the actual pass-phrase without listening to the target pass-phrase. All other subjects except the authentic speaker will be the adversaries.

4.4.3 The Human Type with Assumption III (H-III)

We assume that the attackers know the pass-phrase and are acquainted with the authentic speaker. Then, they try to mimic the target pass-phrase. In our experiment, we re-play the pass-phrases of target speakers to the adversary (*informed imposter*) and then the informed imposter repeats the pass-phrases. Note that we use the term "informed" instead of "skilled" because the attackers have only been given useful information for creating a forgery. Taken literally, "skill" means that someone has demonstrated a proven talent; we have done nothing to prove that the test subjects actually have a real talent for forgery.

4.4.4 The Algorithmic Type with Assumption I (A-I)

We assume that the attackers know the pass-phrase and have acquired sufficient voice samples of the target user to build a speech synthesizer tuned to the user's voice. Then, they synthesize the pass-phrase. In this type, we use HMM-based speech synthesizer to create *synthetic pass-phrases* [92]. We use 270 phrases in the first round in the LDS dataset for training the speaker-dependent models to synthesize pass-phrases. The phonemes in each phrase are labeled to form HMM models of each

phoneme in the training phrase. The HMM models are parameterized by spectrum (MFCC) and excitation (fundamental frequency, and duration) parameters. To synthesize the sound, the pass-phrase to be synthesized is analyzed then the phoneme HMM models are concatenated based on unit clustering. Finally, the concatenated HMM models output the parameters to synthesize the sound by passing these parameters to the synthesis filter. For each speaker, five pass-phrases are synthesized corresponding to their pass-phrases. For all processes in synthesizing the sound, we set the sampling rate at 8000 Hz.

4.4.5 The Algorithmic Type with Assumption II (A-II)

We assume that the attackers know the pass-phrase and have acquired the template of the target user. Moreover, they know the system's construction and use this information to create *regenerated pass-phrases*.

We refer the algorithmic type attacker (A-I and A-II) to as an *informed adversary*.

Attack against DTW template.

We store 13 order MFCCs of the first utterance as the reference template. Hence, we have to transform this template to a signal. We first transform MFCCs to DFTs using Auditory Toolbox [81]. Then the DFTs are transformed to the speech signal used as the forgery.

Attack against VQ and GMM template.

For VQ and GMM, we use 13 order MFCCs for training and verification. The authentication system consists of two units as indicated in Figure 4.1. The

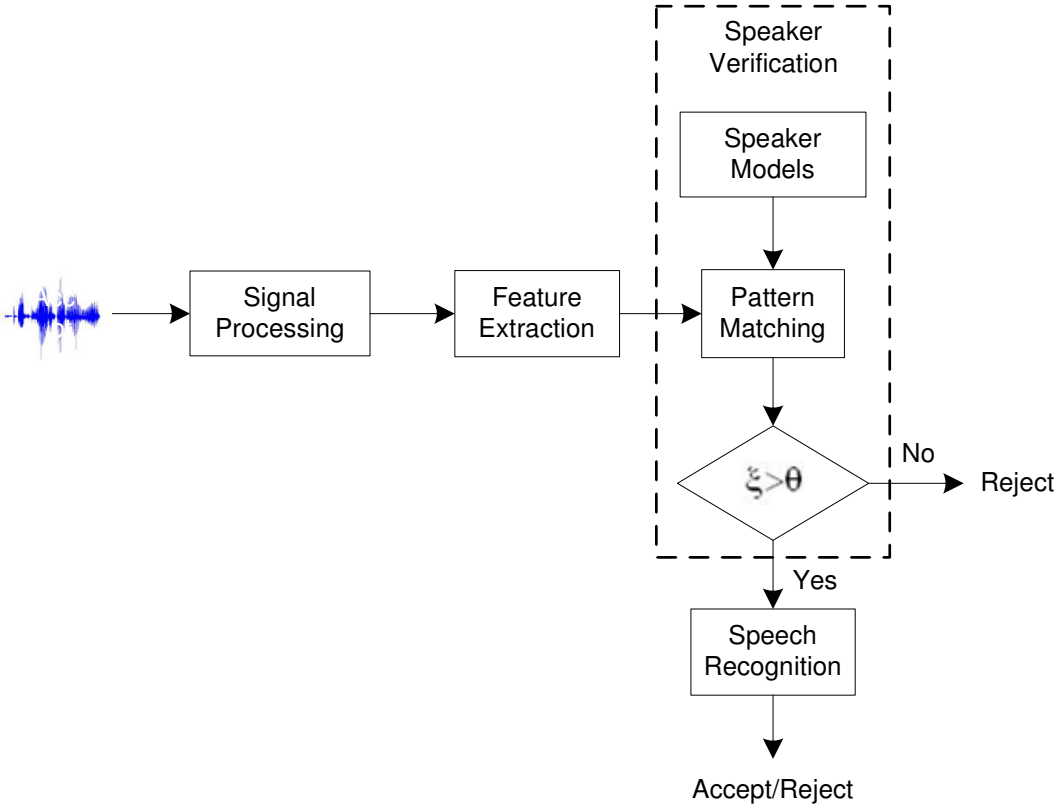


Figure 4.1: Block diagram of the VQ and GMM speech biometric user authentication.

first is speaker verification which aims to determine whether the person is who he/she claims to be. The second is speech recognition which is used to check whether user utter the registered pass-phrases. For the speech recognition unit, the implementation is based on DTW. A set of pass-phrases from the speakers is used to create templates of M classes. Each class consists of R reference templates. An input vector will be aligned to the same range of the set of reference templates in each class. Hence, we employ k-nearest neighbor for classification [29]. Next, the unknown spoken input will be classified into one of M classes. For our datasets, M is 10 for the MDS and 30 for the LDS. The accuracy of the recognition unit to recognize pass-phrases is 90.58% and 94.64% on average.

The VQ template consists of a codebook $\mathcal{C}$ and a decision threshold θ, for each speaker. These parameters will be used to calculate the distortion of a set of input vectors X (a set of range m or number of frame of speech). The system will accept the speaker, if the distortion of the X is lower than the threshold. Hence, we will search a set of vector $x_i \in X$ that yield a distortion close to the decision threshold.

Let the distortion of x_i be $d(x_i)$. We want to select a set of vectors $v = \{x_i | d(x_i) < T_a, i = 1, 2, \ldots, n\}$ where $n \leq m$ and T_a is the appropriate threshold to re-synthesize the pass-phrase. Basically, the verification performance will be degraded if T_a is high. On the other hand, by setting T_a too low, it will degrade the speech recognition performance. Hence, the appropriate threshold will be determined by experimentation. More precisely, we will set $T_a = \theta + \kappa\theta$ where $\kappa \in [-a, a]$ is *a tuning parameter*. Then, we select κ which yields the best result.

A possible problem is the case of a null set of v. In this case, we use the source's vectors (X).

For the GMM attack, the priori probability π_j, mean vectors μ_j, and covariance matrices Σ_j are used to calculate the log-likelihood of the input vectors. We will select a set of vectors based on the decision threshold of log-likelihood the same way as we did for VQ attack.

4.5 Experiments and Results

4.5.1 Experimental Setup

For the LDS, we use five pass-phrases from each speaker in our experiment, a total of 5*6 = 30 different pass-phrases. Six recordings from the first round are used to train the system. Five recordings from the second round are used for verification. We randomly select 25 other pass-phrases from other speakers that do not correspond to the verification pass-phrase to evaluate H-I's trial. Five recordings of the same pass-phrase uttered by other speakers in the second round are used to evaluate H-II's trial, in total of 5*5 = 25 recordings for each pass-phrase. For H-III's trial, five mimicked recordings are used. The synthesized pass-phrase is used for A-I's trial. For A-II's trial, we use five recordings from H-II's trial as the sources of acoustic features and change them to target pass-phrases.

For the MDS, we use six recordings to train the systems. Two recordings are used for verification. As this dataset includes gender information, we conduct two experiments based on gender: Same and Mixed. For the same-gender experiment,

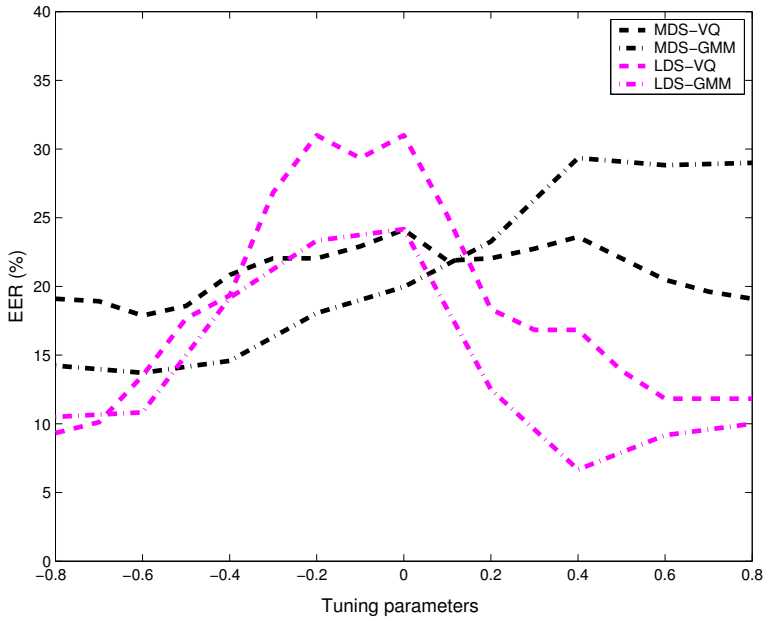


Figure 4.2: The error rates of regenerated pass-phrases by varying κ in VQ and GMM system.

the number of H-II's pass-phrases that are available in the dataset varies from 1 to 6. For the mixed-gender experiment, it varies from 5 to 11. For both experiments, these pass-phrases are used as sources of acoustic features for the A-II. For H-I's trial, we use six pass-phrases from other speakers that are different from the verification pass-phrase (based on gender). For the other classes, we do not have enough voice samples to synthesize reasonably high quality sound and we do not have mimicked utterances. Hence, we do not investigate the H-III and A-I.

4.5.2 Experimental Results

Same-gender experiment

The results in Figure 4.2 illustrate the error rates of regenerated pass-phrases by varying $\kappa \in [-0.8, 0.8]$. Maximum points of each plot optimize the tradeoff between the recognition and verification performance of the systems. Thus, we set κ to the maximum point for each system.

Figure 4.3 depicts the graphical results of EERs against the various attacks for the LDS and MDS. For each attack, we repeat the experiment 30 times. Each time, we randomly select an adversary pass-phrase from a set of dedicated imposters and assign it to each user. In general, if a number of samples is greater than or equal to 30, the sample variance (s^2) will be close to the population variance (σ^2) [68]. Therefore, we can determine the confidence interval on the mean (μ) by the following equation where Z is the standard normal distribution, s the standard deviation of the sample, x is a set of samples, n_s is sample size, μ is the mean of the population, and α is the confidence coefficient.

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n_s}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n_s}} \tag{4.3}$$

These results are shown in Figure 4.3 where we set α to 0.05; the results are the 95% confidence interval. It is clear that the informed adversary utilizing the template information (A-II) is the most successful adversary in gaining access to all systems. In particular, for the DTW system, the error rate is the highest. The results of the A-II from the MDS and LDS seem to conflict (Figure 4.3).

The A-II algorithm for the GMM did better in the MDS, but in the LDS the result is reversed. This may be possible for two reasons. First, we vary the tuning parameters coarsely. Thus, the better value of κ may be missed. The other reason is that we just utilize imposters' pass-phrases (H-II) in the case of a null set of s. Hence, these pass-phrases may affect the results. For the other attack models, the EERs of the DTW are the lowest. In particular, for the H-III and A-I in the LDS, the EERs of the DTW are noticeably lower than the EERs of the VQ and GMM. These results suggest that the DTW will yield a good performance if the template is protected properly. Thus, the template protection is the critical issue for the DTW approach.

Assuming that the practitioners do not take the informed imposter and adversary (H-III, A-I, and A-II) into account, a decision threshold may be determined to be at an operating point of the H-II. We further assume that the systems do not check whether the pass-phrase is correct because for text-dependent speaker verification systems if the pass-phrase is incorrect, the matching score will be greater than the threshold and eventually be rejected. The results are summarized in Table 4.1 which illustrates the error rates (FAR) of various attacks. The figures of the H-I and H-II in the table reflect the standard (traditional) evaluation of biometric authentication systems. Beyond the standard evaluation, the FARs of other attack models are very high. In particular, the FARs of the A-II are the highest.

Table 4.1: FARs (%) of speaker verification systems (DTW, VQ, and GMM) against various attacks using decision thresholds at operating points of imposters (H-II).

Datasets	Attack models	DTW	VQ	GMM
LDS	H-I	0.27	3.53	2.22
	H-II	7.20	11.56	8.89
	H-III	8.67	25.47	24.05
	A-I	20.00	26.67	60.00
	A-II	**90.00**	**55.00**	**65.00**
MDS	H-I	0.00	4.08	2.08
	H-II	11.86	16.40	13.12
	A-II	**100.00**	**47.22**	**89.93**

Mixed-gender experiment

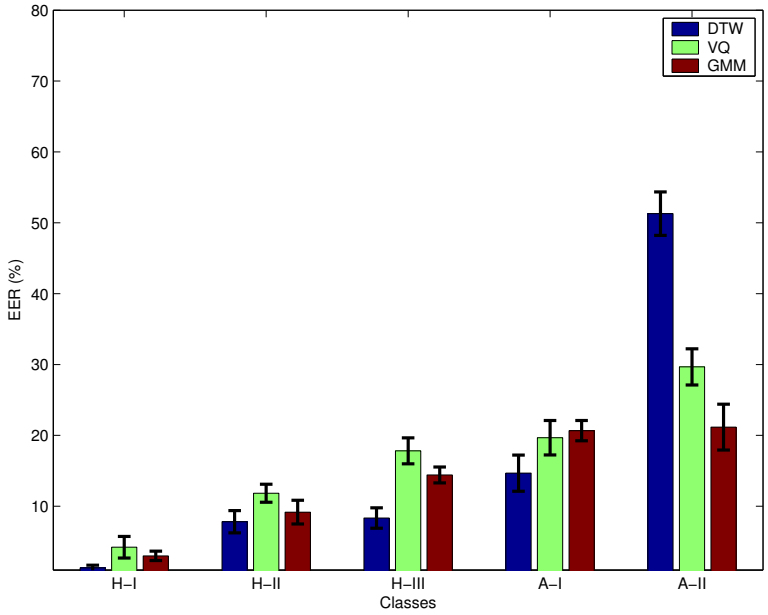
The same methodology which we use for the same-gender experiment is applied for the mixed-gender experiment. Figure 4.4 illustrates comparisons of the H-I and H-II results for the same-gender and mixed-gender experiments for the DTW, VQ and GMM system. The EERs from the two experiments are not significantly different for the VQ and GMM but, for the DTW, it is noticeably different. One possible reason may be that the length of random pass-phrases for males and females are different. For a set of male pass-phrases, there are many short pass-phrases, for example "rocky road." Therefore, these pass-phrases may affect the EER of the mixed-gender experiment.

For the A-II results (Figure 4.5), we do not show results for the DTW system because the DTW template does not depend on gender. For the VQ and GMM, the results are also not significantly different.

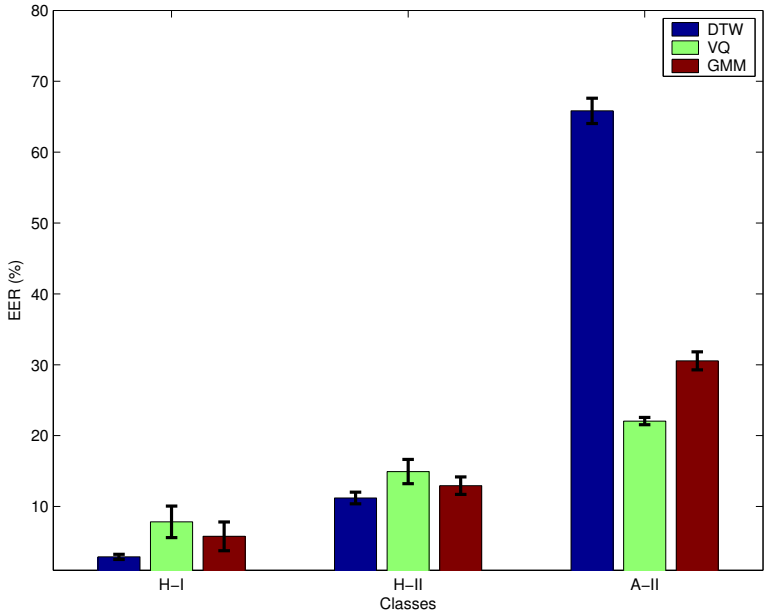
4.6 Summary

In this work, we have shown that the adversary can exploit the DTW, VQ or GMM template and use them to attack the systems. We developed an algorithm to regenerate the pass-phrases that can be used in remote or on-line authentication. We compared our algorithmic attack with the traditional (human imposters) and the more sophisticated attack (an adversary utilizing a synthetic pass-phrase). The EERs of the regenerated pass-phrases were higher than the other attack models. Then, we have demonstrated that the traditional approach to evaluate the security of speech biometric speaker verifications was insufficient. The results indicated that the FARs of other attack models beyond the traditional approach were very high. We also investigated the results based on gender information. There were no significant differences.

We hope that these results raise important issues for researchers when attempting to demonstrate the security of speech biometric systems. For future work, we are considering ways to address the weaknesses we have identified in this chapter.

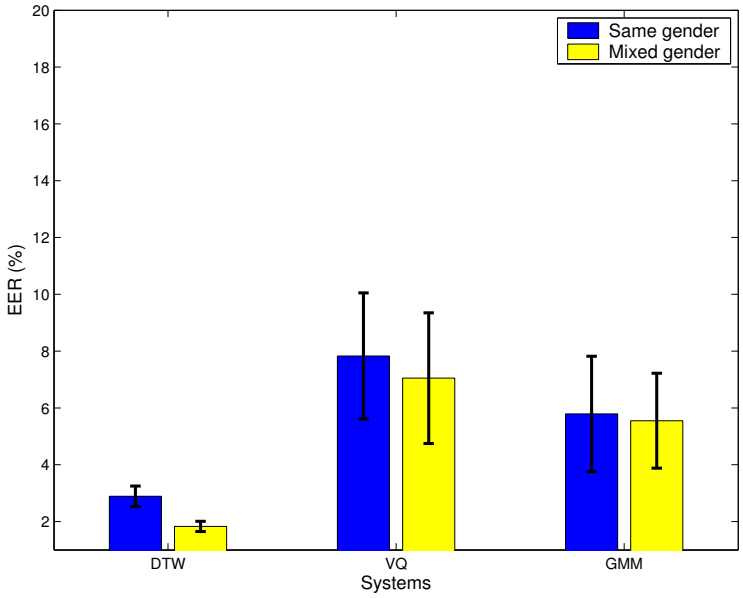


(a)

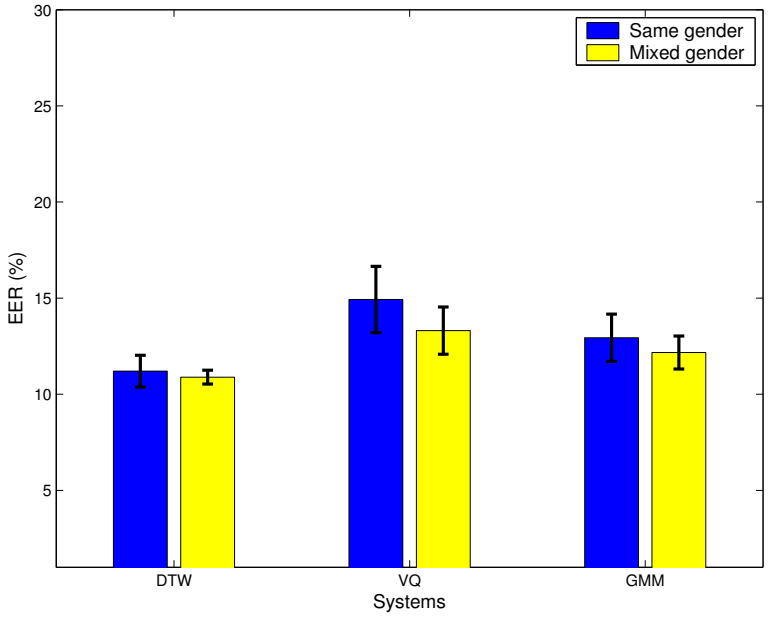


(b)

Figure 4.3: The EERs against various attacks and models with the 95% confidence interval for the same-gender experiment (a) the LDS and (b) the MDS



(a)



(b)

Figure 4.4: Comparisons of the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS (a) the H-I (b) the H-II

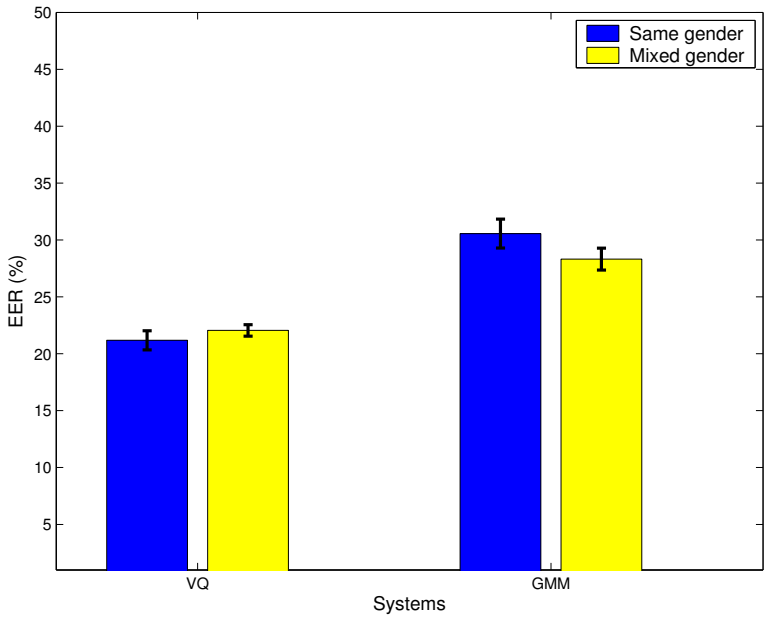


Figure 4.5: Comparisons of the A-II for the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS.

