

**DESIGN AND ANALYSIS OF VOICE TEMPLATE PROTECTION  
SCHEMES BASED ON WINNER-TAKES-ALL HASHING**

By

**CHEE KONG YIK**

A dissertation submitted to the Department of Electrical and Electronics  
Engineering,  
Lee Kong Chian Faculty of Engineering and Science,  
Universiti Tunku Abdul Rahman,  
in partial fulfilment of the requirements for the degree of  
Master of Science  
April 2018

## **ABSTRACT**

### **DESIGN AND ANALYSIS OF VOICE TEMPLATE PROTECTION SCHEMES BASED ON WINNER-TAKES-ALL HASHING**

**CHEE KONG YIK**

The emergence of mobile technologies has greatly popularised the usage of speaker recognition based applications (e.g., biometric lock, voice banking and attendance system). However, with the increase in the usage of these biometric based applications, major concerns of the users will be on the security and the privacy of the information stored in the database. Template protection scheme is designed to secure the user biometric from being recovered by attackers. However, current voice template protection schemes do not completely provide the required security properties, such as non-invertibility, unlinkability and revocability. In this dissertation, two voice template protection schemes that are inspired from Winner-Takes-All hashing are proposed. These two newly proposed schemes are named as Random Binary Orthogonal Matrices Projection hashing and two-dimensional Winner-Takes-All hashing. The former scheme is designed for one-dimensional input while the latter scheme is designed for two-dimensional input. To further increase the security of the proposed schemes, additional factor of authentication is incorporated (i.e., random token and additional biometric modality). Extensive analysis is performed to justify the trade-off between the performance and the security of the proposed schemes. The experimental results and analysis have demonstrated that both of the

proposed schemes are able to survive against major privacy and security attacks (e.g., attack-via-record multiplicity and stolen token attack) while preserving the performance of the proposed schemes.

## **ACKNOWLEDGEMENT**

I would like to thank my supervisors, Dr. Yap Wun She and Prof. Ir. Dr. Goi Bok Min for their guidance in providing theoretical and practical knowledge throughout my research. Moreover, I would also like to express my deepest appreciation to my previous co-supervisor, Dr. Jin Zhe. Without his constructive advices and guidance, I would not be able to complete my master's study. Next, I would express my gratitude to my peers, Lai Yen Lung and Badiul Alam for providing support and motivation and willing to share their ideas in helping me to overcome my difficulties. Lastly, I would like to thank my family members for their continuous support and encouragement in pursuing my master degree.

**LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE**

**UNIVERSITI TUNKU ABDUL RAHMAN**

Date: \_\_\_\_\_

**SUBMISSION OF DISSERTATION**

It is hereby certified that **CHEE KONG YIK** (ID No: **16UEM01194**) has completed this dissertation entitled “DESIGN AND ANALYSIS OF VOICE TEMPLATE PROTECTION SCHEMES BASED ON WINNER-TAKES-ALL HASHING” under the supervision of Dr. Yap Wun She (Supervisor) and Prof. Ir. Dr. Goi Bok Min (Co-supervisor) from the Department of Electrical and Electronic Engineering, Lee Kong Chian Faculty of Engineering and Science.

I understand that University will upload softcopy of dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

\_\_\_\_\_  
(Chee Kong Yik)

## APPROVAL SHEET

This dissertation entitled “**DESIGN AND ANALYSIS OF VOICE TEMPLATE PROTECTION SCHEMES BASED ON WINNER-TAKES-ALL HASHING**” was prepared by CHEE KONG YIK and submitted as partial fulfilment of the requirements for the degree of Master of Science at Universiti Tunku Abdul Rahman.

Approved by:

---

(Dr. YAP WUN SHE)

Date:.....

Supervisor

Department of Electrical and Electronics Engineering

Lee Kong Chian Faculty of Engineering and Science

Universiti Tunku Abdul Rahman

---

(Prof. Ir. Dr. GOI BOK MIN)

Date:.....