Chapter 5

Dynamic Time Warping-based Biometric Key Binding

The results of previous chapter showed that the security of the speech biometric templates was relatively low. In this chapter, we propose a cryptosystem to address this problem. We present a new scheme to transform speech biometric measurements (feature vector) to a binary string which can be combined with a pseudo-random key for cryptographic purposes. We utilize Dynamic Time Warping (DTW) in our scheme. The challenge of using DTW in a cryptosystem is that a template must be useful to create a warping function, while it must not be usable for an attacker to derive the cryptographic key.

5.1 Introduction

The template protection approaches that are proposed in the literature can be classified into two categories [44] (see Figure 5.1): *feature transformation* and *biometric cryptosystem*. For the first approach, a template is transformed using some transformation functions to form a transformed template. The scheme utilizing a one-way function is called *non-invertible transforms* (e.g. [75, 88, 91]). The main drawback of this scheme is that the performance is degraded since the matching algorithm takes place in a transformed domain. Another scheme for this approach is called *salting*. The salting scheme uses an invertible transformation function that is parameterized by a random key or a password to transform a template (e.g., [6, 19, 80, 89]). This scheme also suffers from transformed features. Furthermore, if a random key or a password is compromised, it can be used to recover the biometric template. For the second approach, the public information that does not significantly reveal the biometric template is stored. This information is referred to as *helper data*. During the matching process, the helper data and the biometrics are used to derive a cryptographic key. The system that directly uses the helper data and the biometrics to generate the cryptographic key is called a *key generation cryptosystem* (e.g., [1, 12, 13, 17, 18, 27, 55, 86, 87, 96, 100]). If the biometrics is used to extract the cryptographic key from the helper data, the system is called a *key binding cryptosystem* (e.g., [11, 20, 21, 23, 28, 36, 47, 48, 51, 64, 93, 94, 98]). The system that uses more than one scheme will be referred to as *hybrid schemes* (e.g., [10, 70, 82, 84]).

Our approach falls under the hybrid schemes. We protect the DTW template using the idea similar to the non-invertible transformation scheme. The Hardening

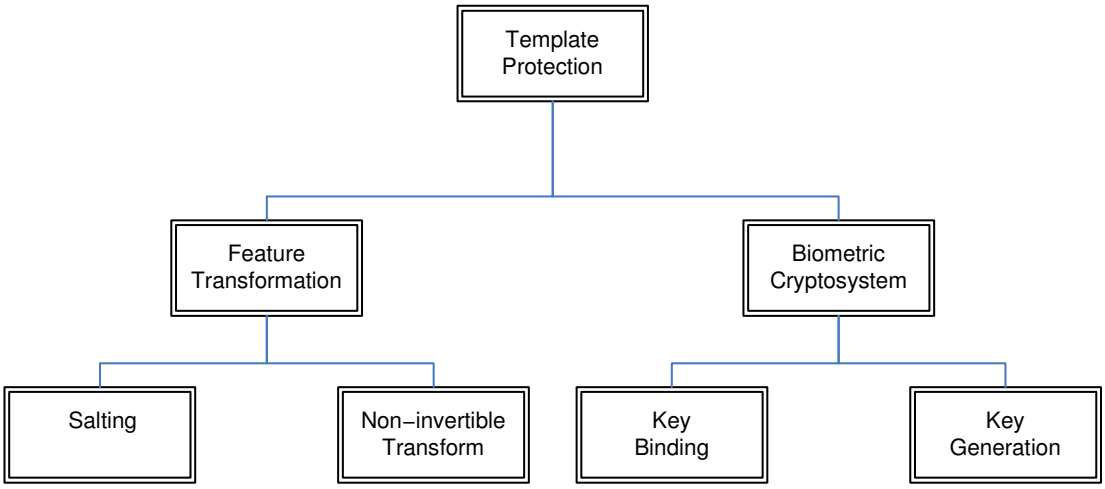


Figure 5.1: Categorization of template protection schemes.

algorithm (Section 5.2.1) is proposed to perturb the original template by removing some frequency-domain features from the template. Finally, the rest of features will be transformed to a time-domain template that we refer to as a *hardened template* (Definition 1). This template will be used as a keying signal in DTW process. The Discrete Fourier Transform DFT and the inverse DFT (IDFT) will be used to create a stored or hardened template.

Definition 1 Given a DFT vector (full template) $X = \{x_i,\ i = 1, \ldots, F\}$, *A hardened template* $\mathcal{H}_T$ is an IDFT of a hardened vector $\mathcal{H} = \{X | \exists x_i = 0\}$ such that the hardened template must be useful to create a warping function, while it must not be usable for an attacker to derive the cryptographic key.

The next step is to regenerate a cryptographic key. The key binding approach is used to protect the key. We refer this template (key binding) to as a lock data $\mathcal{L}$ or *a binary template.*

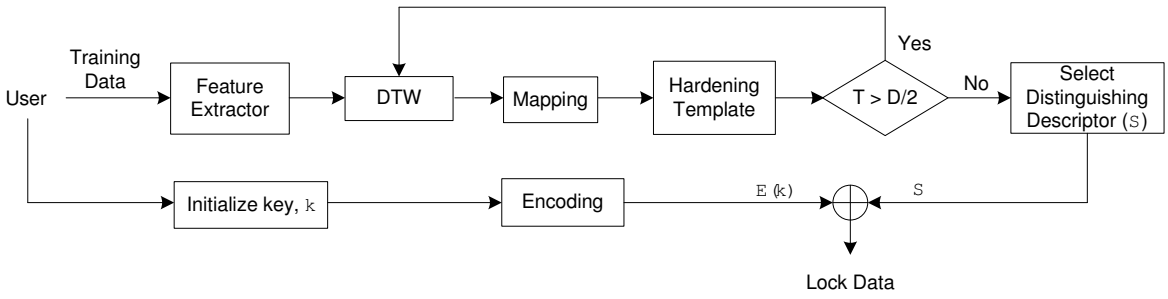


Figure 5.2: Dynamic time warping-based biometric key binding in training phase.

The other problem is the correlation among features. Hao reported that "an iris code usually has a run length of eight consecutive '1's or '0's [34]." For speech (e.g., Monrose et al.'s scheme [66]), we cannot specify the exact length of repetition. It depends on the number of phonemes in a pass-phrase and the idiosyncrasy of each user when he/she utters the pass-phrase. However, consecutive '1's or '0's will lower the randomness of the template. We address this problem by proposing the *Mapping* algorithm (Section 5.2.2) using *a multi-threshold template* $\mathcal{T}$ that are determined from pseudo-random bits (Section 5.2.3). Hence, the algorithm can generate a binary string that an observer cannot predict.

We focus on how to reliably, securely, and randomly (in the context of cryptography) generate a binary string from biometrics. The DTW will make our scheme more reliable while the hardened template maintains security. Finally, a multi-thresholds scheme will help our system generate a binary string unpredictably to maximize the entropy of the template. The following section describes our approach to regenerate a cryptographic key.

5.2 Dynamic Time Warping-based Biometric Key Binding (DBKB)

Our design can be viewed as two phases: training and verification. The biometric key binding is in the training phase indicated in Figure 5.2. Users provide their training pass-phrases that are repeated *l+1* times to the system. Feature extraction is the first process to derive feature vectors and Discrete Fourier Transform (DFT) features. This process involves digital signal processing which we have described in the previous chapter. The system is initialized by using one of the training utterances as the keying signal which is stored as 121 DFT features of m frames. Then the system performs DTW to the rest of training utterances. The feature vectors of each utterance (m frames) will be mapped, a frame per bit, to a binary string of length m called *a set of feature descriptors*. Lastly, l sets of feature descriptors are used to define *distinguishing features*: features of length D that the user can reliably generate. The binary string of distinguishing features derived from the training utterances is called *distinguishing descriptors*. The mapping and defining the distinguishing features procedure are detailed in Section 5.2.2.

We initialized the template by using a full set of DFT features. However, we are not able to use the full template as attackers can utilize it to derive the cryptographic key. Hence, the template has to be perturbed which is what we call *hardening the template* and we refer the result to as a *hardened DTW template*. We set the goal of hardening the template by the following statement: the attacker directly utilizing a hardened template should not be better than *the simplest attack* where the attacker randomly guesses the distinguishing descriptors.

Specifically, let the total number of bit derived from the hardened template that corresponds to the distinguishing descriptors be T; the system should yield T as less than or equal to *D/2*. The motivation is due to the hardening goal. For a simplest attack, any random bits are equally likely to be 1 or 0. Hence, the expected proportion of agreeing bits between the simplest attack and the template (distinguishing descriptors) is 0.5 or $D/2$.

It is worth noting that we assume that a binary string derived from the hardened template does not correlate to the distinguishing descriptor in the training phase. In practical, some information may leak from the hardened template because of correlation of features. For this issue, in Section 5.3.3, we will investigate the information leakage and show security analysis where the attackers have perfect knowledge of the correlation of the features.

If this condition hold, $T \leq D/2$, the template will not help the attackers as they just using a simplest attack is easier (better). For this reason, if T is greater than *D/2*, the template will be hardened (see Section 5.2.1). After each step in hardening the template, the new hardened DTW template will be the keying signal of the training pass-phrases and the process will be re-started until the condition is met. Finally, the IDFT of the latest hardened DTW template is stored as a hardened template $\mathcal{H}_T$ and 2^n-1 distinguishing descriptors, where $n = 3, 4, \ldots$, will be selected based on feature variation to form a binary string S.

Once the hardened template is set, a pseudo-random key k is generated and then encoded properly denoted by *E(k)*. In our case, we use Bose and Ray-Chaudhuri (BCH) code [56] . The encoding code *E(k)* has to tolerate error within Hamming distance (H), a maximum number of bit differences between the distinguishing de-

scriptors and the feature descriptors of a legitimate user. For the next step, the S and the encoding code *E(k)* will be hidden using an XOR operation and then stored as a lock data denoted by $\mathcal{L}$. Only the user with feature descriptors S' that is sufficiently similar to the S within Hamming distance ($|S - S'| \leq H$) can unlock the $\mathcal{L}$ and correctly decode the key. We refer to the fuzzy commitment scheme [48] for more detail.

5.2.1 Hardening Template

As described earlier, the DFT features should be used to create a template to be a keying signal. The template is m frames of 121 DFT features each. We need to store a hardened template in order to set the time alignment to the input signal using DTW technique. This template should not be used to derive the key. The straightforward way is to enumerate over m frames of the original template then choose a set of optimal features that yield $T \leq D/2$, but the computational time is not possible. Hence, the optimal search algorithm should be employed. We choose a Sequential Backward Search (SBS) that is a top down search procedure starting from the full set of features and remove one feature per step until the condition is met [71]. By using SBS, it is easy to terminate the program under the assumption we described earlier.

To start, a user presents $l + 1$ training pass-phrases to the system. Then, the sets of DFT features, $\beta_1, \ldots, \beta_{l+1}$, are extracted from the pass-phrases and these sets are used as the inputs of the *Hardening* algorithm (see Algorithm 1). Next, the threshold is initialized with Ω, the mean of the linear combination of all components in the DFT features (vectors) of the biometric samples. The β_1 is used as the initialized

Algorithm 1 Specification of the Hardening algorithm

Input: The biometric samples $\beta_1, \ldots, \beta_{l+1}$
Output: The lock data $\mathcal{L}$, multi-thresholds $\mathcal{T}$, hardened template $\mathcal{H}_T$, selected relevant indexes Ψ
Initialize: $\mathcal{T} \leftarrow \Omega$, $\psi = \{1, \ldots, 121\}, \zeta \leftarrow 121, [D, T] \leftarrow m$, $\mathcal{H} \leftarrow \beta_1$

Hardening$(\psi, \mathcal{H})$
  for $j \leftarrow 1$ *to* ζ
    $\mathcal{H}' \leftarrow \mathcal{H}$
    $\mathcal{H}'(\psi(j)) \leftarrow 0$
    $[T', D'] \leftarrow$ **Mapping**$(\mathcal{T}, \mathcal{H}')$
    if $T' < T$
      $T \leftarrow T'$, $D \leftarrow D'$, $index \leftarrow j$
  $\mathcal{H}(\psi(index)) \leftarrow 0$, **Remove**$(\psi(index))$, $\zeta \leftarrow \zeta - 1$
  if $T > D/2$ **and** $\zeta > 1$
    Hardening$(\psi, \mathcal{H})$
  return $\mathcal{H}$
$\mathcal{T} \leftarrow$ **MultiThreshold**(μ, σ, κ)
$\mathcal{H}_T \leftarrow$ **IDFT**$(\mathcal{H})$
$[B, indexes] \leftarrow$ **Mapping**$(\mathcal{T}, \mathcal{H}_T)$
$\Psi \leftarrow \{\Psi(1), \ldots, \Psi(2^n - 1)\}$ **such that** $\sigma(\Psi(i)) < \sigma(\Psi(i+1))$ **and** $\Psi \subset indexes$
$S \leftarrow B(\Psi)$
$\mathcal{L} \leftarrow E(k) \oplus S$
Delete$(\{\beta_1, \ldots, \beta_{l+1}\}, p, \mu, \sigma, \kappa,\ B,\ S,\ indexes)$

hardened DTW template ($\mathcal{H}$) and the ψ is a list of DFT indexes. For the hardening process, the algorithm will search for a DFT feature in $\mathcal{H}$ that yields the least T when that DFT feature is substituted with 0. Then the index is removed from the ψ (see Algorithm 1, lines 2-8). The hardening process includes the *Mapping* algorithm that will return T and D (more detail will be explained in Section 5.2.2). Thus, the system can check whether T is greater than $D/2$. The above described steps are iterated until T less than or equal to *D/2* (see lines 9-11).

When the recursion is terminated, the algorithm will generate multi-thresholds $\mathcal{T}$ (see Section 5.2.3). Next, the IDFT of $\mathcal{H}$ is stored as the hardened template ($\mathcal{H}_T$ in line 13). The algorithm then inputs the $\mathcal{H}_T$ and the multi-thresholds $\mathcal{T}$ into the *Mapping* algorithm. Consequently, it yields the distinguishing descriptors and their relevant indexes (B and *indexes* in line 14). The last step, we select 2^n-1 the least variation of the distinguishing features (selected relevant indexes Ψ), where $n = 3, 4, \ldots$, to form a binary string S and a lock data $\mathcal{L}$ (see lines 15-17). Finally, the system securely deletes a set the training parameters using a *Delete* function and stores $\mathcal{L}$, $\mathcal{T}$, $\mathcal{H}_T$, and Ψ in the database.

5.2.2 Mapping the Biometric to a Binary String

Algorithm 2 is used to map feature vectors to a binary string. First, the algorithm performs DTW between $\mathcal{H}$ and β_k, $k = 1, \ldots, l+1$. The results are represented with f_k. For each frame of f_k, let $f_k(i)$ represents a feature vector, where i = *1, ..., m*, is the number of the frame. We compute $f'_k(i)$ from the linear combination of all components in $f_k(i)$ and then set a biometric feature $\phi_k(i) = f'_k(i) - \mathcal{T}(i)$ where $\mathcal{T}$ is

a set of thresholds (see lines 2-5). Binarization is the next step. The $\phi_k(i)$ is mapped to a feature descriptor, $b_k(i)$, by testing whether $\phi_k(i)$ is positive or negative. It will be mapped to 1 if it is positive and 0 otherwise (see lines 6-7). The last step is to define distinguishing features that the user can reliably generate. It means that any binary strings derived from the distinguishing features of any β_k should be identical. Therefore, a bitwise XORing of the distinguishing descriptors will be 0. For this reason, we determine XORing of, $b_k(i)$, *k = 2, . . . , l+1.* If the XORing of $b_k(i)$ is 0, the i^{th} feature will be a distinguishing feature and we set $B(i) = b_2(i)$ (see lines 8-13). Here, the B is a set of distinguishing descriptors of length D (see line 14). Next, the algorithm returns a set of indexes of the distinguishing features (*indexes*), B, and D. Finally, the hardened template is examined and the algorithm returns the number of bits T that corresponds to distinguishing descriptors (see line 15).

5.2.3 Multi-thresholds Generation

We select a set of thresholds in such a way that the entropy of the biometric template is maximized. According to Jain et al., the entropy of the biometric template can be understood as a measure of the number of different identities that are distinguishable by a biometric system [44]. Hence, the set of thresholds that is used in mapping process should yield a binary string that appears to be random in a context of cryptography.

We first generate pseudo-random bits $p \in \{0,1\}^m$ using Blum Blum Shub (BBS) algorithm [9]. Next, a set of thresholds is selected based on the criteria that query biometrics will be mapped to a binary string that is close to p. Finally, the pseudo-

Algorithm 2 Specification of the Mapping algorithm

Input: $\mathcal{T}, \mathcal{H}$
Output: $T, D, B, indexes$
$indexes \leftarrow \{\}, \beta_1 \leftarrow \mathcal{H}$

for $k \leftarrow 1$ *to* $l+1$
  $f_k \leftarrow$ **perform DTW of** $\mathcal{H}$ **and** β_k
  for $i \leftarrow 1$ *to* m
    $f'_k(i) \leftarrow$**linear combination of all components in** $f_k(i)$
    $\phi_k(i) \leftarrow f'_k(i)$-$\mathcal{T}(i)$
    if $\phi_k(i) > 0$
      $b_k(i) \leftarrow 1$ **otherwise** $b_k(i) \leftarrow 0$
for $i \leftarrow 1$ *to* m
  $b(i) \leftarrow 0$
  for $k \leftarrow 3$ *to* $l+1$
    $b(i) \leftarrow b(i) + (b_2(i) \oplus b_k(i))$
  if $b(i) = 0$
    $B(i) \leftarrow b_2(i)$**,** $indexes \leftarrow indexes \cup i$
$D \leftarrow$ **range** $indexes$
$T \leftarrow$ **the number of bits such that** $b_1(indexes(i)) \oplus B(indexes(i)) = 0$
return $T, D, B, indexes$

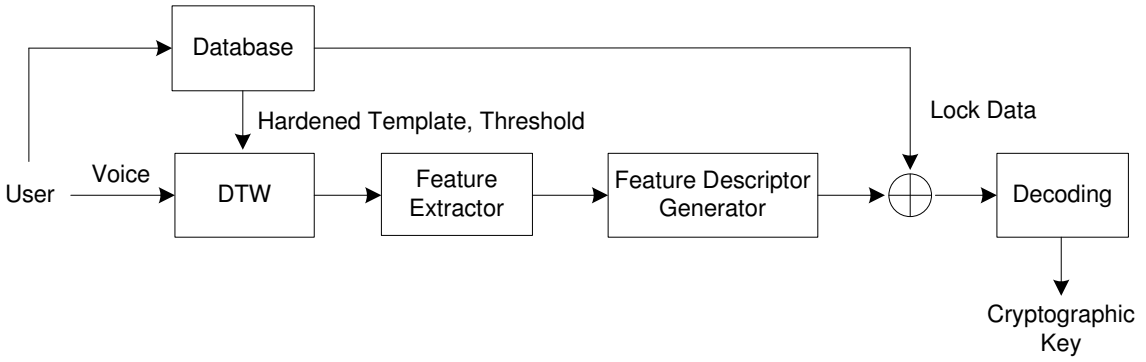


Figure 5.3: Dynamic time warping-based biometric key retrieval in verification phase.

random bits will be securely deleted. As the *Mapping* algorithm simply maps a feature to 1 if the feature is greater than a threshold and 0 otherwise, hence we select a threshold to be lower than the mean of that feature if a corresponding pseudo-random bit is 1 and greater than the mean otherwise. Specifically, to generate the multi-thresholds for any users, the *MultiThreshold* function is used in *Hardening* and *Mapping* algorithm. Let $\mu(i)$ and $\sigma(i)$ be the mean and standard deviation of the linear combination of all features of i^{th} frame over l training utterances, the function executes as follows:

1. Generate pseudo-random bits $p \in \{0,1\}^m$ using BBS algorithm [9].

2. Set the multi-thresholds $\mathcal{T}(i) = \mu(i) + (-1^{p(i)})\ \kappa_i \sigma(i)$ for some parameter $\kappa_i > 0$ which optimize the distinguishing descriptor.