Co-supervisor

Department of Mechatronics and Biomedical Engineering

Lee Kong Chian Faculty of Engineering and Science

Universiti Tunku Abdul Rahman

## DECLARATION

I hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

Name \_\_\_\_\_

Date \_\_\_\_\_

## TABLE OF CONTENTS

	<b>Page</b>
<b>ABSTRACT</b>	<b>ii</b>
<b>ACKNOWLEDGEMENTS</b>	<b>iv</b>
<b>PERMISSION SHEET</b>	<b>v</b>
<b>APPROVAL SHEET</b>	<b>vi</b>
<b>DECLARATION</b>	<b>vii</b>
<b>TABLE OF CONTENTS</b>	<b>viii</b>
<b>LIST OF TABLES</b>	<b>x</b>
<b>LIST OF FIGURES</b>	<b>xi</b>
<b>CHAPTER</b>	
<b>1.0 INTRODUCTION</b>	<b>1</b>
1.1 Problem Statement	4
1.2 Objectives	7
1.3 Contributions	8
1.4 Organisation of the Dissertation	8
<b>2.0 LITERATURE REVIEW</b>	<b>10</b>
2.1 Cancellable Biometrics	10
2.1.1 Biometric Salting	10
2.1.2 Non-Invertible Transformation	11
2.2 Biometric Cryptosystem	14
2.3 Hybrid Biometric Cryptosystem	18
2.4 Winner-Takes-All Hashing (WTA)	20
<b>3.0 RANDOM BINARY ORTHOGONAL MATRICES PROJECTION (RBOMP)</b>	<b>22</b>
3.1 Introduction	22
3.2 The Specification of RBOMP	23
3.2.1 Determining the Range of $\omega$	26
3.2.2 Matching	27
3.3 Experimental Setup and Performance Analysis	29
3.3.1 Effect of $\omega$ and $k$ on the Recognition Performance of the Proposed Method	29
3.3.2 Comparison of the Recognition Performances for Different Methods	30
3.3.3 Comparison of the Recognition Performances for Different Databases	32
3.4 Security Analysis	33
3.4.1 Revocability Analysis	33
3.4.2 Unlinkability Analysis	35
3.4.3 Brute Force Attack	36



3.5	Security Analysis against Attack-via-Record Multiplicity (ARM)	37
3.6	Summary	49
<b>4.0</b>	<b>TWO DIMENSIONAL WINNER-TAKES-ALL HASHING (2DWTA)</b>	<b>50</b>
4.1	Introduction	50
4.2	The Specification of 2DWTA	51
4.2.1	Architecture of the Multimodal Biometric System	51
4.2.2	Fusion of Fingerprint and Voice Template	52
4.2.3	Two-dimensional Winner-Takes-All Hashing (2DWTA)	54
4.2.4	Matching	56
4.3	Experimental Setup and Performance Analysis	57
4.3.1	Effect of k and h on the Recognition Performance of the Proposed Method	59
0		60
0		62
	Security Analysis	
4.4.1	Revocability Analysis	62
4.4.2	Unlinkability Analysis	64
4.4.3	Brute Force Attack	66
4.4.4	Attack-via-Record Multiplicity (ARM)	68
4.5	Summary	70
<b>5.0</b>	<b>CONCLUSION</b>	<b>71</b>
5.1	Conclusion	71
5.2	Future Work	73
	<b>REFERENCES</b>	<b>74</b>
	<b>LIST OF PUBLICATION</b>	<b>82</b>

## LIST OF TABLES

<b>Table</b>		<b>Page</b>
2.1	Overview of the existing template protection schemes	19
3.1	Comparison of different speech template protection methods	29
3.2	Comparison of recognition performance for different datasets	32
3.3	Comparison between number of intermediate indices compromised versus time complexity required to access the system	45
4.1	Performance of different parameters on dataset combination of $V$ and $F1$ in EER (%)	59
4.2	Performance of different parameters on dataset combination of $V$ and $F2$ in EER (%)	59
4.3	Performance of different parameters on dataset combination of $V$ and $F3$ in EER (%)	60
4.4	Performance comparison between unimodal biometric and WTA in stolen token scenario	61

## LIST OF FIGURES

Figures		Page
1.1	Generic biometric authentication system	2
1.2	Points of attacks on a biometric authentication system	5
2.1	Example of WTA computation	20
3.1	One round RBOMP hashing	25
3.2	Example of similarity score computation for RBOMP hashing	27
3.3	EER versus number of random binary orthogonal matrices for different lengths of window	29
3.4	Distribution of genuine, impostor and pseudo-impostor scores	34
3.5	Distribution of pseudo-genuine scores and pseudo-impostor scores	35
3.6	Mapping of intermediate index, $C$ , at $\omega = 4$ and $\omega = 6$ respectively	38
3.7	Guessing the index, $C$ , at $\omega = 6$ when both hashed code and random token are compromised	39
3.8	Deriving the order of $i$ -vector by comparing two desired windowed vector	41
3.9	Expected number of trials needed to collect $m$ distinct binary matrices	44
3.10	Minimum number of templates required to compromise to access the system for different data size	48
4.1	Overview of the proposed feature level fusion template protection method	52
4.2	Fusion of voice template and fingerprint template at feature level via matrix multiplication	54
4.3	Process of one round 2DWTA	56

4.4	Example of similarity score computation for 2DWTA	57
4.5 (a)	The genuine, impostor and pseudo-impostor distribution for database combination: $V$ and $F1$	63
4.5 (b)	The genuine, impostor and pseudo-impostor distribution for database combination: $V$ and $F2$	63
4.5 (c)	The genuine, impostor and pseudo-impostor distribution for database combination: $V$ and $F3$	64
4.6 (a)	The pseudo-genuine and pseudo-impostor distribution for database combination: $V$ and $F1$	65
4.6 (b)	The pseudo-genuine and pseudo-impostor distribution for database combination: $V$ and $F2$	65
4.6 (c)	The pseudo-genuine and pseudo-impostor distribution for database combination: $V$ and $F3$	66
4.7	Example of hashed code, $k$ and permutation seeds being compromised by the adversary	69

## CHAPTER 1

### INTRODUCTION

Authentication is the process of proving or confirming the user's claimed identity. The emergence of e-commerce business, information and communication technologies have increased the demand for more secure and reliable security systems to prevent any breaches in database, hacking, or unauthorised use of information. In 2016, Morgan (2016) reported that United States government increased its budget in cybersecurity by 35% from \$14 million in year 2016 to \$19 million in year 2017. In addition, Cybersecurity Ventures had projected that \$1 trillion will be spent globally in cybersecurity from year 2017 to year 2021. On the other hand, the cost of cybercrime in year 2015 had been increased by 19% and it had been projected to reach \$2 trillion by year 2019 (Ponemon Institute, 2015; Morgan, 2016). To ensure that only the true user is eligible to access the data, authentication procedure are often implemented to protect the privacy of these data. Generally, the authentication mechanism may include the following authentication factors:

- 1) Knowledge. The user is required to prove knowledge of a secret to gain access, e.g., access key or passcode.
- 2) Possession. The user possesses the physical token or key to gain access.
- 3) Inheritance. The user provides his biometric traits to authenticate himself.

Of the aforementioned factors, biometric provides a more convenient and reliable measure of authentication due to the fact that biometric has a stronger representation or association with the identity of an individual. Unlike other factors of authentication (e.g., passcode or token), biometric authentication does not require the user to memorise any password or carry any tokens, hence this greatly reduces the concern of password being stolen or misplaced of physical token (Dharavath et al., 2013).

Biometric utilises the physical or behavioural traits of an individual in authentication process. Well-known instances of biometric modalities include fingerprint, face, voice, signature and iris. A generic biometric authentication system consists of five major components as described in Figure 1.1.

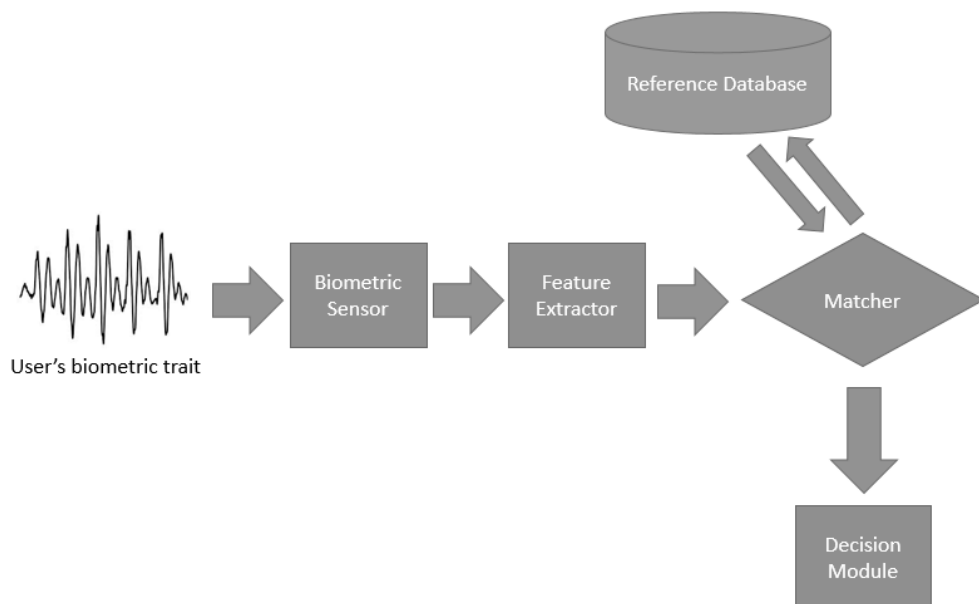


Figure 1.1: Generic biometric authentication system

In a biometric authentication system, there are two stages of operation, namely enrolment stage and verification stage (Jain et al., 2016). In the enrolment stage, a sensor will be used to scan the user's biometric traits and output the digital representation of the biometric traits (e.g., image or audio). Since the quality of the scanned biometric traits may be affected by different factors (e.g., background noise, brightness, camera resolution etc.), feature extractor is needed to extract important/useful information that can best express the identity of the user. Feature extractor will generate a compact representation of the user with sufficient discriminative features. Subsequently, the compact representation of the user will be stored as a template in the database as reference for future comparison. In the verification stage, the user will provide his query biometric traits to authenticate or identify himself. A template is then generated by feeding the new biometric trait to the feature extraction. Finally, the resultant template will be compared with the template pre-stored in the database for matching purpose. Notice that the matching of the query template and enrolled template will be conducted based on the objective of the user recognition system. If the user wishes to identify himself without claiming his identity, the query template needs to be compared with all the enrolled templates stored in the database (i.e., one-to-many matching). In this case, the operation is referred as *identification* mode. In contrast, if the objective of the user recognition system is to verify himself using a claimed identity, the system will match the query template with the template of the claimed identity (i.e., one-to-one matching). In this situation, the operation of the system is referred as *verification* mode. The results of the authentication will depend on the similarity scores obtained from the matching of the templates and the threshold of

authentication. If the similarity scores of the templates exceed the threshold, the user is deemed as the true user, else the user is denied of his access.