3. Securely delete pseudo-random bits

5.2.4 Biometric Key Retrieval

The biometric key retrieval process is in the verification phase indicated in Figure 5.3. The user requests the template from the database that contains the hardened template, the multi-thresholds, and the lock data. Then the system performs DTW employing a user's pass-phrase. The signal that results from DTW is executed using the algorithm similar to Section 5.2.2 to generate feature descriptors, and the feature descriptors of the distinguishing features (feature descriptors of the relevant indexes in the Ψ) will be XORed with the lock data. The next step is the decoding process. If the error is within the tolerance, the key can be correctly reconstructed. To check whether the key is identical to the key generated in the training phase, a number of researchers [3, 34, 67] checked the hash function. In the training phase, the initialized key, k, was stored as $\mathrm{h}(k)$. Once the key k', is regenerated from the verification phase, the system checks to see whether $\mathrm{h}(k) = \mathrm{h}(k')$. If $\mathrm{h}(k) = \mathrm{h}(k')$, the key, k', is correct.

5.3 Experiments and Results

5.3.1 Experiments Setup

We compare the DBKB with other speaker verification systems: Dynamic Time Warping (DTW) [32], Vector Quantization (VQ) [83], and Gaussian Mixture Model with Universal Background Model(GMM-UBM) [76].

For the DBKB, 121 DFT elements of a full template are reduced to an average of nine and 11 for the MDS and LDS. We set the length of the binary string to 511 bits. For the MDS, we can generate 139 bits on average for each feature; we need

four features to generate 511 bits. For our setting, four features are the Short-Term Energy, the 13 order MFCC, the 12 order Linear Prediction Coefficient (LPC), and the DFT. Nevertheless, some pass-phrases cannot generate a binary string of length 511. In this case, we use a zero padding scheme to adjust the lengths of the binary string of these pass-phrases to that length even if these pass-phrases may degrade the recognition performance of our approach. For LDS, we can generate 221 bits for each feature. However, we use the same features in the MDS.

For DTW, VQ, and GMM, the attack models and the parameters are set to be the same as the previous chapter. For the DBKB, the difference is for the A-II which we are going to describe now.

For the DBKB template attack, the attacker can exploit a hardened and multi-thresholds template. We consider two approaches for the attack. For the first approach, we directly invert the hardened template (DFT vectors) to the signal using Inverse Discrete Fourier Transform (IDFT). For the second approach, we will search in sources' pass-phrases similar to VQ and GMM attack. Therefore, the average of multi-thresholds is used as the decision threshold for analysis. Upon examination of the first approach, the EER was 0%. Hence, we employ the second approach for the DBKB attack. The results are illustrated in Figure 5.4.

5.3.2 Performance

The same datasets in the previous chapter are used in the experiments. Figure 5.5 shows the recognition performance (EERs) of the DTW, VQ, GMM-UBM, and DBKB for same-gender experiments. From all attack models except the A-II, the DTW

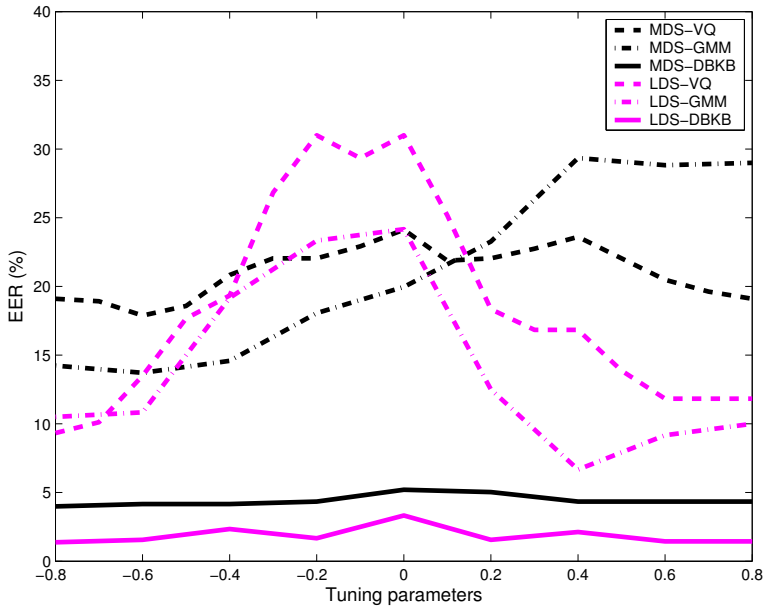
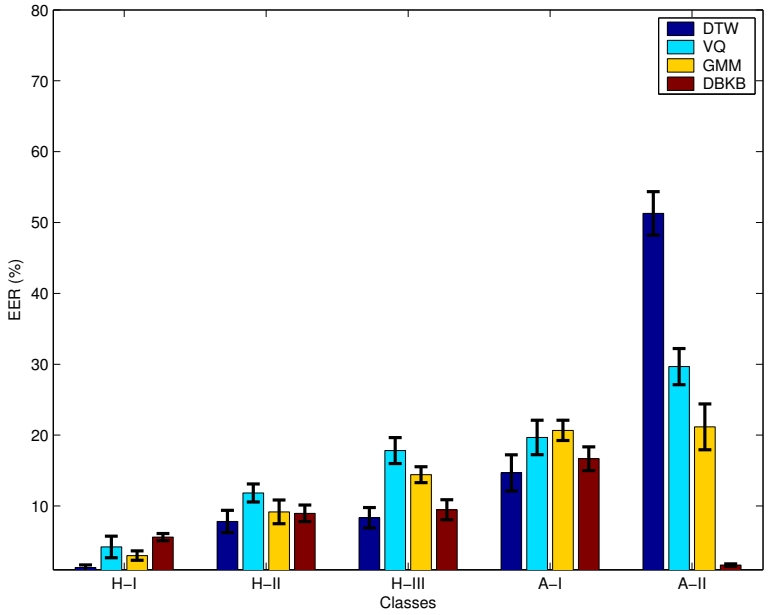


Figure 5.4: The error rates of regenerated pass-phrases by varying κ in VQ, GMM, and DBKB.

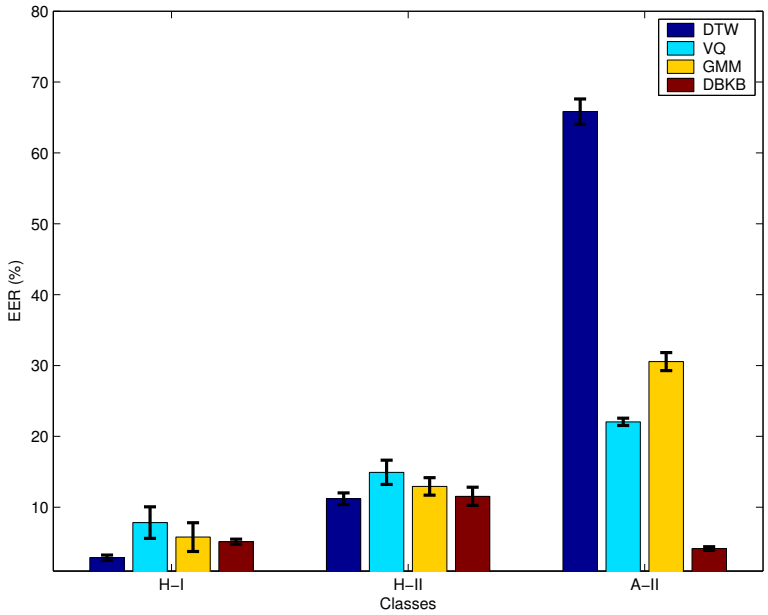
yields the best performance while the DBKB has the second best results. However, the difference between the DTW and DBKB is slight. As indicated in Figure 5.5, the EER of the DTW method against the A-II is significantly higher than others, therefore, its slightly better performance has no merit because the security and privacy are significantly lost.

The EERs of the DBKB against the A-II are the lowest when we compare them with other attack models and other systems. This is due to the design. The security of the template is the main issue for the DBKB. We previously mentioned that, in the hardening process, the system examines the template before storing it. Therefore, it is not surprising that the EER of our construction against the A-II is the lowest.

The operating points of the DBKB for imposter trial are 38 and 54 bits for the MDS and LDS. We use t-error-correcting BCH [56] which denoted by *BCH(n, k, t)*

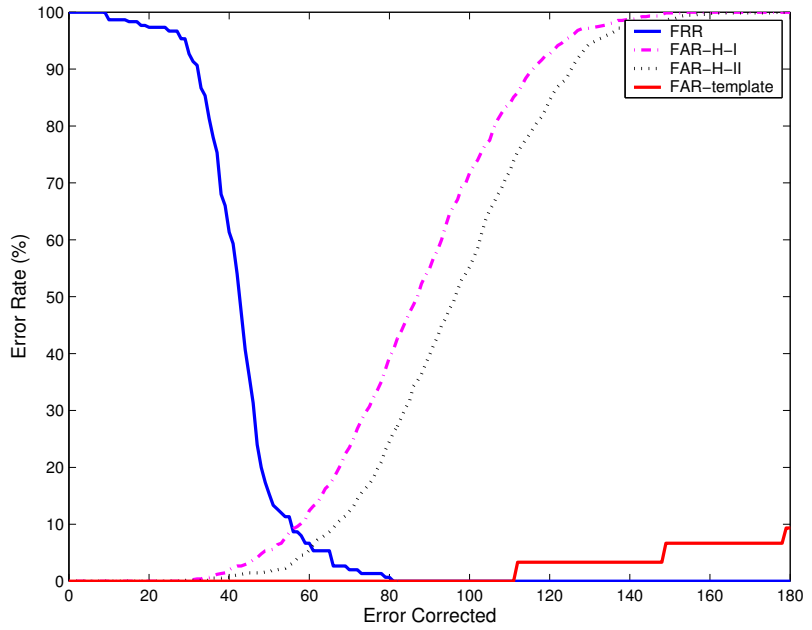


(a)

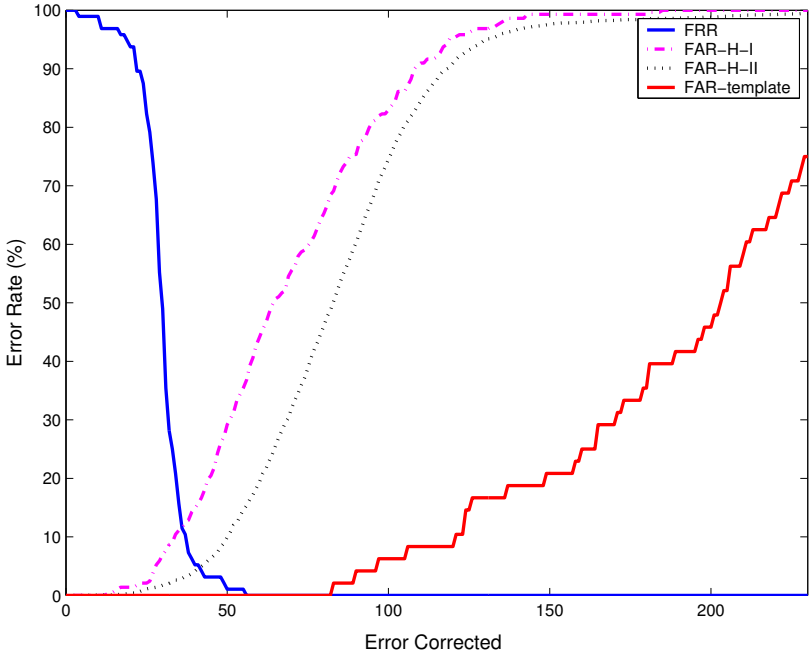


(b)

Figure 5.5: The EERs against various attacks and models with the 95% confidence interval for same-gender experiments on the DTW, VQ, GMM, and DBKB (a) the LDS and (b) the MDS



(a)



(b)

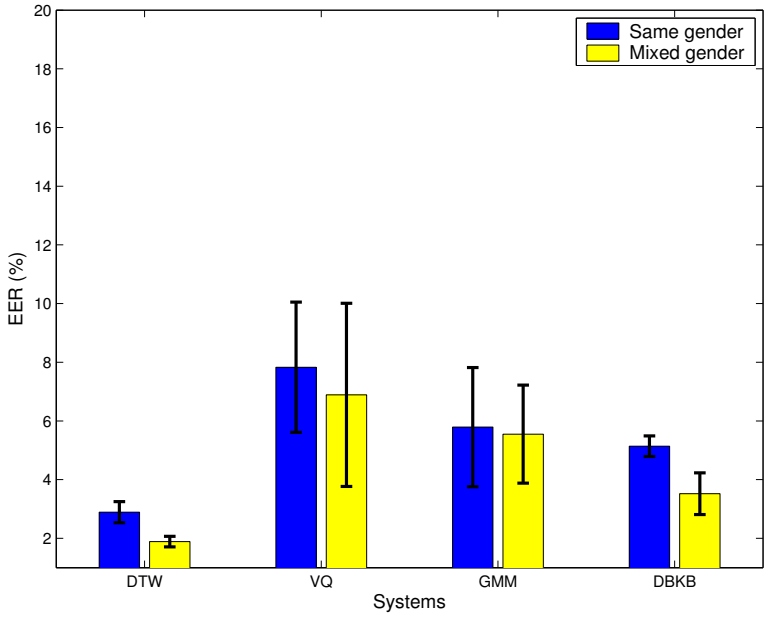
Figure 5.6: The performance of the DBKB against attackers using random pass-phrases (Random), true pass-phrases (Imposter), and the templates (Template): (a) the LDS (b) the MDS.

where n is a block length, k is the key, and t is correctable bits. Hence, *BCH(511, 229, 38)*, and *BCH(511, 139, 54)* are employed for the MDS and LDS. By testing with pass-phrases of 1.87 and 3.05 seconds on average in the MDS and LDS, we can generate the cryptographic key up to 229 and 139 bits that exceed the requirement of 128-bit Advanced Encryption Standard (AES).

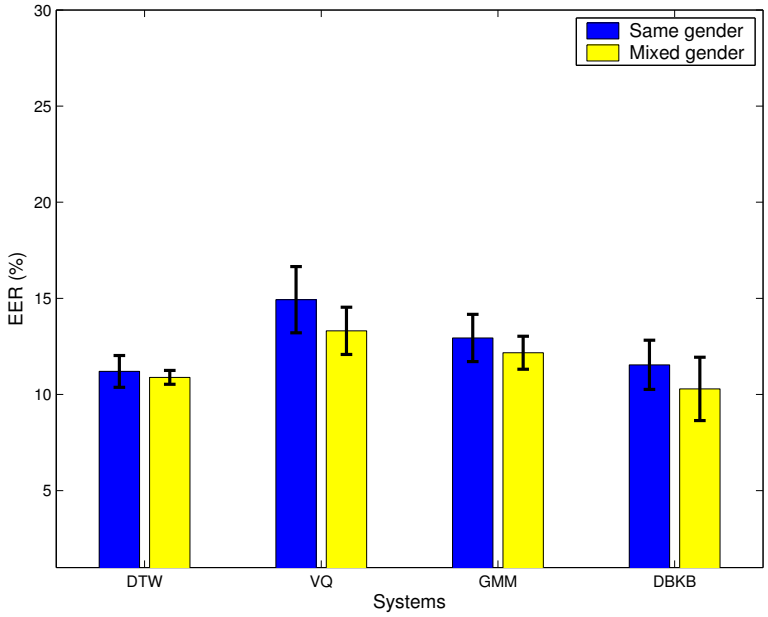
Figure 5.6 shows the plots of the recognition performance of the DBKB against various attacks including the attackers who acquire the hardened template. Assuming that the attackers use a hardened template to derive the key the same way as in the random pass-phrase and imposter trial, the EERs of the template attack are 0% in both datasets. However, more analysis of the security of the template is provided in Section 5.3.3 where the attackers have perfect knowledge of the correlation of the features.

We also show the FARs when the decision thresholds are set to be the operating point of the H-II's EER. These results are indicated in Table 5.1 which illustrates the error rates (FAR) of various attacks. The figures of the H-I and H-II in the table reflect the standard (traditional) evaluation of biometric authentication systems. Beyond the standard evaluation, the FARs of other attack models are very high for the VQ and GMM. In particular, the FARs of the A-II are the highest. For the DTW and DBKB, the FARs of the H-I, H-II, H-III, and A-I are close, but the A-II's FARs of the DBKB are significantly lower than the DTW. These results are also another evidence to demonstrate that the security of our scheme is better.

Furthermore, we show gender-based results under the same setting of the previous chapter. These results are shown in Figure 5.7 and 5.8. Figure 5.7 illustrates comparisons of the H-I and H-II results for the same-gender and mixed-gender experiments



(a)



(b)

Figure 5.7: Comparisons of the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS (a) the H-I (b) the H-II

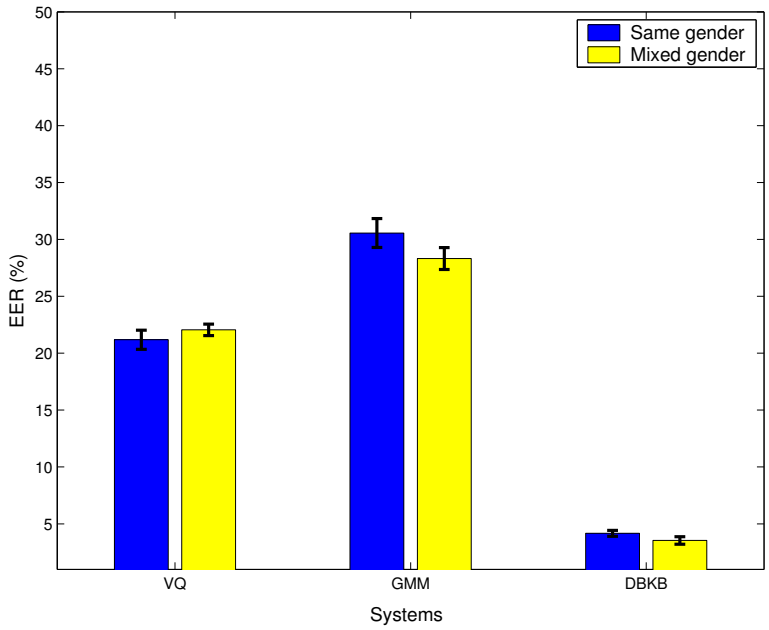


Figure 5.8: Comparisons of the A-II for the same-gender and mixed-gender experiments with the 95% confidence interval on the DTW, VQ and GMM system in the MDS.

for the DTW, VQ, GMM, and DBKB system. Figure 5.8 illustrates comparisons of the A-II results for the same-gender and mixed-gender experiments for the DTW, VQ and GMM system. The EERs of the DBKB are similar to the DTW for the same reason we have mentioned in the previous chapter.

5.3.3 Security Analysis

The security of the scheme is based on the template protection. Our scheme falls under the hybrid schemes. First, the DTW template is protected using a non-invertible transformation scheme. The algorithm will search for a set of features in order to use them as the hardened template. Next, the key binding scheme is applied to protect the key, and then the training data will be securely deleted from the system. It is

Table 5.1: FARs (%) of speaker verification systems (DTW, VQ, GMM, and DBKB) against various attacks using decision thresholds at operating points of imposters (H-II).

Datasets	Attack models	DTW	VQ	GMM	DBKB
LDS	H-I	0.27	3.53	2.22	4.90
	H-II	7.20	11.56	8.89	8.26
	H-III	8.67	25.47	24.05	8.53
	A-I	20.00	26.67	60.00	23.33
	A-II	**90.00**	**55.00**	**65.00**	**0.00**
MDS	H-I	0.00	4.08	2.08	3.52
	H-II	11.86	16.40	13.12	12.26
	A-II	**100.00**	**47.22**	**89.93**	**2.78**

computationally hard to decode the key without any knowledge of biometric data [44].

We can estimate the security of the scheme using the sphere packing bound [63] similar to Hao's work [34]. Let z be the uncertainty of voice and w be the error bits that can be corrected by the system, the lower bound $\mathcal{BF}$ can be set by the following equation.

$$\mathcal{BF} = \frac{2^z}{\sum_{i=1}^{w} \binom{z}{i}} \tag{5.1}$$

To estimate the lower bound, we use two verification recordings of each speaker in the MDS. We carry out 4,512 of inter-speaker comparisons to evaluate the uncertainty. The following steps are the uncertainty analysis [22]. For more detail, we refer to Daugman's work [22]. For each comparison, the Normalized Hamming Distance (*NHD*) between two binary templates, A and B, is given in the following equation where $\mathcal{D}_H(A, B)$ is a function to calculate the Hamming distance between A and B.