Of the aforementioned instances of biometric modalities, biometric authentication based on voice recognition/ speaker recognition is gaining more popularity due to the advancement of technologies and the increase in the usage of mobile devices. Instances of speaker recognition based applications include biometric locks, time and attendance systems and voice banking. Similar to biometric authentication system, speaker recognition can be categorised as speaker identification and speaker verification, where the former determines a registered speaker from a set of known speakers while the latter decides whether the input voice is from the claimed speaker.

## **1.1 Problem Statement**

Despite the convenience and popularity of biometric, biometric authentication may vulnerable to adversarial attacks. Adversary launches his attacks by exploiting the loopholes in the system design and the available resources needed to access the system (Jain et al., 2008). Ratha et al. (2001) categorised the points of attacks in the biometric authentication system into eight different points as illustrated in Figure 1.2.



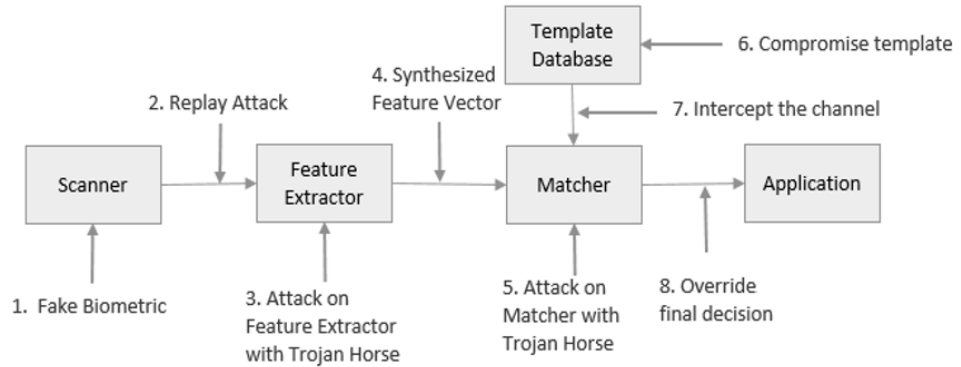


Figure 1.2: Points of attacks in a biometric authentication system

Of those points of attacks, biometric template leakage is considered as one of the most harmful attacks in the biometric system (Jain and Nandakumar, 2012). As biometric traits are strongly associated with the identity of an individual, the compromise of biometric template can lead to the following consequences (Raju et al., 2014):

- i) Impostor can replace the template to gain illegitimate access to the system.
- ii) Physical spoof can be created from biometric template to gain unauthorised access to systems that use the same biometric trait.
- iii) Stolen biometric template can be replayed to the matcher to gain illegal access to the system.
- iv) Stolen biometric template can be used by the adversary to perform cross-matching across other databases to covertly track a person without their consent.

The compromised biometric information/ biometric template stored in the database greatly expose the individual to the threat of identity-theft and misuse of information. To make things worse, biometric traits are irreplaceable

once compromised. In the context of speaker recognition system, the authenticity of a user is determined by matching the voice reference stored in the database. Hence, this raises the concern on the protection of the voice reference stored in the database to prevent security and privacy threats.

Generally, template protection schemes are needed to protect the template stored in the database against privacy and security attacks. Hence, a reliable template protection scheme needs to be designed and must fulfil all of the following requirements (Nandakumar and Jain, 2015):

- i) Irreversibility. It should always be computationally hard for the adversary to invert the protected biometric template.
- ii) Unlinkability. It should always be computationally hard for the adversary to distinguish whether multiple protected biometric templates were generated using the same biometric trait of a user.
- iii) Revocability. It should be feasible to renew or revoke the protected biometric template to replace the old template while the original template should be computationally hard to be inverted from multiple protected biometric templates derived from the same biometric trait of a user.
- iv) Performance. The performance of the biometric recognition rate should not be seriously degraded.

From the existing voice template protection schemes, several issues have been observed and needs to be addressed.

- i) **Robustness to attacks:** It is observed that majority of the speech template protection schemes were vulnerable to different attacks such as attack-via-record multiplicity (ARM) and stolen token attacks. The vulnerability of the scheme is most likely due to the high correlation between the templates generated using the same biometric feature. Hence, the adversary is able to derive the original template by analysing multiple compromised templates. Thus, there is an urgent need to ensure that the generated templates are independent to each another, fulfilling the unlinkability and revocability criteria.
- ii) **Performance degradation:** The use of feature transformation approach to transform biometric from one space to another will cause the loss of the discriminative features. Thus, it will result in the increase of the intra-class variation and eventually lead to the accuracy degradation in the performance. Therefore, the template protection scheme should be able to preserve the performance of the system as much as possible while providing sufficient security protection.

## **1.2 Objectives**

This dissertation is concerned with protection of the voice template stored in the database. Since template protection in biometric authentication system is urgently needed, this work aims to:

- i) Design cancellable template protection schemes that project the biometric feature to ordinal space that is resilient to insignificant intra-class variation while inducing strong non-invertible property.
- ii) Integrate additional factor of authentication (i.e., user-specific token and other biometric traits) to further enhance the security of the proposed schemes.
- iii) Perform rigorous analysis on the performance and the security of the proposed schemes to justify the common trade-off between the performance and security.

### **1.3 Contributions**

This research aims to present reliable voice template protection schemes that are able to survive against major attacks while preserving the performance of the system. Rigorous analyses on the performance and the security aspect of the proposed schemes are performed to ensure all evaluation criteria of template protection are satisfied. Most importantly, exhaustive analysis on attack-via-record multiplicity (ARM), which is considered one of the most harmful attacks (Scheirer and Boulton, 2007), is conducted to ensure the reliability and security of the proposed schemes.

### **1.4 Organisation of the Dissertation**

The remainder of the dissertation is organised as follows. Chapter 2 details the literature review on the existing template protection schemes and

Winner-Takes-All hashing where the proposed schemes are derived from. Chapter 3 provides the details and analyses of the proposed Random Binary Orthogonal Matrices Projection (RBOMP) hashing. Chapter 4 describes the limitation of RBOMP hashing and further introduces two-dimensional Winner-Takes-All (2DWTA) hashing. Finally, Chapter 5 concludes the findings obtained in this dissertation and provides suggestions for future research work.

## CHAPTER 2

### LITERATURE REVIEW

In this chapter, the previous work on speech template protection schemes are discussed and summarised. Generally, biometric template protection schemes can be categorised into cancellable biometrics, biometric cryptosystem and hybrid biometric cryptosystem. Lastly, a summary of the related work of different architectures are presented in a table.

#### 2.1 Cancellable Biometrics

Cancellable biometric refers to the intentional distortion of the biometric features to enhance and protect the user's privacy (Ratha et al., 2007). If the cancellable feature is compromised, the same biometric feature can be mapped into a new distinct template using the pre-designed distortion characteristics. In other words, with the same biometric data, one can generate a new biometric template as there is no correlation between the newly generated templates. Cancellable biometrics can further divided into biometric salting and non-invertible transform.

##### 2.1.1 Biometric Salting

Biometric salting (Labati et al., 2012) blends an auxiliary data to the biometric data to produce a transformed template. Instances of auxiliary data

include user-specific key or password which is derived externally. Hence, the existing biometric template can be easily revoked and replaced with a newly generated template. However, such an approach depends heavily on the security of the external token or password. In the event that the token or confidential key being stolen (i.e., stolen token attack), the recognition performance of the system will degrade significantly. A well-known example of biometric salting is biohashing which is based on a user-specific random projection (Teoh et al., 2008). In the context of speech template protection scheme, a concrete example would be probabilistic random projection approach proposed by Chong and Teoh (2007). Two-dimensional principal component analysis was applied on the speech feature matrix and feature matrix is then gone through a random projection process using an externally derived pseudo-random number. The resulted/projected matrix was then fed into Gaussian Mixture Model (GMM) to obtain probabilistic speaker models. It was shown in their work that the scheme was able to preserve the recognition performance even under the stolen token scenario where the token had been compromised. However, the scheme was vulnerable to attack-via-record multiplicity as the adversary will be able to derive the original feature template by exploiting multiple templates generated using different random projection matrices (Wang and Plantaniotis, 2010).

### **2.1.2 Non-Invertible Transformation**

Non-invertible transform refers to the use of one-way transformation function to convert the biometric data in such a way that it is computationally difficult to be inverted (Labati et al., 2012). Xu and Cheng (2008) proposed the use of fuzzy vault scheme and a non-invertible function in voice template

protection. Chaff points were added to the unordered Mel-Frequency Cepstral Coefficient matrix to create a vault. A prime accumulator was then used to separate the genuine points from the chaff points. For authentication, polynomial reconstruction is used while a non-invertible function is applied to hide the raw feature from the adversary. However, Chang et al. (2006) argued that the selection of the new chaff points depends on the previously selected points. The latecomers, referring to the points added later, are more likely to have more nearby points. This leads to the increase in the likelihood that the adversary will be able to correctly guess the genuine points with the increase in the chaff points. Besides, this scheme is not resistant to stolen token attack as the adversary will be able to obtain the genuine points easily once the prime accumulator is compromised.