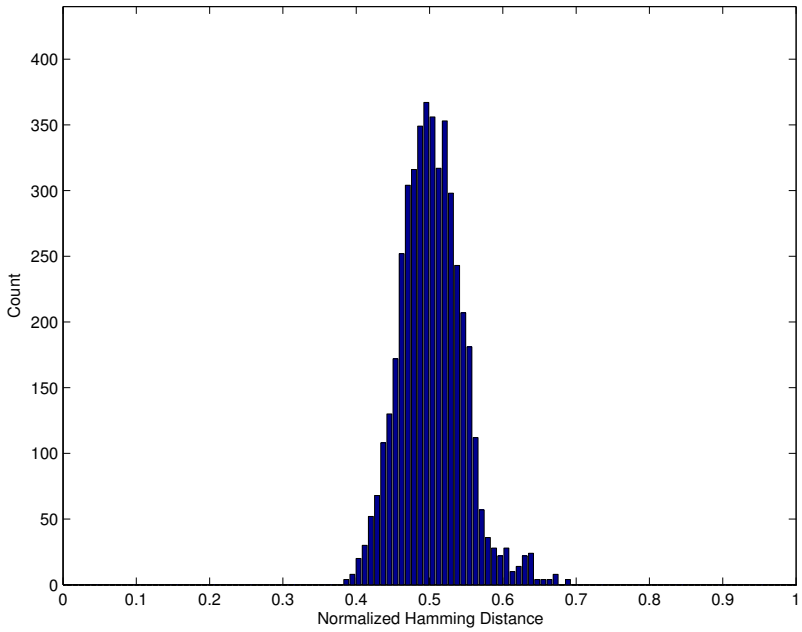
$$NHD = \frac{\mathcal{D}_H(A,\ B)}{511} \tag{5.2}$$

Hence, $NHD = 0$ would represent a perfect match. Figure 5.9 (a) shows the distribution of the *NHD* of inter-speaker comparisons where $\mu = 0.5281$ is mean and $\sigma = 0.0455$ is standard deviation of *NHD*. The result in this figure is close to a binomial distribution as shown in Figure 5.9 (b) which has the the fractional function in the following equation where $N = 120$ and $\mu = 0.5$.

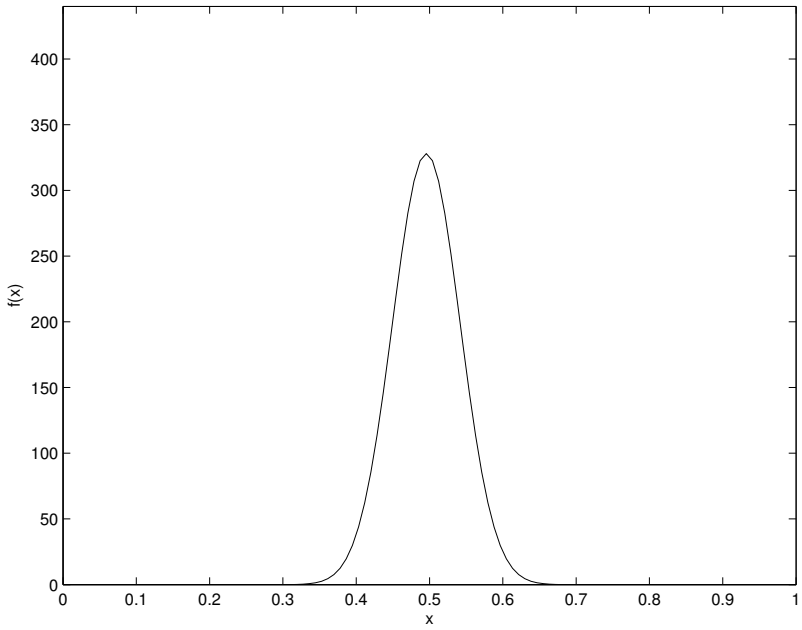
$$f(x) = \binom{N}{y}\mu^y(1-\mu)^{N-y},\ x = y/N \tag{5.3}$$

These results indicate that the difference between the binary templates is likely distributed to be a binomial experiment of 120 repeated trials with $\mu = 0.5$. Therefore, for a binary string of 511 bits, it has approximately 120 degrees-of-freedom. For the H-II, the system should be able to correct the error up to 38 bits (imposter trial), that is approximately 8%. Here, z is 110 bits and w is 10 bits. The estimated entropy is 73 bits.

However, some information may leak from the hardened template and multi-thresholds. Recall that any bits in the binary template were set to be close to pseudo-random bits p, any given bits in the binary template were equally likely to be 1 or 0. If a binary string generated from the hardened template is random, the expected agreement between the binary template and the binary string derived from the hardened template should be close to 256 bits. Upon testing in the MDS, the expected agreement is 303 bits, the result implies that some information leaks from



(a)



(b)

Figure 5.9: The comparison of a distribution of Normalized Hamming Distance and a binomial distribution (a) Distribution of Normalized Hamming Distances obtained from 4,512 comparisons of inter-speaker in the MDS and (b) A binomial distribution with $\mu = 0.5$ and $N = 120$ degrees-of-freedom.

the hardened template. Furthermore that the attackers randomly guess the remaining $511 - 303 = 208$ bits can get 104 bits correct. In the worst case, we assume that the attacker can correctly locate these $303 + 104 - 256 = 151$ bits. Hence, 151 bits or $\frac{151}{511} \times 100 = 29.54\%$ leak from the hardened template. As a result, the estimated entropy will be 51 bits.

We further carry out 4,512 of inter-speaker comparisons using the global-threshold scheme where the threshold is fixed. The estimated entropy is 16 bits. Table 5.2 summarizes the security when we compare the multi-thresholds to the global-threshold scheme in the MDS. It is clear that the entropy of the multi-thresholds scheme is significantly improved.

Table 5.2: The security of the multi-thresholds and the global-threshold scheme in the MDS.

	Multi-thresholds	Global-threshold
Estimated Entropy (bits)	**51**	16

5.4 Summary

We addressed two problems in a cryptosystem. First, the problem of the feature correlation could be mitigated by using the proposed multi-thresholds. As a result, the randomness of the key (entropy) was increased from 16 to 51 bits. Second, we addressed the challenge in using DTW in a cryptosystem, more specifically, that the template must be useful to create a warping function, while it must not be usable for an attacker to derive the cryptographic key. A solution, the hardened template, was proposed. We showed that the EERs against the attackers using the hardened template were 0%. We compared our system with DTW, VQ, and GMM-UBM speaker

verifications. The DTW yielded the best performance while ours had the second best results. However, the differences between the DTW and ours were slight. We noted that the DTW speaker verification is not secure and it leaves all the biometric information (a full set of DFT templates) in the system. Hence, its slightly better performance has no merit because the security and privacy are significantly lost. We also investigated the results based on gender information. There were no significant differences.

Chapter 6

Performance and Security of the Hardened Template

A DTW-based biometric user verification system needs a DTW template to set up a warping function for query biometrics. In addition, a matching template is required to examine similarity. A transformation approach utilized a transformation function to protect a DTW template. Unfortunately, the matching template was not protected properly. In this chapter, we first show that an adversary can exploit the matching template to gain access to the system. We also compare our scheme (hardened template in the previous chapter) with a transformation approach and an unprotected method. Moreover, we continue to demonstrate the security of the hardened template. First, we prove that it is hard to recover the original template. Then, we focus on an algorithmic attack.

6.1 Introduction

To our knowledge, the first approach to protect a DTW template was applied to online handwriting [35]. The authors protected the DTW template by using only static features while only dynamic features were utilized for verification. More precisely, the static feature used in this work was (x,y) coordinates. The dynamic features were pen-down time, Root Mean Square (RMS) of V_x, RMS of V_y, etc. Moreover, the authors protected a matching template by utilizing a cryptographic framework. The entropy of their scheme was approximately 40 bits which was better than 18-30 bits for eight characters password-based system [14]. However, the authors did not compare the error rates when the template was not protected.

Maiorana et al. [58] proposed a scheme which utilized a convolution function to transform a template. The best result was reported that the error rate only degraded from 4.07% to 5.22% when they compared the protected template with unprotected. In their work, the transformed versions were stored as templates and used as a DTW template and a matching template. The system performed DTW with each template to query biometrics and then the minimum distance was selected. The system decided whether to accept that biometrics by comparing the minimum distance to a decision threshold. Even though they proved that to recover the original templates was computationally as hard as random guessing [59], the system left the transformed templates which could be used in gaining access to the system.

In this chapter, we compare the recognition performance of the transformation approach with ours (DBKB). We conduct two experiments which are described in Section 6.2. For the first experiment, we aim to demonstrate that the transformed

template (detailed in Section 6.2) can be used to gain access to the system which does not differ from the traditional approach (Unprotected template). For the second experiment, we compare performance of the transformed template and the hardened template when they are utilized in our scheme. The results are also compared with the unprotected approach.

6.2 Transformation Approach

In this section, we describe the transformation approach as a scheme to compare it with ours (DBKB). The transformation approach is based on Maiorana et al.'s scheme [58]. Let the original template $\mathcal{RF}$ be a set of vectors $\mathcal{RF} = \{r(n), n = 1, \ldots, N\}$. The $r(n)$ is an F-element vector $r(n) = [r_1(n), \ldots, r_F(n)]^T$. The transformation version $\mathcal{TF}$ is another set of vectors, $\mathcal{TF} = \{f(n), n = 1, \ldots, K\}$. The $f(n)$ is also an F-element vector $f(n) = [f_1(n), \ldots, f_F(n)]^T$. To derive $\mathcal{TF}$, the $\mathcal{RF}$ is first partitioned into W segments.

Let $b_j = \lfloor (\frac{d_j}{R} N) \rfloor$ for $j = 0, \ldots, W$ where d_j is selected randomly from a set of integer $d = [1, R-1]$ such that $d_j > d_{j-1}$ and R is an upper bound of d_j. In addition, d_0 and d_W are set to 0 and R. The original sequence $r_{i \in [1,F]}(n)$ is divided into W segments of length $N_j = b_j - b_{j-1}$. Each segment is represented by equation 6.1 for $n = 1, \ldots, N_j$ and $j = 1, \ldots, W$.

$$r^j_{i \in [1,F]}(n) = r_{i \in [1,F]}(n + b_{j-1}) \tag{6.1}$$

The $f_{i \in [1,F]}(n)$, $n = 1, \ldots, K (K = N - W + 1)$ is then obtained through the linear

convolution of the function $r^{j}_{i\in[1,F]}(n)$, $j = 1, \ldots, W$ represented by equation 6.2.

$$f_{i\in[1,F]}(n) = r^{1}_{i\in[1,F]}(n) * \ldots * r^{W}_{i\in[1,F]}(n) \tag{6.2}$$

6.2.1 Experimental Setup

We investigate the performance of three systems: the unprotected approach, transformation approach, and our approach (DBKB) with the MDS. For all constructions, we use a low-pass digital filter with a cut-off at 4 kHz. The signal is pre-emphasized by passing the signal to a first order digital filter $H(z) = 1\text{-}\gamma z^{-1}$, where we set $\gamma = 0.98$. Framing is the next step. The signal is framed into the short time analysis interval. Each frame is multiplied by a window function (Hamming). For the sampling rate of 8 kHz, we use 240 samples per frame that are shifted every 80 samples.

Unprotected Approach

For the unprotected approach, we employ a DTW user verification system where the DTW and matching template are a set of 13 order Mel-Frequency Cepstrum Coefficients (MFCCs). More precisely, we use the first utterance in the training set as the reference signal (DTW template) and then perform DTW to the rest. The averaged result is stored as the matching template. The distance between an input and the matching template is determined by using the Euclidean distance. The system decides whether to accept or reject the speaker by comparing the Euclidean distance to the decision threshold.

Transformation Approach

For the transformation approach, the parameter W in Section 6.2 can be understood as the security and error rate index. From Maiorana et al.'s report, the error rate was the best when they set W to the lowest value. For the security issue, the attackers have to figure out W unknown functions to invert the transformed template. Thus, the security may be harmed if we set the W too low. However, in this experiment, we set $W = 2$, the lowest, to guarantee that it will yield the best recognition performance. In addition, the R is set to 100. The features and construction applied to this system are the same as the unprotected approach.

DBKB

For the DBKB, 121 DFT elements of a full template are reduced to an average of nine. We set the length of the binary string to 511 bits. For the MDS, we can generate 139 bits on average for each feature; we need four features to generate 511 bits. For our setting, four features are the Short-Term Energy, the 13 order MFCC, the 12 order Linear Prediction Coefficient (LPC), and the DFT. Nevertheless, some pass-phrases cannot generate a binary string of length 511. In this case, we use a zero padding scheme to adjust the lengths of the binary string of these pass-phrases to that length even if these pass-phrases may degrade the recognition performance of our approach.

6.2.2 Experimental Results

For the experiments in this chapter, we evaluate the systems with the MDS. We use six recordings to train the systems. Two recordings are used for verification. To investigate the performance of the system, we use the same pass-phrase uttered by other speakers to evaluate the *imposter trial*. The number of imposters available in the dataset varies from 1 to 6 (same-gender experiment). In addition, we use six pass-phrases of other speakers that are different from the verification pass-phrase to evaluate the *random trial*.

Experiment I

For the first experiment, we employ a DTW user verification system. We compare the recognition performance of the transformation approach (transformed template) with the unprotected approach (unprotected template). Beyond the random and imposter trial, we present a *generative attack* which we are going to describe now.

We know that the 13 order MFCCs of the training utterances are stored as the matching template. Hence, we have to transform this template to a signal. We first transform MFCCs to DFTs using Auditory Toolbox [81]. Then the DFTs are transformed to the speech signal used as a forgery. We refer to this attack as the *generative trial*.

For each attack, we repeat the experiment 30 times. Each time, we randomly select an adversary pass-phrase from a set of dedicated imposters and assign it to each user. Therefore, we can determine the confidence interval on the mean using equation 4.3.

Table 6.1: Equal Error Rates (EERs) with the 95% confidence interval of the transformation approach (transformed template) and the unprotected approach (unprotected template) for the DTW-based systems against random attack, imposter, and generative.

DTW Template	EER ($\bar{x} \pm Z_{\frac{0.05}{2}} \frac{s}{\sqrt{n_s}}$ %)		
	Random	Imposter	Generative
Transformed	8.06 ± 2.28	16.46 ± 2.73	45.31 ± 1.95
Unprotected	**2.89 ± 0.36**	**11.20 ± 0.83**	65.83 ± 1.78

The experimental results are illustrated in Table 6.1 with the 95% confidence interval. The EERs of the random and imposter trial are noticeably degraded when we compare the transformation approach with the unprotected. These results are consistent the Maiorana et al.'s work [58]. Let $\mathcal{DF} = \frac{E_T - E_B}{E_B}$ be a degradation factor where E_B is an EER of the original template and E_T is an EER of the transformed template. The degradation factor in Table 6.1 (imposter) and the Maiorana's work (Table 1 in [58], $W = 2$) are 0.46 and 0.47.

For the generative attack, even if the EER of the transformed template is significantly better than the unprotected, it is still very high when we compare it with the other trials. These results demonstrate in a very convincing way that the matching template must be protected.

Experiment II

We employ our construction (DBKB) to protect the matching template. For the DTW template, the DBKB uses the hardened template to protect the reference signal (Section 5.2). Hence, this template (hardened) can be replaced with the transformed version (Section 6.2). In this experiment, we compare the performance of the DBKB when the hardened and transformed template are

Table 6.2: Equal Error Rates (EERs) with the 95% confidence interval of the DBKB when the transformed template and hardened are applied.

DBKB Template	EER ($\bar{x} \pm Z_{\frac{0.05}{2}} \frac{s}{\sqrt{n_s}}$ %)		Error Corrected {Random, Imposter}(bits)
	Random	Imposter	
Transformed	10.69 ± 1.35	17.27 ± 0.42	{41, 37}
Hardened	**5.14 ± 0.35**	**11.54 ± 1.28**	{42, 38}

used as the DBKB's DTW template. The results are illustrated in Table 6.2 with the 95% confidence interval. The performance of the hardened template noticeably outperforms the transformed version.

We also compare the results with the unprotected approach. The comparisons are illustrated in Figure 6.1. For the imposter trial, the performance of our approach is not far from the unprotected approach and it noticeably outperforms the transformation approach. For the random trial, the performance of our approach is slightly degraded, but it is significantly better than the transformation approach.

The operating points of the DBKB for imposter trial are 38 and 37 bits for the hardened and transformed template (Table 6.2). We use t-error-correcting BCH [56] which denoted by *BCH(n, k, t)* where n is a block length, k is the key, and t is correctable bits. Hence, *BCH(511, 229, 38)*, and *BCH(511, 238, 37)* are employed for the hardened and transformed template. By testing with passphrases of 1.87 seconds on average, we can generate the cryptographic key up to 229 and 238 bits that exceed the requirement of 128-bit Advanced Encryption Standard (AES).

For security against the generative attack, we illustrate the results in Figure 6.2,

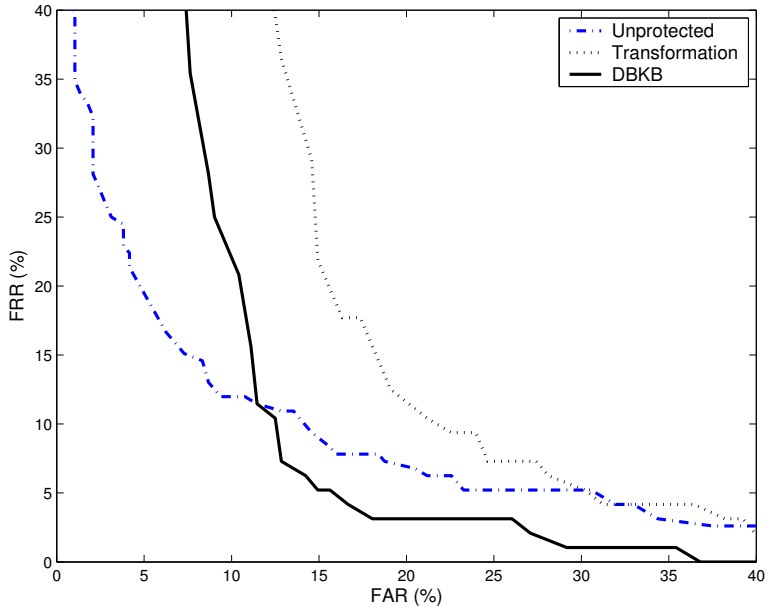
which shows the plots of the recognition performance of the DBKB against various attacks including the attackers who acquire the DBKB's DTW templates. Assuming that the attackers use the hardened and transformed template to derive the key the same way as in the random and imposter trial, the EERs of the template attack are 0% and 2.08%. However, more analysis of the security of the template is provided in Section 6.3.

6.3 Security of the Hardened Template

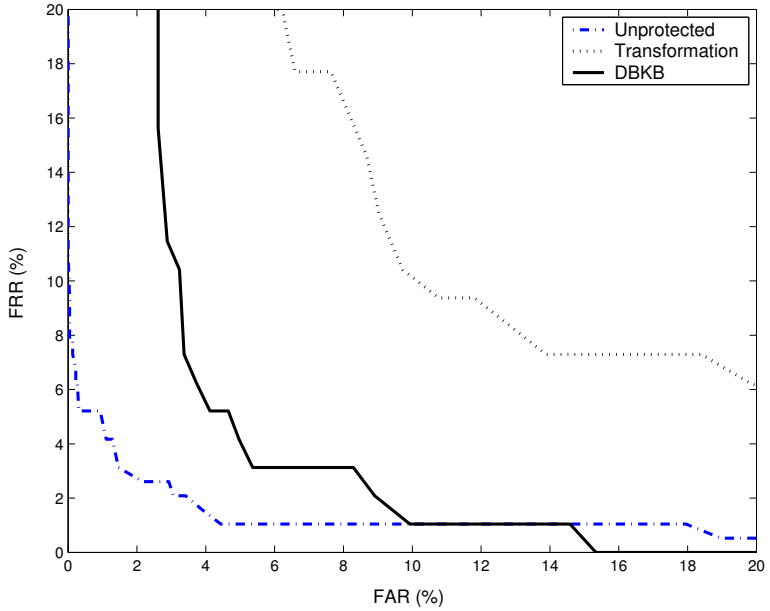
We have shown that the attacker utilizing the hardened template to derive the key cannot access the system. In this section, we will show whether the attacker can invert the hardened template.