In 2016, Pandey et al. proposed a new technique named deep secure encoding (DSE) for face template protection. The face feature was first extracted and trained using deep convolutional neural networks (CNN) to generate an unprotected binary template. The binary template was then divided into  $n$   $k$ -bit blocks where each  $k$ -bit block was used as an input of a cryptographic hash function, namely SHA-256. The resulting  $n$  outputs of the hash function will then be stored in the database. For verification process, the query face image will go through similar extraction and hashing process, and the hashed output from the query will be compared with the hashed output stored in the database. The authentication is successful if the number of the matches between the enrolled hashed output and the query hashed output exceeds the predefined threshold. Different from a typical template protection



scheme, where the key is used in securing the unprotected template for cancellable property, the random key is used and embedded during the face extraction and training processes. Hence, the user is able to generate a new key to replace the embedded key. However, exhaustive training process is required in order to regenerate a new template using the new key.

Yang et al. (2018) had proposed a fingerprint and finger-vein based cancellable multi-biometric system where both fingerprint and finger-vein modalities are fused at feature level. Three different fusion options are introduced which integrate the use of enhanced partial discrete Fourier transform (EP-DFT) to cater for different demands on accuracy and security. Wavelet transform is first applied to the biometric feature that is in binary representation followed by partial discrete Fourier Transform. (Wang et al., 2013). This process will convert the biometric feature into real-value representation which will provide non-invertible property to the system as it will be difficult to invert the transformed feature back to binary form. The author has claimed that this work is able to provide satisfactory recognition results while able to satisfy the biometric template protection evaluation criteria. In addition, the author also provides justification that the proposed EP-DFT is able to withstand attack-via-record multiplicity through the adjustment of the parameters in wavelet transform to generate different outputs.

While cancellable biometric can satisfy both unlinkability and revocability requirement of template protection, cancellable biometric still consists of several limitations such as not resistant against stolen token attack

and suffers performance degradation when transforming from one space to another.

## 2.2 Biometric Cryptosystem

Biometric cryptosystem uses helper data to generate or secure the cryptographic key by the means of biometric (Jain and Nandakumar, 2008). Helper data refers to the public information that is derived from the biometric without revealing any information regarding the original biometric data. Generic biometric cryptosystem uses fuzzy-based schemes, such as fuzzy commitment (Juels and Wattenberg, 1999) and fuzzy vault (Juels and Sudan, 2006).

Fuzzy commitment was first proposed by Juels and Wattenberg (1999). Fuzzy commitment involves a two-step algorithm, i.e., commitment and decommitment. The fuzzy commitment scheme  $F$  commits a random codeword  $c$  using a one-way hash function  $h$  and a template  $x$ , where both  $c$  and  $x$  are represented in binary form of  $n$ -bit strings. Mathematically,  $F(c, x) = (h(c), x - c)$  and the output is stored in the database. To decommit a query,  $x'$ , denoted as the witness, is used such that the extracted commitment  $c' = f(x' - (x - c))$ , where  $f$  is the decommit function. Decommitment is successful if  $h(c)$  is similar to  $h(c')$  within the capacity of error-correcting code (ECC).

A well-instance of fuzzy commitment scheme was realised by Inthavasis and Lopresti (2011). Inthavasis and Lopresti proposed a password based

cryptographic key generation. Dynamic time warping (DTW) was performed on the extracted voice feature and it was then mapped onto a binary string named feature descriptor, which was subsequently used to define distinguishing features. The template would then be perturbed and a stable feature will be extracted for each perturbations. The extracted feature would be the key and this process will be continued until the length of the distinguishing descriptor has less than or equal to half of the length of the feature vector. Lastly, the template would go through transformation, permutation and key binding process under fuzzy commitment framework. The main limitation of this work is that the security of this work is actually dominated by the security management of the password instead of the biometric itself.

Meanwhile, Billeb et al. (2015) proposed the use of fuzzy commitment scheme on voice template protection under the universal background model (UBM) framework. The proposed scheme binarised the supervector derived from UBM and an adapted fuzzy commitment scheme was used as the basis for the template protection scheme. Even though security analysis against unlinkability and privacy protection was provided, the proposed scheme still suffers from ARM when both key and the difference vectors are compromised. The adversary can exploit the compromised information to reconstruct the template stored in the database.

On the other hand, Paulini et al. (2016) presented the use of multi-bit allocation in contrast to the single-bit allocation used by Billeb (2015). The proposed scheme divided the feature space into  $2^k$  intervals and each interval

was encoded with  $k$  bits. A modified fuzzy commitment scheme was further applied on the binarised feature and it was shown that their scheme was able to outperformed Billeb (2015) with lesser degradation in the recognition ability. However, the issue on the vulnerability against ARM was not addressed.