Given N points of speech signal $x[n]$, we can derive the Discrete Fourier Transform DFT $X[k] = \sum_{n=0}^{N-1} x[n] \exp \frac{-j2\pi nk}{N}$ for $k = 0, \ldots, N-1$. If we write the series expression for $X[k]$ for each value of k, we obtain a set of N equations as shown in the following where $W_N = e^{-j\frac{2\pi}{N}}$.

$$\begin{aligned}
X[0] &= \tfrac{1}{N}\big[x[0] + x[1] + \ldots + x[N-1]\big] \\
X[1] &= \tfrac{1}{N}\big[x[0] + x[1]W_N^1 + \ldots \\
&\quad + x[N-1]W_N^{N-1}\big] \\
\vdots &\quad \vdots \\
X[N-1] &= \tfrac{1}{N}\big[x[0] + x[1]W_N^{(N-1)} + \ldots \\
&\quad + x[N-1]W_N^{(N-1)(N-1)}\big]
\end{aligned}$$

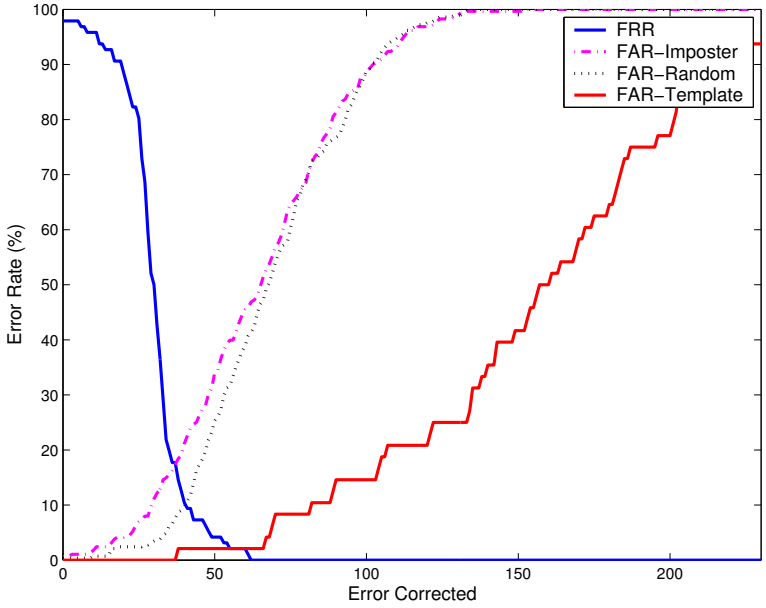


(a)

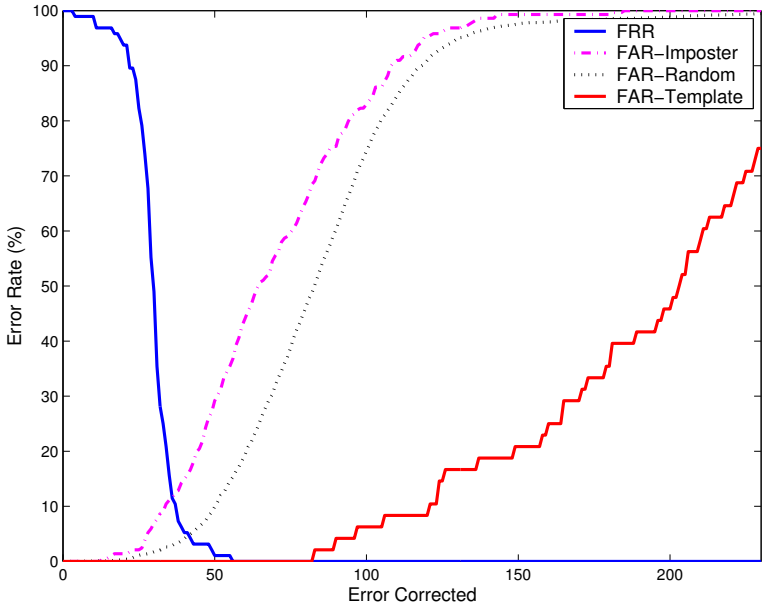


(b)

Figure 6.1: The ROC curves of three approaches: Unprotected, transformation, and our approach (DBKB). (a) the imposter trial and (b) the random trial.



(a)



(b)

Figure 6.2: The EERs of attackers using true pass-phrases (Imposter), random pass-phrases (Random), and generative attack (Template): (a) the transformation approach is applied to the DBKB's DTW template (b) the hardened template is utilized.

From the experiment's result, the 121 DFT coefficients were reduced to nine. According to the real function and symmetric property [72], the attackers are left with 18 equations with 240 unknown variables. Thus, these equations cannot be resolved. Even if it is hard to recover the original template, the template may be useful for attackers to devise algorithms for creating forgeries to gain access to the system. Hence, a mathematical proof may not work here (nor with the transformation approach). The following section uses the experiments to demonstrate the security of the template in this case.

6.3.1 Experimental Setup

We also use the MDS in this experiment. We assume that the attackers acquire the hardened template which consists of U unknown elements out of N and $K = N - U$ known elements. In addition, they know the pass-phrases and they collect samples of $\lceil \frac{N}{K} \rceil$ pass-phrases for analysis.

By definition, the hardened vector $\mathcal{H} = \{X^T | \exists X^T[k] = 0\}$ where $X^T[k]$, $k = 1, \ldots, N$ is a full template of a target speaker, the attackers know that any elements $X^T[k] = 0$ are likely to be the elements which was removed. Then they arrange the index of those elements in ascending order, such that $k(j) < k(j+1)$, $j = 1, \ldots, U$, in $\nu = \{k(j)\}$. In the same way, the index of known elements is $\bar{\nu} = \{k(i), i = U+1, \ldots, N\}$.

Modification

Given the DFT $X^T_{k\in\bar{\nu}}[k] = \sum_{n=0}^{N-1} x[n] \exp \frac{-j2\pi nk}{N}$ where $x[n], 1, \ldots, N$ is an original signal, we have shown in the previous section that the attacker cannot derive $x[n]$. They have to modify the pass-phrase samples $x^s[n]$ denoted by $x^{\bar{s}}[n]$ such that

$$X^T_{k\in\bar{\nu}}[k] = \sum_{n=0}^{N-1} x^{\bar{s}}[n] \exp \frac{-j2\pi nk}{N} \tag{6.3}$$

Let $a[n]$ be an estimation function such that

$$x^{\bar{s}}[n] = a[n]x^s[n] \tag{6.4}$$

Now, we substitute $x^{\bar{s}}[n]$ in 6.3; we obtain

$$X^T_{k\in\bar{\nu}}[k] = \sum_{n=0}^{N-1} a[n]x^s[n] \exp \frac{-j2\pi nk}{N} \tag{6.5}$$

We have to determine $a[n]$ to estimate the transformed signal $x^{\bar{s}}[n]$. For each pass-phrase sample, we obtain K equations with N unknown variables; we need $\lceil \frac{N}{K} \rceil$ pass-phrase samples to determine $a[n]$. For the MDS, N is 121 and K is 9; we need $\lceil \frac{121}{9} \rceil = 14$ pass-phrase samples. Hence, these samples are selected from mixed-gender imposter's pass-phrases.

Table 6.3: Equal Error Rates of the modified pass-phrase attack, the original imposters' pass-phrases, the adversary using template information only.

Attack Models	EER (%)
Template	1.86
Imposters' pass-phrases	11.96
Modified pass-phrases	5.43

6.3.2 Experimental Results

Table 6.3 shows the EERs of the modified pass-phrase attack when we compare it with the EER of the original imposters' pass-phrases and the EER of the adversary using template information only. The experimental results show that the EER of the modified pass-phrase attack is noticeably better than the EER of the adversary using template information, but it is significantly lower than the EER of the original imposters' pass-phrases. Hence, the attack in this scenario is not useful for the adversaries as the original imposters' pass-phrases attack is better.

6.4 Summary

Even though the DTW transformed template is computationally hard to invert to the original template, we have shown that the adversary can exploit the transformed version to attack the system on-line. We compared our approach (DBKB) with the transformation approach and the unprotected method. The experimental results showed that the recognition performance (EER) was almost the same when we compared the DBKB with the unprotected approach. In addition, our system noticeably outperformed the transformation approach. We have also demonstrated that the transformation approach can be applied in the DBKB; the EER was only slightly

increased.

We have shown that the attacker directly utilizing the hardened template cannot gain access to the system. The EER of this attack was 0%. Even if it is impossible to recover the original template, the attacker may devise an algorithm by exploiting the hardened template and imposter's pass-phrase information to re-synthesize the forgeries. We evaluate this attack by devising an algorithm which exploits the template information and imposter's pass-phrases to re-synthesize the pass-phrases. The results show an improvement in gaining access to the system when we compare them with template information only attack. However, percentage to be accepted by the system is still lower than imposter's pass-phrase only attack.

Chapter 7

Speech Cryptographic Key Regeneration based on Password

Thus far, we have shown that the security of biometric templates we proposed is significantly improved. In doing so, we demonstrated the security of the templates both mathematical proof and empirical experiments. Furthermore, we have evaluated the recognition performance against several attacks. The experimental results showed that its recognition performance of our scheme is marginally degraded when we compare ours with the DTW system and noticeably improved when we compare ours with the VQ and GMM system. However, the error rates were still high. In this chapter, we address this problem by proposing a way to combine a password with a speech biometric cryptosystem. We present two schemes to enhance verification performance in a biometric cryptosystem using password. Both can resist a password brute-force search if biometric is not compromised. Even if the biometric is compromised, attackers have to spend many more attempts in searching for cryptographic

keys when we compare ours with a traditional password-based approach. In addition, the experimental results show that the verification performance is significantly improved.

7.1 Introduction

Personal authentication systems that yield high security and verification performance are desired. In addition, convenience-to-use systems are preferred. The traditional knowledge-based (e.g., password) and token-based (e.g., smartcard) authentication systems meet the requirement of verification performance, but their security is a concern. In analyzing the security of a knowledge-based system, one of the factors is the complexity of passwords. Unfortunately, users tend to select a password which is easy to guess [30]. To select more complex passwords, users have to follow advice and rules which are different for each system. Thus, it is difficult and inconvenient to remember every systems' password. For a token-based system, the security of the system may be compromised when tokens are stolen. Moreover, carrying tokens all the time may be inconvenient. For these reasons, biometric-based systems have been proposed by a number of researchers to address the mentioned issues. Furthermore, they offer properties, such as proof of identity.

To date, it is well known that biometric systems are vulnerable to attack [70]. In particular, the security of a stored template is a serious concern. To alleviate this problem, researchers proposed a biometric cryptosystem to secure the template. However, the verification performance is degraded. Moreover, the error rate is unacceptable, for example the work in [39]. Even though the authors showed that

the verification performance was slightly degraded when it was compared with the unprotected template approach, its error rate was still high.

Thus far, one of the promising ways to authenticate users is to combine a biometric cryptosystem with the other factors: knowledge or token. Therefore, the performance is improved in the case that the biometrics and the input factors are not compromised simultaneously.

We consider the knowledge-based approach to be another factor in improving a biometric cryptosystem because the users do not need to carry a token. However, we need to deal with weak passwords selected by users. For this issue, we will show that the proposed scheme offers better properties than a traditional password-based approach when the biometrics is compromised.

For a biometric system based on a password, it must be ensured that the attackers must not be able to discriminate the correct password from incorrect when they utilize a brute-force search to find the key without knowledge of the biometrics. On the other hand, even if the password is compromised, it cannot be used to reveal the key. Hence, in this case (the password is compromised), the security of such a construction is still the same as before the password is used.

In this chapter, we present Speech Cryptographic Key Regeneration based on user's Passwords (SCKRP). The SCKRP is a cryptographic framework that binds a biometric template with a pseudo-random key to create a protected template. We propose two schemes to enhance verification performance in a biometric cryptosystem using password. The proposed schemes are: transformation and permutation. Both can resist password brute-force search if biometric is not compromised. Even if the

biometric is compromised, the security meets the same level of the password approach. On the other hand, the security provided by the biometric cryptosystem is not affected even when the password is compromised. We utilize Dynamic Time Warping (DTW) in our scheme. A DTW-based biometric user authentication system needs a DTW template to set up a warping function for query biometrics. In addition, a matching template is required to examine similarity. We utilize a hardened template proposed in Chapter 5 to protect the DTW template. For the matching template, it is protected by cryptographic framework. Next, the hardened template and query biometrics will be transformed using a password. We then introduce a scheme for mapping behavioral biometric measurements (feature vector) to a binary string which can be combined with a pseudo-random key for cryptographic purposes. These steps are detailed in Section 7.2.

We evaluate SCKRP verification performance using Equal Error Rate (EER) with a public database: The MIT mobile device speaker verification corpus [97] available from MIT.

We consider three different scenarios in evaluating the SCKRP: I) Genuine: When an adversary does not access genuine biometrics and passwords. II) Compromised passwords: When an adversary accesses genuine passwords. III) Compromised biometrics: When an adversary acquires genuine biometrics. Then, we compare the system with unprotected Dynamic Time Warping-based speaker authentication [32]. Next, we compare ours with the protected approach in [39]. These experiments are detailed in Section 7.3. Finally, the results and security analysis are illustrated in Section 7.4.

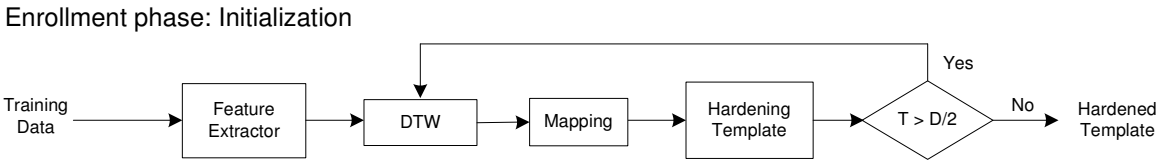


Figure 7.1: Enrollment phase: Initialization.

7.2 Speech Cryptographic Key Regeneration based on Password (SCKRP)

The SCKRP can be overviewed as two phases: Enrollment and Verification. The biometric key regeneration is in the enrollment phase that comprises of two stages: Initialization and Regeneration. The first stage is used to protect the DTW template through a hardening process illustrated in Figure 7.1. The second stage illustrated in Figure 7.2 is used to protect the matching template; we apply a password in the second stage.

7.2.1 Enrollment: Initialization

For the Initialization stage (Figure 7.1), we follow the processes in Section 5.2.1. For a quick overview, the processes are detailed in the following.

To start, a user presents training utterances to the system. Then, the first utterance is used as the initialized hardened template. The algorithm performs DTW to match it to the other training utterances. Next, the results are mapped to binary strings by comparing with the multi-threshold. Next, the bits from a binary string that the speaker can reliably generate are defined and refer to as distinguishing de-

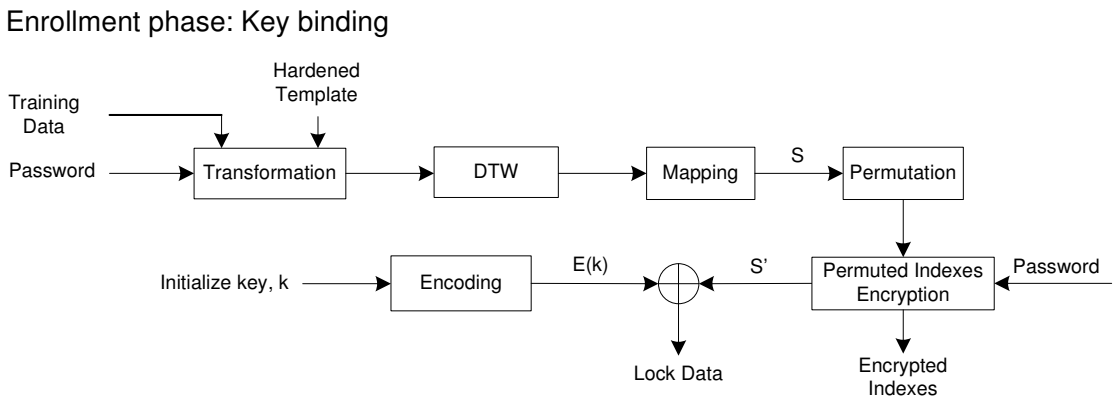


Figure 7.2: Enrollment phase: Regeneration.

scriptor D. The next step is hardening process that is the process to remove a feature from the DTW template. Specifically, let the total number of bit derived from the hardened template that corresponds to the distinguishing descriptors be T; the system should yield T as less than or equal to D/2. In Chapter 5, we showed that, under this condition, the hardened template was secure even if the attacker acquired the templates and had perfect knowledge of correlation of features. For this reason, if T is greater than D/2, one of templates feature vectors will be removed. After each step in hardening the template, the hardened DTW template will be the keying signal of the training pass-phrase and the process will be re-started until the condition is met. Finally, the result is stored as a hardened template in using for the next stage.

7.2.2 Enrollment: Regeneration

This stage (Figure 7.2) consists of three main steps: transformation, permutation, and key binding. Firstly, random numbers derived from the user's password are used to transform the hardened template and training pass-phrases. The transformed

biometrics is then mapped to binary strings. Then, a distinguishing descriptor (a binary string that users can reliably generate) is defined. Secondly, the distinguishing descriptor is encrypted with a password. Finally, the encrypted binary string is used to secure the cryptographic key using fuzzy commitment framework [48]. These steps are detailed in the following.

Transformation

The system first generates two sets of pseudo-random numbers, $S = \{-1, 1\}^{2m}$ where m (the same m in the initialization stage) is the number of frames of the hardened template. Hence, if training biometrics is greater than $2m$ frames, the users will be asked to re-utter their pass-phrases.

Two sets, say S_1 and S_2, are arranged in a two-column table: S_1 in the first column and S_2 in the second column. Then the system uses an eighth-character password to generate a $2m$ bit binary string $R = \{r_i \in \{0, 1\}, i = 1, \ldots, 2m\}$. Next, the R will be used to select the random numbers in the table: select the number in the first column if $r_i = 1$ and otherwise in the second column. Lastly, the selected random numbers will be used to transform feature vectors. More precisely, let $\mathcal{H} = \{f_i, i = 1, \ldots, m\}$ be a set of feature vectors of the hardened template where the vector $f_i = [f_i(1), \ldots, f_i(121)]^T$. The transformed version of $\mathcal{H}$ can be represented by $\mathcal{T} = \{(-1)^{r_i} \cdot f_i + f_i, i = 1, \ldots, m\}$. For a set of training vectors $X = \{x_i, i = 1, \ldots, n \leq 2m\}$ where the vector $x_i = [x_i(1), \ldots, x_i(121)]^T$, the transformed version is $\mathcal{Q} = \{(-1)^{r_i} \cdot x_i + x_i, i = 1, \ldots, n\}$. Using $\mathcal{T}$ as a reference template, the system performs DTW to the $\mathcal{Q}$. The result will be used in mapping and generating distinguishing descriptor the same way as

described in Section 5.2.1, 5.2.2, and 5.2.3. Next, we select 2^n-1 bits, where $n = 3, 4, \ldots$, based on feature variation to form a binary string $S = \{b_i \in \{0, 1\}, i = 1, \ldots, L = 2^n - 1\}$.

Permutation

For the second step, the indexes of the S will be randomly permuted in the context of cryptography. Next, the permuted indexes are used to arrange the binary string S and we refer the result to as an arranged binary string S'. Finally, we employ a prefix cipher [8] with domain and range in $[1, L]$ where $[1, L]$ denote a set of integers from 1 to L in encryption and decryption the permuted indexes with a password $P \in \mathcal{K}$.

The prefix cipher consists of two functions: $E : \mathcal{K} \times [1, L] \rightarrow [1, L]$ and $D : \mathcal{K} \times [1, L] \rightarrow [1, L]$. Therefore, if we refer the encrypted permuted indexes to as $\mathcal{M}$, every possible password when it is used to decrypt $\mathcal{M}$, will yield an integer string that consists of non-repeated random integers in $[1, L]$. By utilizing this scheme, the attackers cannot discriminate the correct password from brute-force search as the decrypted template appears as a random permutation on a subset of the indexes.

Key binding

To combine the arranged binary string S' with cryptographic key is the last step. The system first generate a pseudo-random bit k and then encoded properly denoted by $E(k)$ of length L (see Figure 7.2). In our case, we use BCH code [56]. The encoding code *E(k)* has to tolerate error within Hamming distance (H), a maximum number of bit differences between the distinguishing descriptors and

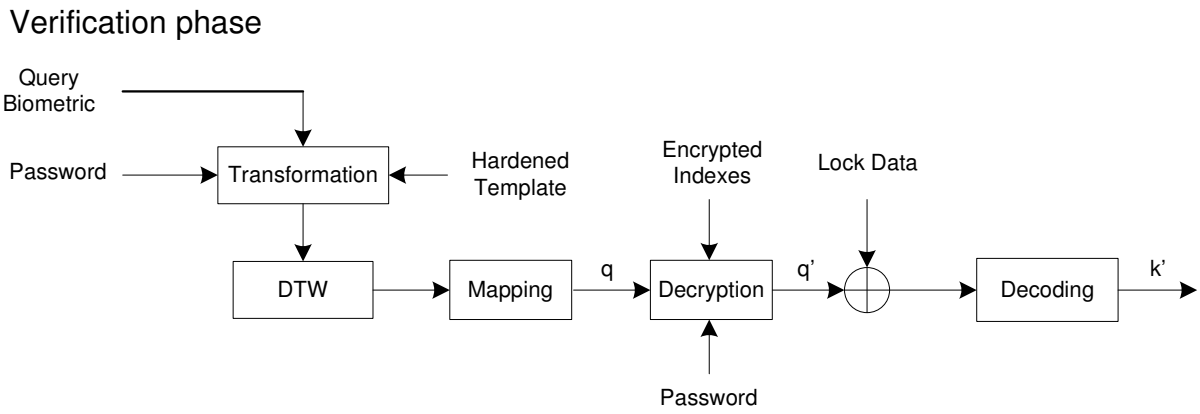


Figure 7.3: Biometric key retrieval in verification phase.

the feature descriptors of a legitimate user. For the next step, the S and the encoding code *E(k)* will be hidden using an XOR operation and then stored as a lock data denoted by $\mathcal{L}$. Only the user with feature descriptors S' that is sufficiently similar to the S within Hamming distance ($|S - S'| \leq H$) can unlock the $\mathcal{L}$ and correctly decode the key.