Fuzzy vault scheme proposed by Juels and Sudan (2006) locks the secret key  $k$  under an unordered set  $k$ . A polynomial  $p$  was selected in such a way that it is able to encode  $k$  into variable  $x$ . Random chaff points (i.e., dummy points that are added to confuse the adversary) that do not lie on  $p$  were then added to set  $A$ , creating a vault which consists of collection of points lying on  $p$  and chaff points. If the query represented by another unordered set  $B$  overlaps considerably with  $A$ , the collection of points that lie on polynomial  $p$  can be identified. With adequate amount of identified points and error correction ability, the polynomial  $p$  can be reconstructed and thereby key  $k$ .

A vaulted verification protocol was proposed by Johnson et al. (2013). A challenge-respond protocol and fuzzy vault were used in their security scheme. By using the work proposed by Inthavasis and Lopresti (2011) as their baseline system, the proposed method was able to outperform Inthavasis and Lopresti's (2011) work under stolen token scenario. Firstly, the voice feature was separated into several blocklets and chaff blocklets were added to each real blocklets forming many pairs of chaff and real blocklets. These pairs were then encrypted by password and stored in database. During authentication, the template was first decrypted and challenging bitstring was generated such that real blocklet represents "0" and chaff blocklet represent "1". The pairs were

then randomly swapped and score computation was performed by matching the bitstring response provided by the user with the template. However, it was stated in Johnson (2013) that the limited biometric information, such as limited voice input, will not be able to vary the data in challenge-response process due to lesser pairings of the real and chaff blocklets, thus making it easier for the adversary to guess the correct response.

An alternative approach using Homomorphic Encryption (HE) had been proposed by Gomez-Barrero et al. (2017) recently without involving any auxiliary data or helper data. HE allows computation to be performed on ciphertexts, which will in turn generate encrypted results. When decrypting the encrypted result, it will match the results of the operations carried out on the original plaintext (Fontaine and Galand, 2007). In the proposed work, the client would first extract the probe template and received the encrypted enrolled template from the database. The similarity scores were then computed and then encrypted to be passed to the authentication server to be decrypted using the key. Paillier homomorphic probabilistic encryption scheme (Paillier, 1999) was used to encrypt the data and Gomez-Barrero had made several adaptations to his model such as the comparator was moved to the authentication server to prevent hill-climbing attack and there is no collusion between the authentication server and the database server. Their experiment were conducted using multimodal biometric framework, using on-line signature and fingerprint modalities, with different type of fusions (i.e., feature level, score level and decision level). Their work showed promising experimental results with no degradation in the performance of the recognition system and further analyses

on the biometric template protection requirement were performed (i.e., unlinkability analysis, revocability analysis and irreversibility analysis). However, as mentioned by Gomez-Barrero, the use of such sophisticated schemes required a higher computational load and the proposed scheme did not provide any analysis on security against ARM.

Recent work by Badiul et al. (2018) proposed the use of bit-toggling strategy incorporated with random projection and discrete Fourier Transform in protecting fingerprint templates. The work consisted of three stages, namely condensed polar grid base 3-tuple quantisation, protected PGTQ and Cancellable PGTQ where quantisation of raw biometric sample, non-invertible transform and cancellable procedures are performed in the respective stages. Raw fingerprint feature is first binarised into a binary template and bit-toggled. Discrete Fourier Transform is then applied, followed by random projection through a random matrix to offer cancellable property. Analysis on the biometric template protection evaluation is well described in this work, however as mentioned by the author, the use of large dimension random matrix may increase the operation time of the system.

Similar to cancellable biometric, biometric cryptosystem also possesses several drawbacks such as exploitation of the helper data may expose the system to privacy and security threats. Besides that, biometric cryptosystem is not able to generate multiples unlinkable templates, hence unable to satisfy the unlinkability criteria of the template protection requirement.

### 2.3 Hybrid Biometric Cryptosystem

As both cancellable biometric and biometric cryptosystem have limitations, a hybrid approach of combining both cancellable biometric and biometric cryptosystem is proposed to overcome such limitation (Jain and Nandakumar, 2012). Hybrid biometric cryptosystem combines of two or more biometric template protection schemes to meet more or all the biometric template protection requirement. Instances of hybrid biometric cryptosystem are fuzzy vault with bioHash and key-generation with cancellable biometrics (Chandra and Kanagalakshmi, 2011). Hybrid biometric cryptosystem reaps the benefit of cancellable properties from cancellable biometric while providing stronger security and privacy protection inherited from biometric cryptosystem. An instance of hybrid biometric cryptosystem is the cancellable speech template based on chaff point mixture method proposed by Zhu et al. (2012) where a two-step hybrid approach (i.e., random projection and fuzzy vault) was used. The voice feature matrix was first randomly projected into another feature space and chaff points were added to the projected space instead of directly to the original feature matrix. Binary indices were used to bind the points and accumulator of genuine indices (i.e., key) were calculated using OR operator. The key will be sent to the matcher to filter out the genuine points from query using AND operator. The proposed work had shown that it was able to preserve the performance of the recognition system, however the security of the proposed work is not analysed in detail as ARM analysis and lost key scenario were not considered. In the event that the binary indices and the key are compromised, the adversary will be able to differentiate the genuine points from randomly

added chaff points. The summary of the literature review is presented in Table 2.1.