7.2.3 Verification

The biometric key retrieval process is in the verification phase illustrated in Figure 7.3. The user requests the template from the database that contains the hardened template, the multi-thresholds, and the lock data. A user's password will be used to transform the hardened template and query biometrics the same way in Section 7.2.2. Once the transformed versions are set, the system performs DTW. Then, the result will be mapped to a feature descriptor q. Next, the encrypted permutation indexes $\mathcal{M}$ will be decrypted with the password; the result is used to re-arrange the feature

descriptor q and we refer the re-arranged result to as q'. Then, the q' will be XORed with the lock data. The next step is the decoding process. If the error is within the tolerance, the key can be correctly reconstructed. To check whether the key is identical to the key generated in the training phase, a number of researchers [3, 36, 67] checked the hash function. In the training phase, the initialized key was stored as $h(k)$. Once the key k', is regenerated from the verification phase, the system checks to see whether $h(k) = h(k')$. If $h(k) = h(k'))$, the key, k', is correct. The system authenticates the user.

7.3 Experimental Setup

We compare the SCKRP with other speaker verification systems: Dynamic Time Warping (DTW) [32] detailed in Section 4.3.1 and Dynamic Time Warping-based Biometric key Binding (DBKB) detailed in Chapter 5.

For the SCKRP, the same parameters in the DBKB are utilized except the correctable bits parameter t illustrated in Table 7.1 is varied to an operating point of each scenario. We will evaluate verification performance using Equal Error Rate (EER). However, the dataset does not include users' passwords. In our experiments, we have to select passwords which are likely to be used in the real world application. For this reason, we select eight character users' passwords based on difficulty levels indicated in Figure 7.4 [16]. Six classes of users' passwords and their distribution are: 1) one word (23%), 2) combination of two or more word (6%), 3) familiar numbers, such as a social security number, street address, birth date, etc. (21%), 4) unfamiliar numbers (10%), 5) string of numbers and letter (34%), 6) string of numbers, letter, and

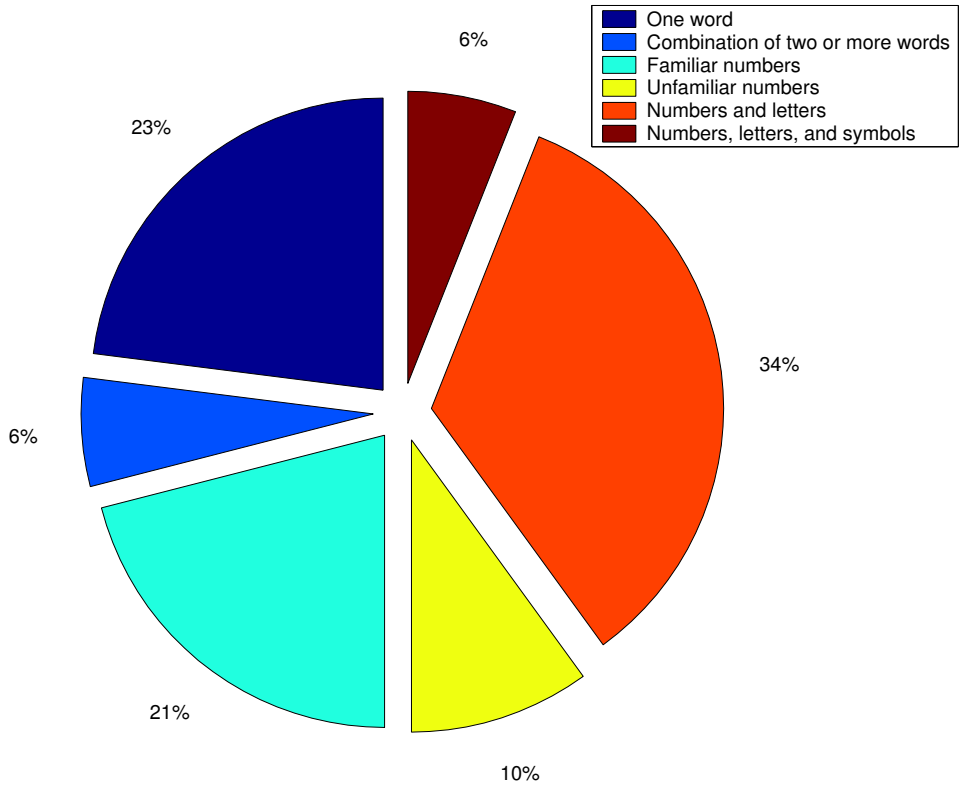


Figure 7.4: Distribution of users' passwords that are comprised of one word, combination of two or more words, unfamiliar numbers, familiar numbers, string of numbers and letters, or string of numbers, letters, and symbols.

symbols (6%).

7.4 Experimental Results

We will first investigate in the case that one of the applied password schemes is excluded (one-layer scheme). Then, we will investigate the two-layer scheme SCKRP (transformation and permutation schemes).

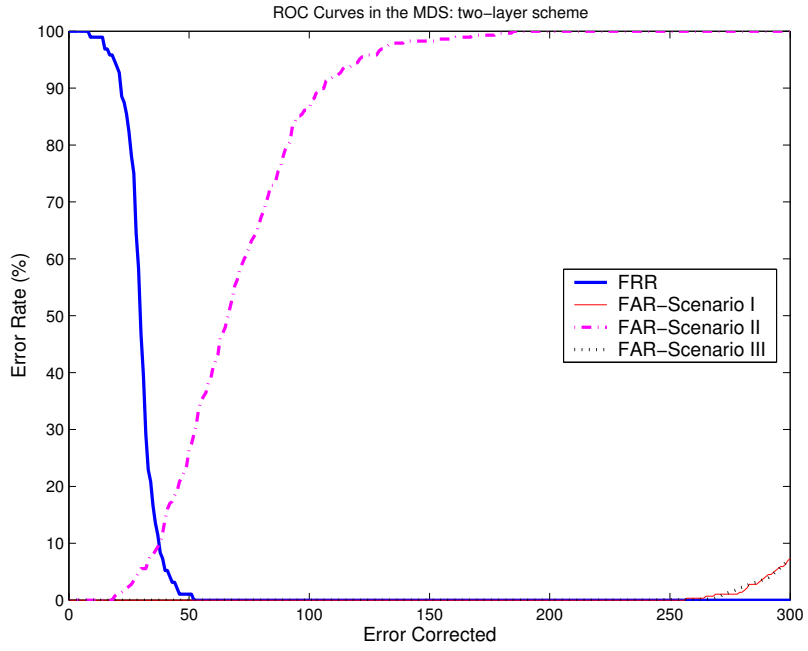
7.4.1 One-layer Scheme

Table 7.1 shows the EERs of the DTW [32], DBKB [39], and SCKRP in the MDS and LDS against imposter attack (H-II). In the case of the permutation scheme only, the error rate of the compromised password scenario (II) does not differ from the DTW and DBKB system. In the other scenarios (I and III), the verification performance meets the same level of the password approach.

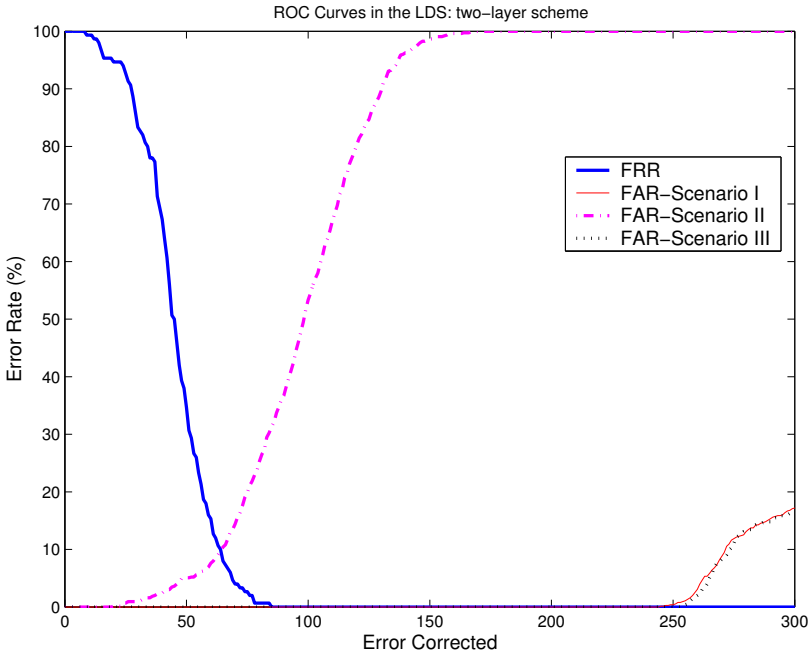
In the case of the transformation scheme only, the error rate of the compromised password scenario (II) also does not differ from other systems, but the error rates in the other scenarios (I and III) are noticeably degraded when we compare them with the previous case. As we introduced, the transformation layer is designed to slow down attackers who try to brute-force search the key. Therefore, it is necessary to keep this layer. In the next section, we will show that we can address this drawback when two schemes are combined.

7.4.2 Two-layer Scheme

Table 7.2 shows the EERs of the two-layer scheme SCKRP in the MDS and LDS against imposter attack (H-II). In the case that the password and biometrics are not compromised (scenario I), the verification performance of the SCKRP clearly outperforms the other systems. For the compromised biometric case (scenario III), the error rate is still the same as scenario I. For the compromised password case (scenario II), the verification performance of the SCKRP does not differ when we compare it with other systems. These results are also illustrated in Figure 7.5.



(a)



(b)

Figure 7.5: ROC curves of two-layer scheme. Scenario I: genuine, Scenario II: compromised password, and Scenario III: compromised biometrics (a) the MDS (b) the LDS

Table 7.1: EERs of speaker verification systems in the MDS and LDS against imposter attack for Dynamic Time Warping-based (DTW), Dynamic Time Warping-based Biometric key Binding (DBKB), and our approach (SCKRP) in the case that one of the applied password layer is excluded. Scenario I: genuine, Scenario II: compromised password, and Scenario III: compromised biometrics

Dataset	Method	Scenario	EER ($\bar{x} \pm Z_{\frac{0.05}{2}} \frac{s}{\sqrt{n_s}}$ %)	Error Corrected
MDS	DTW[32]	I	11.20 ± 0.83	-
	$DBKB$[39]	I	11.54 ± 1.28	38 bits
	$Permuted$ $SCKRP$	I	**0.00**	53 bits
		II	10.69 ± 1.11	38 bits
		III	**0.00**	53 bits
	$Transformed$ $SCKRP$	I	3.61 ± 0.24	46 bits
		II	11.61 ± 1.65	38 bits
		III	9.05 ± 0.62	38 bits
LDS	DTW[32]	I	7.82 ± 1.57	-
	$DBKB$[39]	I	8.96 ± 1.17	38 bits
	$Permuted$ $SCKRP$	I	**0.00**	53 bits
		II	8.40 ± 1.71	38 bits
		III	**0.00**	53 bits
	$Transformed$ $SCKRP$	I	4.43 ± 1.72	46 bits
		II	9.35 ± 2.58	38 bits
		III	7.40 ± 1.12	38 bits

7.4.3 Security Analysis

In this section, we investigate the security of two-layer SCKRP with three scenarios. For the case of the genuine (scenario I), we use the same approach presented in Chapter 5 to estimate the entropy. Hence, the security of the scheme can be estimated using the sphere packing bound $\mathcal{BF}=\frac{2^z}{\sum_{i=1}^{w}\binom{z}{i}}$ where z is the uncertainty of voice and w is the error bits that can be corrected by the system [36]. We carry out 4,512 of inter-speaker comparisons (the same dataset as used in Chapter 5) to evaluate the uncertainty. For a binary string of 511 bits, the uncertainty of our template is 125 bits. From Table 7.2, the system should be able to correct the error up to 39 bits,

Table 7.2: EERs of two-layer SCKRP in the MDS and LDS against imposter attack (H-II). Scenario I: genuine, Scenario II: compromised password, and Scenario III: compromised biometrics

Dataset	Method	Scenario	EER ($\bar{x} \pm Z_{\frac{0.05}{2}} \frac{s}{\sqrt{n_s}}$ %)	Error Corrected
MDS	$Two-layer$ $SCKRP$	I	**0.00**	53 bits
		II	10.72 ± 1.65	39 bits
		III	**0.00**	53 bits
LDS	$Two-layer$ $SCKRP$	I	**0.00**	53 bits
		II	8.24 ± 1.13	39 bits
		III	**0.00**	53 bits

that is approximately 8%. Here, z is 125 bits and w is 10 bits. The estimated entropy is 76 bits, which is much better than the DBKB (51 bits).

For the case of the compromised password (scenario II), the estimated entropy is 77 bits. However, 33.07%, which is determined using the analysis technique proposed in [39], leaks from the hardened template. Therefore, the estimated entropy is 51 bits, which is the same as reported in [39]. Even though 51 bits of entropy can easily be enumerated using today's computational resources, this space is determined under the assumption that an attacker knows every users' password in the system. However, it is very difficult for the attacker to do.

For the case of the compromised biometrics (scenario III), the estimated entropy is between 18-30 bits [14]. However, the attackers have to spend many more attempts for two reasons. First, the SCKRP is a biometric-based system; it prevents the attacker who is content to find the password of any users in the system (the weakest link). More precisely, the attackers randomly try the most probable password with every user in the system and try other passwords until they find the first match. For the SCKRP, they cannot do that as applying the same password to different biometrics

yields different results. Second, the transformation process forces the attackers to run dynamic programming every time they try other different passwords. In contrast, if the transformation process is excluded, they can run dynamic programming only once. Then, they apply passwords to the warping signal and check to see whether the result matches the template. As a result, this case (the transformation layer is excluded) does not differ from the traditional password-based approach.

Overall processes create a greater computational load for an attacker. Even if this also makes users wait more time for authentication, it makes much more time for the attackers as they have to try every possible password.

7.5 Summary

We have proposed a way to combine a speech biometric cryptosystem with a password. The system consists of three layers. For the first layer, the biometrics is transformed using a password. Then, we map the transformed version to a binary string. For the second layer, the result from the second layer is permuted using a password in such a way that the attackers cannot discriminate the correct password from brute-force search if the biometrics is not compromised. For the third layer, a cryptographic key and the binary string are hidden using a fuzzy commitment framework. The experimental results show that the verification performance of the system meets the same level of a traditional password-based approach if biometrics and password are not compromised simultaneously. Furthermore, the system increases the computational time for attackers to search for the key. Even if the attackers acquire the biometrics, they have been forced to align query biometrics each time they guess the password.

In further research, we plan to investigate the impact of user passwords on the system.

Chapter 8

Conclusion and Future Work

8.1 Conclusions

The DTW, VQ, and GMM techniques have been used in speaker verification systems. The DTW was the first technique developed. Even though it yields good recognition performance, researchers are concerned about the security of the template because the system stores a full set of feature vectors. After that, VQ and GMM were developed to address this problem and have been popular since then.

In this thesis, we have investigated the security of the aforementioned methods against various attacks. The attacks included the traditional attack reported in the literature (human imposters), the more sophisticated attack (an informed adversary utilizing synthetic pass-phrases), and the algorithm we developed (an informed adversary utilizing biometric templates). We have shown that these attacks were devastating to the biometric systems. In particular, the most effective was the biometric template attack. Then, we have demonstrated that the traditional approach to eval-

uate the security of speech biometric user verification was insufficient. The results indicated that the FARs of the other attack models beyond the traditional approach were significantly high. We also investigated the results based on gender information. There were no significant differences.

We developed the cryptographic-based speaker verification to protect the biometric template. We utilized Dynamic Time Warping (DTW) in our system. We presented a hardened template which was useful for creating a warping function, but it was not usable for an attacker to derive the cryptographic key. We have shown that the recognition performance of the hardened template was not far from the unprotected template. In addition, the EER against the attackers utilizing the hardened template to generate a cryptographic key was 0%. The results were also compared with the transformation approach where the one-way function was used to protect the template. The experimental results showed that our approach performed better than the transformation approach. In addition, we showed that even if a template was stored as a transformed version, the attacker could gain access to the system which did not differ from an unprotected template.

We addressed the problem of feature correlation which reduced the security of the biometric template by proposing a multi-thresholds scheme. As a result, the randomness of the key (entropy) was increased from 16 to 51 bits. Lastly, we showed that the EER of the proposed system against the template attack was significantly lower than the other methods.

Finally, we used a password to protect stored templates, enhance security, and reduce error rates in biometric cryptosystems. We addressed the problem of the security of the system dropping to the same level as that of a password approach

if the biometrics is compromised. We have shown that the EER of our scheme was the same as a traditional password-based approach even when the biometrics was compromised but the scheme increased the computational time for attackers to search for the key. Even if the attackers acquire the biometrics, they have been forced to align the biometrics each time they guess the password.

Through careful study of biometrics and their flaws, we believe that our work can ultimately lead to stronger, more usable biometric security. We proposed a new algorithmic attack based on template information to demonstrate that the traditional approach to evaluate the security of speech biometric user verification was insufficient. Then, we developed the cryptographic-based speaker verification to protect the biometric templates. Lastly, we used a password to protect stored templates, enhance security, and reduce error rates in biometric cryptosystems.

8.2 Future Work

Future work is suggested as follows.

1. Even though the error rates of the proposed system against synthetic speech are the lowest when we compare our scheme with the other systems, it is still high when compared with other attack models. Hence, the design of a speaker verification system should have the ability to discriminate between synthetic and real pass-phrases. This issue is related to the construction of speech synthesizers. For example, with HMM-based synthetic pass-phrases, the fact that the synthesizer always produces the same optimal waveform in terms of likelihood score should be used as another feature in a speaker verification system. If the

likelihood score is greater than the threshold, the speaker verification system will reject the pass-phrase. For other speech synthesizers, we believe that there are similar likelihood-score features. It is also possible that a challenge-response scheme or some kind of dialog-based scheme might work. The natural variation that arises when a human says the same phrase twice might be hard to duplicate in some synthesis methods. In further research, these features should be identified to improve the recognition performance of a speaker verification system.

2. By assigning an appropriate tuning parameter to the proposed generative model, we have shown that it offered great potential in gaining access to the systems. In this work, we have set a tuning parameter as a global threshold. In further research, we should focus on investigation of a local threshold, one per a target user.

3. We have shown that our techniques offer great potential to protect the speech biometric template with slightly degraded recognition performance in the case of a compromised password. In further research, we should investigate security and performance of other behavioral modalities which have temporal features, such as a signature or handwriting.

4. The permutation technique we proposed in Chapter 7 is a general technique which we believe can be used in other biometric modalities for the key binding scheme. In further research, we should investigate security and performance of our technique for physiological biometrics, such as fingerprints, faces, and iris codes.

Appendix A
Datasets

Appendix A.1

A list of pass-phrases in the MDS, F = Female and M = Male

Enrolled subject	Pass-phrase		Enrolled subject	Pass-phrase
F00	mint chocolate chip		F11	mint chocolate chip
F01	pralines and cream		F12	chunky monkey
F02	chocolate fudge		F13	chocolate fudge
F03	pralines and cream		F14	peppermint stick
F04	pralines and cream		F15	mint chocolate chip
F05	pralines and cream		F16	chunky monkey
F06	mint chocolate chip		F17	chunky monkey
F07	chunky monkey		F18	pralines and cream
F08	chocolate fudge		F19	chunky monkey
F09	chocolate fudge		F20	chocolate fudge
F10	peppermint stick		F21	mint chocolate chip

A list of pass-phrases in the MDS (continue), F = Female and M = Male

Enrolled subject	Pass-phrase		Enrolled subject	Pass-phrase
M00	peppermint stick		M13	rocky road
M01	rocky road		M14	rocky road
M02	peppermint stick		M15	peppermint stick
M03	chocolate fudge		M16	rocky road
M04	pralines and cream		M17	chocolate fudge
M05	mint chocolate chip		M18	pralines and cream
M06	chunky monkey		M19	chunky monkey
M07	peppermint stick		M20	rocky road
M08	rocky road		M21	rocky road
M09	peppermint stick		M22	peppermint stick
M10	chocolate fudge		M23	peppermint stick
M11	pralines and cream		M24	chunky monkey
M12	mint chocolate chip		M25	mint chocolate chip

Appendix A.2

A list of dedicated users' pass-phrases in the MDS, F = Female, M = Male, and i = Dedicated imposter

Enrolled subject	Pass-phrase		Enrolled subject	Pass-phrase
F00i	chocolate fudge		M03i	rocky road
F01i	chocolate fudge		M04i	rocky road
F02i	chocolate fudge		M05i	pralines and cream
F03i	chocolate fudge		M06i	rocky road
F04i	chunky monkey		M07i	pralines and cream
F05i	chunky monkey		M08i	rocky road
F06i	mint chocolate chip		M09i	peppermint stick
F07i	mint chocolate chip		M10i	pralines and cream
F08i	chocolate fudge		M11i	mint chocolate chip
F09i	mint chocolate chip		M12i	peppermint stick
F10i	mint chocolate chip		M13i	pralines and cream
F11i	mint chocolate chip		M14i	mint chocolate chip
F12i	chunky monkey		M15i	peppermint stick
F13i	chunky monkey		M16i	chunky monkey
F14i	peppermint stick		M17i	mint chocolate chip
F15i	pralines and cream		M18i	peppermint stick
F16i	pralines and cream		M19i	chunky monkey
M00i	rocky road		M20i	mint chocolate chip
M01i	pralines and cream		M21i	mint chocolate chip
M02i	rocky road		M22i	chocolate fudge

Appendix A.3

A list of pass-phrases in the LDS

Enrolled subject		Pass-phrase
subject1		If you can't beat them, join them.
		When it rains, it pours.
		The person who has no opinion will seldom be wrong.
		I kept getting a busy signal.
		I'd like to reserve a table for dinner.
subject2		The first step is always the hardest.
		Blood is thicker than water.
		The secret of being tiresome is to tell everything.
		We seem to have a bad connection on this phone.
		Do you have an apartment available?
subject3		There's no place like home.
		A fool and his money are easily parted.
		Only the suppressed word is dangerous.
		Would you care to leave a message?
		We could do it first thing tomorrow morning.
subject4		If you can't beat them, join them.
		When it rains, it pours.
		The person who has no opinion will seldom be wrong.
		I kept getting a busy signal.
		I'd like to reserve a table for dinner.
subject5		You have to take the good with the bad.
		All in the same boat.
		The important thing is never to stop questioning.
		I want it to be very, very lean.
		I hope there's nothing serious.
subject6		Absence makes the heart grow fonder.
		A taste of your medicine.
		Doubt is not a pleasant condition, but certainty is absurd.
		Let me get back to you in a few minutes.
		I left the keys in the car.

Appendix A.4

A list of phrases to build the speech corpus in the LDS

	Pass-phrase
001	Gad, do I remember it.
002	You got out by fighting, and I through a pretty girl.
003	I can see that knife now.
004	When I can't see beauty in woman I want to die.
005	His slim fingers closed like steel about Philip's.
006	He seized Gregson by the arm and led him to the door.
007	Hear the Indian dogs wailing down at Churchill.
008	I'd say there was going to be a glorious scrap.
009	He turned the map to Gregson, pointing with his finger.
010	His eyes never took themselves for an instant from his companion's face.
011	Lakes and rivers, hundreds of them, thousands of them.
012	Whitefish, Gregson, whitefish and trout.
013	They robbed me a few years later.
014	He chuckled as he pulled out his pipe and began filling it.
015	Everything was working smoothly, better than I had expected.
016	I was completely lost in my work.
017	His slim hands gripped the edges of the table.
018	Philip dropped back into his chair.
019	If I was out of the game it would be easily made.
020	It is growing, every day, every hour.
021	Now, you understand.
022	You have associated with some of these men.
023	All operations have been carried on from Montreal and Toronto.
024	Gregson held a lighted match until it burnt his fingertips.
025	Gregson had seated himself under the lamp and sharpening a pencil.
026	He caught himself with a jerk.
027	How does your wager look now.
028	He confessed that the sketch had startled him.
029	After all, the picture was only a resemblance.
030	Philip knew that she was not an Indian.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
031	In her haste to get away she had forgotten these things.
032	Philip took a step toward Gregson, half determined to awaken him.
033	But if Pierre did not return, until tomorrow.
034	Ten minutes had not elapsed since he had dropped the handkerchief.
035	It won't be for sale.
036	For a few moments he ate in silence.
037	Philip did not pursue the subject.
038	Philip produced a couple of cigars and took a chair opposite him.
039	Suppose you saw me at work through the window.
040	He looked like one who had passed through an uncomfortable hour.
041	There was nothing more, except a large ink blot under the words.
042	All this day Gregson remained in the cabin.
043	The sixth day he spent in the cabin with Gregson.
044	The flush was gone from her face.
045	That is why I am, am rattled, he laughed.
046	He understood the meaning of the look.
047	She was even more beautiful than when I saw her, before.
048	I'll give a thousand if you produce her, retorted Gregson.
049	They have won popular sentiment through the newspapers.
050	We must achieve our own salvation.
051	In moments of mental energy Philip was restless.
052	He would keep his faith with Gregson for the promised day or two.
053	Something about it seemed to fascinate him, to challenge his presence.
054	Now it was missing from the wall.
055	He boiled himself some coffee and sat down to wait.
056	I'm going down there with you, and I'm going to fight.
057	Now have you got anything to say against me, Mr Philip.
058	If I meet her again I shall apologize, said Eileen.
059	Below him the shadow was broken into a pool of rippling starlight.
060	Only the chance sound had led him to observe them.
061	Could the incident have anything to do with Jeanne and Pierre.
062	There was no chance to fire without hitting him.
063	There was no answer from the other side.
064	Then he hastened on, as Pierre had guided him.
065	With these arguments he convinced himself that he should go on alone.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
066	Yet, behind them there was another and more powerful motive.
067	In that case he could not miss them, if he used caution.
068	Before he could recover himself Jeanne's startled guards were upon him.
069	It is the nearest refuge.
070	There was pride and strength, the ring of triumph in his voice.
071	Tomorrow it will be strong enough for you to stand upon.
072	You were going to leave after you saw me on the rock.
073	He bit his tongue, and cursed himself at this fresh break.
074	In it there was something that was almost tragedy.
075	Your face is red with blood.
076	Her eyes smiled truth at him as he came up the bank.
077	He can care for himself.
078	They will search for us between their camp and Churchill.
079	Until I die, he exclaimed.
080	Her beautiful hair was done up in shining coils.
081	The Churchill narrowed and its current became swifter as they progressed.
082	For a full half minute Jeanne looked at him without speaking.
083	I want to die in it.
084	Darkness hid him from Jeanne.
085	And yet if she came he had no words to say.
086	He heard a sound which brought him quickly into consciousness of day.
087	Within himself he called it no longer his own.
088	Besides, that noise makes me deaf.
089	Philip looked back from the crest and saw Jeanne leaning over the canoe.
090	Fifty yards ahead of her were the first of the rocks.
091	There was one chance, and only one, of saving Jeanne.
092	You're a devil for fighting, and will surely win.
093	I'll only be in the way.
094	He lifted his eyes, and a strange cry burst from his lips.
095	Shooting pains passed like flashes of electricity through his body.
096	I know that you are in charge there, and Jeanne knows.
097	For a full minute the two men stared into each other's face.
098	He was sure, now, of but few things.
099	It was a miracle, and I owe you my life.
100	Philip ate lightly of the food which Pierre had ready for him.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
101	Such men believe, when they come together.
102	The journey was continued at dawn.
103	Jeanne and Pierre both gazed toward the great rock.
104	There was something pathetic in the girl's attitude now.
105	He moved his position, and the illusion was gone.
106	For two hours not a word passed between them.
107	I have hunted along this ridge, replied Philip.
108	We saw your light, and thought you wouldn't mind a call.
109	Billinger may arrive in time.
110	I want my men to work by themselves.
111	He destroyed everything that had belonged to the woman.
112	Philip bent low over Pierre.
113	She saw the answer in his face.
114	There is no need of further detail, now – for you can understand.
115	There followed a roar that shook the earth.
116	Blind with rage, he darted in.
117	In it was the joy of life.
118	Swiftly his eyes measured the situation.
119	But this little defect did not worry him.
120	And then, steadily, he began to chew.
121	Together they ate the rabbit.
122	They edged nearer, and stood shoulder to shoulder facing their world.
123	It was beating and waiting in the ambush of those black pits.
124	Something vastly more thrilling had come into it now.
125	It took him half an hour to reach the edge of it.
126	But there was no longer the mother yearning in his heart.
127	Besides, had he not whipped the big owl in the forest.
128	After all, it was simply a mistake in judgment.
129	Had it struck squarely it would have killed him.
130	The Indian even poked his stick into the thick ground spruce.
131	Pebbles and dirt flew along with hair and fur.
132	And he was filled with a strange and foreboding fear.
133	It was steel, a fisher trap.
134	OW, a wild dog, he growled.
135	That is the strange part of it.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
136	His freshly caught furs he flung to the floor.
137	In the crib the baby sat up and began to prattle.
138	She obeyed, shrinking back with the baby in her arms.
139	His teeth shut with a last click.
140	It was over when he made his way through the ring of spectators.
141	In a flash he was on his feet, facing him.
142	He thought he saw a shudder pass through the Factor's shoulders.
143	The moon had already begun its westward decline.
144	They laughed like two happy children.
145	He pulled, and the log crashed down to break his back.
146	Fast, but endure.
147	A little before dawn of the day following, the fire relief came.
148	The Indian felt the worship of her warm in his heart.
149	He drew in a deep breath as he looked at them.
150	Then he shouted, Shut up.
151	He changed his seat for a steamer reclining chair.
152	To these he gave castor oil.
153	Hatred and murder and lust for revenge they possessed to overflowing.
154	Sheldon glanced at the thermometer.
155	Also, I want information.
156	Let them go out and eat with my boys.
157	I, I beg pardon, he drawled.
158	And you preferred a cannibal isle and a cartridge belt.
159	I was in New York when the crash came.
160	No, I did not fall among thieves.
161	Such things in her brain were like so many oaths on her lips.
162	Your being wrecked here has been a godsend to me.
163	I can't go elsewhere, by your own account.
164	Her achievements with coconuts were a revelation.
165	He glanced down at her helplessly, and moistened his lips.
166	That is what distinguishes all of us from the lower animals.
167	He also contended that better confidence was established.
168	Outsiders are allowed five minute speeches, the sick man urged.
169	So was Packard's finish suicide.
170	Joan cried, with shining eyes.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
171	Nobody knows how the natives got them.
172	How can you manage all alone, Mr Young.
173	The planters are already considering the matter.
174	I use great trouble advisedly.
175	Dear Sir, Your second victim has fallen on schedule time.
176	We leave the eventuality to time and law.
177	Similar branch organizations have made their appearance in Europe.
178	Society is shaken to its foundations.
179	A month in Australia would finish me.
180	Down through the perfume weighted air fluttered the snowy fluffs.
181	You were destroying my life.
182	Horses and rifles had been her toys, camp and trail her nursery.
183	I'm as good as a man, she urged.
184	You read the quotations in today's paper.
185	He's terribly touchy about his black wards, as he calls them.
186	Whatever he guessed he locked away in the taboo room of Naomi.
187	This is eighteen eighty.
188	Death is and has been ever since old Maui died.
189	Let us talk it over and find a way out.
190	It is a good property, and worth more than that.
191	I wish you were more adaptable, Joan retorted.
192	Such is my passage engaged on the steamer.
193	The issue was not in doubt.
194	Well, there are better men in Hawaii, that's all.
195	Harry Bancroft, Dave lied.
196	It's a Yankee, Joan cried.
197	He was the leader, and Tudor was his lieutenant.
198	They likewise are disinclined to being eaten.
199	But to culture the Revolution thus far had exhausted the Junta.
200	The President of the United States was his friend.
201	Your face was the personification of duplicity.
202	Shorty turned to their employers.
203	You were engaged.
204	I saw it all myself, and it was splendid.
205	Now run along, and tell them to hurry.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
206	What's that grub-thief got to do with it.
207	It was a superb picture.
208	So she said, the irate skipper dashed on.
209	And watch out for wet feet, was his parting advice.
210	They just lay off in the bush and plugged away.
211	The very thought of the effort to swim over was nauseating.
212	And there was a dog that barked.
213	There are four, all low, McCoy answered.
214	The women they carried away with them to the Big Valley.
215	The Japanese understood as we could never school ourselves.
216	They had been on the same lay as ourselves.
217	The boy grew and prospered.
218	He wanted to give the finish to this foe already so far gone.
219	Exciting times are the lot of the fish patrol.
220	I know they are my oysters.
221	By this time Charley was as enraged as the Greek.
222	They must have been swept away by the chaotic currents.
223	It resembled tea less than lager beer resembles champagne.
224	At the same time spears and arrows began to fall among the invaders.
225	Then, again, Tudor had such an irritating way about him.
226	Outwardly, he maintained a calm and smiling aspect.
227	You fired me out of your house, in short.
228	Her mouth opened, but instead of speaking she drew a long sigh.
229	It's worth eight dollars.
230	And he did hurt my arm.
231	Only once did I confide the strangeness of it all to another.
232	I was not to cry out in the face of fear.
233	And now put yourself in my place for a moment.
234	The boy threw back his head with pride.
235	Why not like any railroad station or ferry depot.
236	We could throw stones with our feet.
237	These were merely stout sticks an inch or so in diameter.
238	Then it was that a strange thing happened.
239	From the source of light a harsh voice said.
240	But I did not enjoy it long.

A list of phrases to build the speech corpus in the LDS (continue)

	Pass-phrase
241	We were now good friends.
242	Two of the Folk were already up.
243	He gave one last snarl and slid from view among the trees.
244	Again the girls applauded, and Mrs Hall cried.
245	Just the same I'd sooner be myself than have book indigestion.
246	Some of the smaller veins had doubtless been ruptured.
247	But we were without this momentum.
248	There was one difficulty, however.
249	The hyena proceeded to dine.
250	Or have they already devised one.
251	We would not spend another such night.
252	At first his progress was slow and erratic.
253	He placed his paw on one, and its movements were accelerated.
254	The awe of man rushed over him again.
255	The Fire-Men wore animal skins around their waists.
256	Between him and all domestic animals there must be no hostilities.
257	All right, Sir, replied Jock with great regret.
258	Why should a fellow throw up the sponge after the first round.
259	His hand shot out and clutched Crooked-Leg by the neck.
260	Does the old boy often go off at half-cock that way.
261	A flying arrow passed between us.
262	I pulled, suddenly, with all my might.
263	Here we allow our solicitors to look after our legal work.
264	His previous wives had never lived long enough to bear him children.
265	It was our river emerging like ourselves from the great swamp.
266	Cameron looked at his hands with their long, sinewy fingers.
267	We got few vegetables and fruits, and became fish eaters.
268	We never made another migration.
269	Nor was Elam Harnish an exception.
270	A little treatment, massage, with some help from the doctor.

Appendix B
List of passwords in the experiments

Appendix B.1

Passwords in the MDS

Enrolled subject	Password
F00	commands
F01	hatching
F02	jeopardy
F03	kangaroo
F04	metaphor
F05	obligate
F06	offender
F07	quainter
F08	relegate
F09	scooters
F10	tendency
F11	getmoney
F12	whatisit
F13	amIright
F14	35463478
F15	34345434

Passwords in the MDS (continue)

Enrolled subject	Password
F16	32435698
F17	74357023
F18	45367584
F19	12091974
F20	41118015
F21	19865123
M00	50820122
M01	61041912
M02	48427405
M03	09281987
M04	36100008
M05	50218015
M06	72382128
M07	Peteson1
M08	Chevron9
M09	4debby06
M10	john1954
M11	411Webst
M12	Brown311
M13	born1974
M14	3brother
M15	mike3son
M16	access97
M17	where2go
M18	make2002
M19	oikee013
M20	July1234
M21	Honday10
M22	one2tree
M23	@411home
M24	t#197412
M25	fo!%345

Appendix B.2

Passwords in the LDS

Enrolled subject	Pass-phrase	Password
1	1	commands
	2	hatching
	3	jeopardy
	4	kangaroo
	5	metaphor
2	1	obligate
	2	offender
	3	getmoney
	4	whatisit
	5	35463478
3	1	74357023
	2	45367584
	3	34345434
	4	12091974
	5	35463478
4	1	34345434
	2	34345434
	3	74357023
	4	john1954
	5	411Webst
5	1	Brown311
	2	born1974
	3	3brother
	4	mike3son
	5	access97
6	1	where2go
	2	make2002
	3	4debby06
	4	@411home
	5	t#197412

List of Abbreviations

A-I	Algorithmic Type with Assumption I	49
A-II	Algorithmic Type with Assumption II	50
AES	Advanced Encryption Standard	67
BCH	Bose and Ray-Chaudhuri	67
DBKB	Dynamic Time Warping-based Biometric Key Binding	66
DFT	Discrete Fourier Transform	29
DP	Dynamic Programming	38
DTW	Dynamic Time Warping	37
EER	Equal Error Rate	44
EM	Expectation Maximization	40
FAR	False Acceptance Rate	44
FRR	False Rejection Rate	44
GMM	Gaussian Mixture Model	40
H-I	Human Type with Assumption I	48
H-II	Human Type with Assumption II	49
H-III	Human Type with Assumption III	49
HMM	Hidden Makov Model	19
IDFT	Inverse Discrete Fourier Transform	29
LDS	Lehigh Quiet Environment Speaker Verification Dataset	44
LPC	Linear Predictive Coding	30
LPCC	Linear Predictive Coding Coefficients	31
LSF	Line Spectral Frequency	19
MFCC	Mel-Frequency Ceptrum Coefficients	32
MDS	MIT Mobile Device Speaker Verification Corpus Dataset	44

NHD	Normalized Hamming Distance	82
RMS	Root Mean Square	67
SBS	Sequential Backward Search	68
SCKRP	Speech Cryptographic Key Regeneration based on Password	67
SVM	Support Vector Machine	21
UMB	Universal Background Model	74
VQ	Vector Quantization	38

List of Notations

$s_a(t)$	Analogue-time speech signal	26
s[n]	Discrete-time speech signal	27
$H[z]$	Transfer function in z-domain	27
$x[n]$	Discrete-time speech signal after multiplied with window function	27
$w[n]$	Discrete-time window function	27
$X[k]$	Discrete-time fourier transform of $x[n]$	29
$u[n]$	Unit step sequence	30
$S[z]$	z-transform of Discrete-time speech signal	30
$U[z]$	z-transform of $u[n]$	30
$R[p]$	Autocorrelation sequences	31
c_{LPCC}	Linear Predictive Cepstral Coefficient	31
H_m	Frequency response of the m^{th} filter of the filter bank	32
$S[m]$	Log-energy output of H_m	32
c_{MFCC}	Mel-Frequency Cepstrum Coefficient	32
E	Short-term energy	34
$\mathcal{P}$	Short-term power	34
ξ	Similarity score	35
θ	Decision threshold	35
$\mathcal{C}$	Codebook	40
λ	GMM speaker model	40
π	Weight function	40
μ	Mean vector	40
Σ	Covariance matrix	40
L_{GMM}	Log-likelyhood of GMM	40
σ	Standard deviation	46
κ	Tuning parameter	52
γ	Pre-emphasis parameter	55
α	Confidence coefficient	55
s	Standard deviation of the sample	55

$\mathcal{H}_\mathcal{T}$	Hardened template	64
$\mathcal{H}$	Hardened vector	64
$\mathcal{L}$	Lock data	64
k	Pseudo-random key	67
$E(k)$	Encoded k	67
H	Hamming distance	67
β	Biometric sample	68
Ω	Initialized threshold	68
$\mathcal{T}$	Multi-threshold template	70
ψ	List of DFT indexes	70
B	Distinguishing descriptor	70
Ψ	Relevant indexes	70
f	Warped signal	70
ϕ	Feature vector	70
b	Feature descriptor	71
p	Pseudo-random bits	71
$h(k)$	Hash function of key k	74
k'	Decoded key	74
$\mathcal{BF}$	Sphere packing bound	82
z	Uncertainty of voice	82
w	Number of error bits	82
$\mathcal{D}(\mathcal{A}, \mathcal{B})$	Hamming distance between template A and B	83
$f(x)$	A fractional function	83
$_{\mathcal{RF}}$	The original template	89
$\mathcal{TF}$	The transformed template	89
$\mathcal{TP}$	A password-based transformed template	108
$\mathcal{Q}$	Password-based transformed input signal	108
P	User's password space	109
$\mathcal{K}$	Key space	109
$E(\cdot)$	An encryption function	109
$D(\cdot)$	A decryption function	109
$\mathcal{M}$	Encrypted permuted indexes	109
q	Query feature descriptor	110
q'	Re-arranged query feature descriptor	111

Bibliography

[1] A. Arakala, J. Jeffers, and K. J. Horadam. Fuzzy Extractors for Minutiae-based Fingerprint Authentication. In *Proceedings of Second International Conference on Biometrics*, pages 760-769, Seoul, South Korea, August 2007.

[2] L. Ballard. *Robust Technique to Evaluate the Security of Biometric Cryptographic Key Generators*. PhD thesis, The Johns Hopkins University, Baltimore, Maryland, March 2008.

[3] L. Ballard, S. Kamara, F. Monrose, and M. K. Reiter. Towards Practical Biometric Key Generation with Randomized Biometric Templates. In *Proceedings of 15th ACM Conference on Computer and Communications Security*, pages 235-244, Alexandria, VA, October 2008.

[4] L. Ballard, S. Kamara, and M. K. Reiter. The Practical Subtleties of Biometric Key Generation. In *Proceedings of the 17th Annual USENIX Security Symposium*, pages 61-74, San Jose, CA, August 2008.

[5] L. Ballard, D. Lopresti, and F. Monrose. Forgery Quality and Its Implications for Behavioral Biometric Security. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics (Special Edition)*, 37(5):1107-1118, October 2007.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(7):711-720, 1997.

[7] S. M. Bellovin and M. Merritt. Encrypted Key Exchange: Password-based Protocals Secure Against Dictionary Attacks. In *Proceedings of the 1992 IEEE Symposium on Security and Privacy*, pages 72-84, Washington DC, USA, 1992.

[8] J. Black and P. Rogaway. Ciphers with Arbitrary Finite Domains. In *Proceedings of the Cryptographer's Track at the RSA Conference on Topics in Cryptology*, pages 114-130, Springer-Verlag, 2002.

[9] L. Blum, M. Blum, and M. Shub. Comparision of Two Pseudo-Random Number Generators. In *R. L. Rivest, A. Sherman, and D. Chaum. Proc. Crypto'82*, pages 61-78, New York, 1983. Plenum Press.

[10] T. E. Boult, W. J. Scheirer, and R. Woodworth. Fingerprint Revocable Biotokens: Accuracy and Security Analysis. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1-8, Minneapolis, MN, June 2007.

[11] X. Boyen, Reusable Cryptographic Fuzzy Extractors. In *ACM Conference on Computer and Communications Security*, pages 82-91, Washington D.C, USA, October 2004.

[12] I. R. Buhan, J. M. Doumen, P. H. Hartel, and R. N. J. Veldhuis. Fuzzy Extractors for Continuous Distributions. In *Proceedings of ACM Symposium on Information, Computer and Communications Security*, pages 353-355, Singapore, March 2007.

[13] I. R. Buhan, J. M. Doumen, P. H. Hartel, and R. N. J. Veldhuis. Secure Ad-hoc Pairing with Biometrics: SAfE. In *Proceedings of First International Workshop on Security for Spontaneous Interaction*, page 450-456, Innsbruck, Austria, September 2007.

[14] W. E. Burr, D. F. Dodson, and W. T. Polk. Information Security: Electronic Authentication Guideline. *NIST Special Report 800-63*, April 2006.

[15] J. P. Campbell. Speaker Recognition: a Tutorial. In *Proceedings of the IEEE*. Vol.85 No.9, pages 1437-1426, September 1997.

[16] D. S. Carstens, P. R. McCauley-Bell, L. C. Malone, and R. F. DeMara. Evaluation of the human impact of password authentication practices on information security. *Informing science journal*, 7:67-85, 2004.

[17] Y. -J. Chang, W. Zhang, and T. Chen. Biometrics-based Cryptographic Key Generation. In *Proceedings of IEEE Conference on Multimedia and Expo*, vol. 3, pages 2203-22-6, Taipei, Taiwan, June 2004.

[18] E. C. Chang and S. Roy. Robust Extraction of Secret Bits From Minutiae. In *Proceedings of Second International Conference on Biometrics*, pages 750-759, Seoul, South Korea, August 2007.

[19] C. S. Chin, A. B. J. Teoh, and D. C. L. Ngo. High Security Iris Verification System Based On Random Secret Integration. *Computer Vision and Image Understanding*, 102(2): 169-177, May 2006.

[20] Y. Chung, D. Moon, S. Lee, S. Jung, T. Kim, and D. Ahn. Automatic Alignment of Fingerprint Features for Fuzzy Fingerprint Vault. In *Proceedings of Conference on Information Security and Cryptology*, pages 358-369, Beijing, China, December 2005.

[21] T. Clancy, D. Lin, and N. Kiyavash. Secure Smartcard-based Fingerprint Authentication. In *Proceedings of ACM SIGMM Workshop on Biometric Methods and Applications*, pages 45-52, Berkley, USA, November 2003.

[22] J. Daugman. The Important of Being Random: Statistical Principles of Iris Recognition. *Pattern Recognition*, 36(2): 279-291, 2003.

[23] G. I. Davida, Y. Frankel, and B. J. Matt. On Enabling Secure Applications through Off-line Biometric Identification. In *Proceedings of the 1998 IEEE Symposium on Security and Privacy*, pages 148-157, May 1998.

[24] P. L. De Leon, V. R. Apsingekar, and J. Yamagishi. Revisiting the Security of Speaker Verification Systems against Imposture Using Synthetic Speech. In *International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1798-1801, Dallas TX, USA, 2010.

[25] J.R. Deller, Jr., J. H. L. Hansen, and J. G. Proakis. *Discrete-Time Processing of Speech Signals*. Macmilland Pub. Co., New York, 1993.

[26] A. Demster, N. Lair, and D. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Roy. Statist. Soc.*, Vol. 39, pages 1-38, 1977.

[27] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. In *Cryptology ePrint Archive*, Tech. Rep. 235, February 2006.

[28] S. C. Draper, A. Khisti, E. Martinian, A. Vetro, and J. S. Yedidia. Using Distributed Source Coding to Secure Fingerprint Biometrics. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 2, pages 129-132, Hawaii, USA, April 2007.

[29] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. New York: Wiley, 2001.

[30] D. Feldmeier and P. Karn. UNIX Password Security-Ten Years Later. In *Advances in Cryptology-CRYPTO'89*, pages 44-63. Springer Verlag, London, UK, 1989

[31] S. Furui. *Digital Speech Processing, Synthesis and Recognition*. Marcel Dekker, Inc., New York, 2001.

[32] S. Furui. Ceptral Analysis Technique for Automatic Speaker Verification. *IEEE Transactions on Acoustics, Speech, Signal Processing*, ASSP-29(2): 254-272, April 1981.

[33] L. P. Garcia-Perera, J. C. Mex-Perera, and J. A. Nolazco-Flores. Multi-speaker Voice Cryptographic Key Generation. In *the 3rd ACS/IEEE International Conference on Computer System and Application*. Page 93-98, 2005

[34] F. Hao. *On Using Fuzzy Data in Security Mechanisms*. PhD thesis, University of Cambridge, April 2007.

[35] F. Hao, C. W. Chan. Private Key Generation from On-line Handwritten Signatures. *Information Management & Computer Security*, Issue 10, No. 2, pages 159-164, 2002

[36] F. Hao, R. Anderson, and J. Daugman. Combining Cryptography with Biometrics Effectively. *IEEE Transactions on Computer*, 55(9):1081-1088, September 2006.

[37] A. Higgins, L. Bahler, and J. Porter, Speaker Verification Using Randomized Phrase Prompting. *Digital Signal Processing*, 1(2): 89-106, 1991.

[38] X. Huang, A. Acero, and H. Hon. *Spoken Language Processing*. Prentice Hall, New Jersey, 2001.

[39] K. Inthavisas and D. Lopresti. Speech Biometric Mapping for Key Binding Cryptosystem. In *Biometric Technology for Human Identification VIII (SPIE Defense, Security, and Sensing)*, pages 80291P-1 - 80291P-12, Orlando, FL, April 2011.

[40] K. Inthavisas and D. Lopresti. Attacks on Speech Biometric Authentication. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pages 310-316, Las Vecas, NV, July 2011.

[41] K. Inthavisas and D. Lopresti. Biometric Template Protection for Dynamic Time Warping-based User Authentication. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pages 303-309, Las Vecas, NV, July 2011.

[42] K. Inthavisas and D. Lopresti. Speech Cryptographic Key Regeneration Based on Password. In *Proceedings of the International Joint Conference on Biometrics (IJCB 2011)*, Arlington, VA, October 2011.

[43] F. Itakura (1975) Minimum Prediction Residual Principle Applied to Speech Recognition. *IEEE Transaction on Acoustic, Speech, Signal Processing*, Vol. 3(1), pages 67-72, 1975.

[44] A. K. Jain, K. Nandakumar, and A. Nagar. Biometric Template Security. *EURASIP Journal on Advances in Signal Procesing Special Issue on Biometrics*, January 2008.

[45] Q. Jin, A. Toth, A. Black, and T. Schultz. Is Voice Transformation a Threat to Speaker Identification? In *proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008)*, pages 4845-4848, April 2008.

[46] B. Juang and F. K. Soong. Speaker Recognition Based on Source Coding Approaches. In *International conference on acoustics, speech and signal processing.* Vol.1, pages 613-616, 1990.

[47] A. Juels and M. Sudan. A Fuzzy Vault Scheme. In *Proceedings of IEEE International Symposium on Information Theory*, pages 408, Lausanne, Switzerland, 2002.

[48] A. Juels and M. Wattenberg. A Fuzzy Commitment Scheme. In *Proceedings of Sixth ACM Conference on Computer and Communications Security*, pages 28-36, Singapore, November 1999.

[49] S. Kanade, D. Camara, E. Krichen, D. Petrovska-Delacretaz, and B. Dorrizzi. Three Factor Scheme for Biometric-based Cryptographic Key Regeneration Using Iris. In *biometrics Symposium*, pages 59-64, Tampa, FL, September, 2008.

[50] M. Karnjanadecha. *Signal Modeling with Non-uniform Time Sampling of Features for Automatic Speech Recognition*. PhD thesis, Department of Electrical Engineering, Old Dominion University, August 2000.

[51] E. J. C. Kelkboom, B. Gkberk, T. A. M. Kevenaar, A. H. M. Akkermans, and M. van der Veen. 3D Face: Biometric Template Protection for 3D Face Recognition. In *Proceedings of Second International Conference on Biometrics*, pages 566-573, Seoul, South Korea, August 2007.

[52] T. Kinnunen. *Spectral Featrues for Automatic Text-Independent Speaker Recognition*. Licentiate thesis, Department of Computer Science, University of Joensuu, Findland December 2003.

[53] D. Klein, Foiling the cracker. A Survey of, and Improvements to, Password Security. In *Proceedings of the 2nd USENIX Security Work shop*, August 1990.

[54] J. Kominek and A. Black. The CMU Arctic speech databases. In *5th ISCA Speech Synthesis Workshop*, pages 223-224, Pittsburgh, PA, 2004.

[55] Q. Li and E. C. Chang. Robust, Short and Sensitive Authentication Tags Using Secure Sketch. In *Proceedings of ACM Multimedia and Security Workshop*, pages 56-61, Geneva, Switzerland, September 2006.

[56] S. Lin, and D.J. Costello, Jr. *Error Control Coding Fundamentals and Applications*. Prentice-Hall, N.J., 1983.

[57] D. Lopresti, J. Raim. The Effectiveness of Generative Attacks on an Online Handwriting Biometric, In *Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication*, pages 1090-1099, NY, USA, July 2005.

[58] E. Maiorana, P. Campisi, and A. Neri. Template Protection for Dynamic Time Warping-based Biometric Signature Authentication. In *Proceedings of the 16th international conference on Digital Signal Processing*, pages 526-531, Santorini, Greece, 2009.

[59] E. Maiorana, M. Martinez-Diaz, P. Campisi, J. Ortega-Garcia, and A. Neri. Template Protection for HMM-based On-line Signature Authentication. In *Computer Vision and Pattern Recognition Workshops*, pages 1-6, Anchorage, AK, 2008.

[60] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*, Springer-Verlag, 2003.

[61] T. Masuko, T. Hitotsumatsu, K. Tokuda and T. Kobayashi. On the Security of HMM-based Speaker Verification Systems against Imposture Using Synthetic Speech. In *Proceedings of the European Conference on Speech Communication and Technology*, Vol.3, pages 1223-1226, Budapest, Hungary, September 1999.

[62] T. Masuko, K. Tokuda, and T. Kobayashi. Imposture Using Synthetic Speech against Speaker Verification Based on Spectrum and Pitch. In *Proceedings of the International Conference on Spoken Language Processing*, Vol. 3, pages 302-305, Beijing, China, October 2000.

[63] F. J. McEliece and N. J. A. Sloane. *The theory of Error Correcting Codes*, North Holland, 1991.

[64] F. Monrose, M. K. Reiter, and S. Wetzel. Password Hardening Based on Keystroke Dynamics. In *Proceedings of the 6th ACM Conference on Computer and Communications Security*, pages 73-82, November 1999.

[65] F. Monrose, M. K. Reiter, Q. Li , and S. Wetzel. Using Voice to Generate Cryptographic Keys: A Position Paper. In *Proceedings of Odyssey 2001, The Speaker Verification Workshop*, June 2001.

[66] F. Monrose, M. K. Reiter, Q. Li , and S. Wetzel. Cryptographic Key Generation from Voice (Extended Abstract). In *Proceedings of the 2001 IEEE Symposium on Security and Privacy*, May 2001.

[67] F. Monrose, M. K. Reiter, Q. Li, D. Lopresti, and C. Shih. Towards Speech-Generated Cryptographic Keys on Resource Constrained Devices (Extended Abstract). In *Proceedings of the 11th USENIX Security Symposium*, August 2002.

[68] D. C. Montgonery and G. C. Runger. *Applied Statistics and Probability for Engineers.* John Wiley & Sons, New York, 1999.

[69] A. Nagar, K. Nandakumar, and A. K. Jain. Biometric Template Transformation: a Security Analysis. In *Proc.SPIE, Electronic Image, Media Forensics and Security XII*, San Jose, CA, January 2010.

[70] K. Nandakumar, A. Nagar, and A. K. Jain. Hardening Fingerprint-based Fuzzy Vault Using Password. In*proceedings of 2nd International Conference on Biometrics (ICB)*, pages 927-937, Seoul, South Korea, August 2007.

[71] M. Pandit and J. Kittler. Feature Selection for a DTW-based Speaker Verification System. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 2 pages 769-772. Seattle, WA, May 1998.

[72] T. W. Parsons. *Voice and Speech Processing*. McGraw-Hill, New York, 1987.

[73] B. L. Pellom and J. H. L. Hansen. An Experimental Study of Speaker Verification Sensitivity to Computer Voice Altered Imposters. In *Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing*, March 1999.

[74] N. K. Ratha, J. H. Connell, and R. M. Bolle. An Analysis of Minutiae Matching Strength. In *Proceedings of Third International Conference on Audio, Video-Based Biometric Person Authentication (AVBPA)*, Halmstad, Sweden, pages 223-228, June 2001.

[75] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle. Generating Cancelable Fingerprint Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):561-572, April 2007.

[76] D. A. Reynolds, T. F. Quatieri, and Robert B. Dunn. Speaker Verification Using Adapted Guassian Mixture Models. *Digital Signal Processing*. 10(1-3), pages 19-41, 2000.

[77] A. Rosenberg and M. Sambur, New Techniques for Automatic Speaker Verification. *IEEE Trans. Acoustics, Speech, Signal Proceeding*, 23(2): 169-176, 1975.

[78] A. Rosenberg and A. E. Soong, Evaluation of a vector quantization talker recognition system in text independent and text dependent models. *Computer Speech and Language*, pages 143-157, 1987.

[79] H. Sakoe and S. Chiba. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. Acoustics, Speech, Signal Proceeding*, ASSP-26 (1): 43-49, February 1978.

[80] M. Savvides and B. V. K. Vijaya Kumar. Cancellable Biometric Filters for Face Recognition. In *Proceedings of IEEE International Conference Pattern Recognition*, 3: 922-925, August 2004.

[81] M. Slaney. Auditory Toolbox. *Interval Technical Report# 1998-010*, 1998.

[82] O. T. Song, A. B. J. Teoh, and D. C. L. Ngo. Application-Specific Key Release Scheme from Biometrics. *International Journal of Network Security*, vol. 6, no. 2, pages 127-133, March 2008.

[83] F. K. Soong, A. E. Rosenberg, B. Juang, and L. Rabiner. A Vector Quantization Approach to Speaker Recognition. AT&T Technical Journal 65. pages 14-26. 1987.

[84] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. K. Vijaya Kumar. Biometric Encryption. In *ICSA Guide to Cryptography, R. K. Nichols, Ed* McGraw Hill, 1999.

[85] A. Stoianov. Security of error correcting code for biometric encryption. In *8th Annual International Conference on Privacy Security and Trust*, pages 231-235, Ottawa, Canada, August 2010.

[86] Y. Sutcu, Q. Li, and N. Memon. Protecting Biometric Templates with Sketch: Theory and Practice. *IEEE Transactions on Information Forensics and Security*, 2(3):503-512, September 2007.

[87] Y. Sutcu, Q. Li, and N. Memon. Secure Biometric Templates from Fingerprint-Face Features. In *Proceedings of CVPR Workshop on Biometrics*, Minneapolis, USA, June 2007.

[88] Y. Sutcu, H. T. Sencar, and N. Memon. A Secure Biometric Authentication Scheme Based on Robust Hashing. In *Proceedings of ACM Multimedia and Security Workshop*, pages 111-116, New York, USA, August 2005.

[89] A. B. J. Teoh, A. Goh, and D. C. L. Ngo. Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12): 1892-1901, December 2006.

[90] A. B. J. Teoh and L. Chong, Secure Speech Template Protection in Speaker Verification System. *Speech Communication*. 52(2): 150-163, February 2010.

[91] A. B. J. Teoh, K.-A. Toh, and W. K. Yip. 2N Discretisation of BioPhasor in Cancellable Biometrics. In *Proceedings of Second International Conference on Biometrics*, Seoul, South Korea, pages 435-444, August 2007.

[92] K. Tokuda, H. Zen, and A.W. Black. An HMM-based Speech Synthesis System Applied to English. In *Proceedings of IEEE Speech Synthesis Workshop*, 2002.

[93] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevennar, G. J. Schrijen, A. M. Bazen, and R. N. J. Veldhuis. Practical Biometric Authentication with Template Protection. In *Proceedings of Fifth International Conference on Audio- and Video-based Biometric Person Authentication*, Rye Town, USA, July, 2005.

[94] U. Uludag and A. K. Jain. Securing Fingerprint Template: Fuzzy Vault With Helper Data. In *Proceedings of CVPR Workshop on Privacy Research In Vision*, pages 163, New York, USA, June 2006.

[95] U. Uludag, S. Pankanti, and A. K. Jain. Biometric Cryptosystems: Issues and Challenges. In *Proceedings of the IEEE*, Vol. 92, no. 6, pages 948-960, June 2004.

[96] C. Vielhauer, R. Steinmetz, and A. Mayerhofer. Biometric Hash Based on Statistical Features of Online Signatures. In *Proceedings of 16th International Conference on Pattern Recognition*, vol. 1, pages 123-126, Quebec, Canada, August 2002.

[97] R. H. Woo, A. Park, and T. J. Hazen. The MIT Mobile Device Speaker Verification Corpus: Data Collection and Preliminary Experiments. In *Proceedings of Odyssey, The Speaker and Language Recognition Workshop*, San Juan, Puerto Rico, June 2006. ISBN: 81-7252-119-7

[98] S. Yang and I. Verbauwhede. Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pages 609-612, Philadelphia, USA, March 2005.

[99] W. L. Yee, M. Wagner and D. Tran. Vulnerability of Speaker Verification to Voice Mimicking. In *Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pages 145-148, Hong Kong, October 2004.

[100] X. Zhou. Template Protection and Its Implementation in 3D Face Recognition Systems. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification*, vol. 6539, pages 214-225, Orlando, USA, April 2007.

Curriculum Vitae

Keerati Inthavisas was born on December 9, 1974 to Mr. On and Mrs. Rachada Inthavisas in Wieng-sra, Surat-thani, Thailand. In 1998, he earned a B.Eng. from the Prince of Songkla University. He has worked for the Thai government as a Lecturer at the Rajamangala University of Technology Srivijaya (RMUTSV), Faculty of Agriculture since 1999. In 2004, he earned a M.Eng. in Computer Control Systems (RMUTSV scholarship) from the same place where he earned his bachelor degree. In 2005, he has joined the Faculty of Engineering, RMUTSV. In 2006, he got a full scholarship from RMUTSV to pursue his Ph.D. in Computer Engineering at Lehigh University. He has earned a M.S. in Computer Engineering from Lehigh University in 2010.

List of Publications

K. Inthavisas and D. Lopresti. Speech Cryptographic Key Regeneration Based on Password. In *Proceedings of the International Joint Conference on Biometrics (IJCB 2011)*, Arlington, VA, October 2011.

K. Inthavisas and D. Lopresti. Attacks on Speech Biometric Authentication.

In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV 2011)*, pages 310-316, Las Vecas, NV, July 2011.

K. Inthavisas and D. Lopresti. Biometric Template Protection for Dynamic Time Warping-based User Authentication. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV 2011)*, pages 303-309, Las Vecas, NV, July 2011.

K. Inthavisas and D. Lopresti. Speech Biometric Mapping for Key Binding Cryptosystem. In *Biometric Technology for Human Identification VIII (SPIE Defense, Security, and Sensing)*, pages 80291P-1 - 80291P-12, Orlando, FL, April 2011.

K. Inthavisas, M. Karnjanadecha, and T. Khaorapapong. Synthesis of Vowels and Tones in Thai Language by Articulatory Modeling. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2004),* Jeju Island, Korea, October 2004.

K. Inthavisas, T. Khaorapapong. and M. Karnjanadecha. Synthesis of Thai Monophthongs by Articulatory Modeling In *Proceedings of the 4th Information Engineering Postgraduate Workshop,* Phuket, Thailand, January 2004.