Table 2.1: Overview of the existing template protection schemes

Authors	Proposed Scheme	Limitations
Chong et al. (2007)	Probabilistic Random Projection	Vulnerable to ARM
Xu and Cheng (2008)	Fuzzy Vault	Chaff points selection were not independent hence adversary can easily guessed the genuine points
Inthavasis and Lopresti (2011)	Cryptographic key generation based on password	Dependent on password management. System is vulnerable if password is compromised
Zhu et al. (2012)	Random Projection with Fuzzy Vault	Vulnerable to ARM
Johnson et al. (2013)	Vaulted Voice Verification	Data in challenge-response process are dependent on the biometric information
Billeb et al. (2015)	Universal Background Model	Vulnerable to ARM
Paulini et al. (2016)	Multi-bit Allocation	Vulnerable to ARM
Pandey et al. (2016)	Deep Secure Encoding	Exhaustive training process is required
Gomez-Marrero et al. (2017)	Homomorphic Encryption	High computational load required
Badiul et al. (2018)	Cancelable PGTQ	Long operation time



## 2.4 Winner-Takes-All Hashing (WTA)

Winner-Takes-All hashing (WTA) was first proposed by Yagnik et al. in 2011 and was used for fast similarity search, mainly for images by Google (Yagnik et al., 2011; Dean et al., 2013). In brief, WTA used rank correlation measures and record the index of the maximum value of the biometric feature after applying random permutations. By varying the permutation seeds, different index vectors can be generated. Figure 2.1 shows an example of WTA computation.

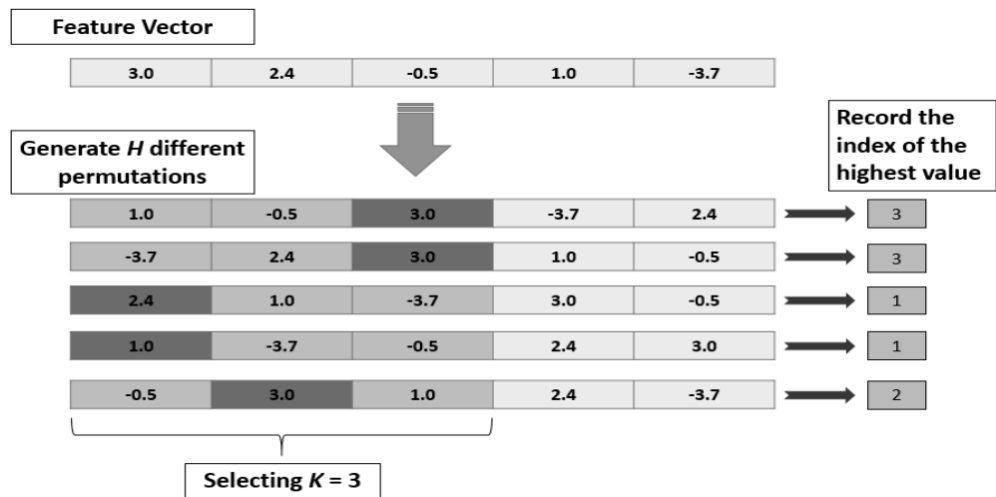


Figure 2.1: Example of WTA computation

The procedure to derive the index vector is described as follows:

- i. *Random Permutation.* The feature vector,  $X$ , is randomly permuted to generate  $X^P$ , where  $X^P$  denotes the permuted feature vector.

- ii. *Selecting  $K$  items.* The first  $K$  items are selected from  $X^P$  for  $2 \leq K \leq n - 1$ . In this step, there will be lost of information due to the reduction in the length of the feature vector.
- iii. *Record the index of the maximum value.* From the  $K$  items, the index of the maximum value is recorded and denoted as  $C$ .
- iv. Step i to iii are repeated using  $H$  different permutation seeds. A series of indexes,  $C_i$  will be generated, where  $i \in [1, H]$  and let  $S$  denotes the set consisting the series of generated  $C_i$  (i.e.  $S = \{C_1, C_2, \dots, C_H\}$ ).

## CHAPTER 3

### RANDOM BINARY ORTHOGONAL MATRICES HASHING

#### (RBOMP)

#### 3.1 Introduction

The use of rank correlation measures in WTA induces a strong non-invertible property as it is computationally difficult to invert from ordinal space back to linear space. In addition, by shifting the focus to the ranking of the feature instead of the value of the feature, WTA allows insignificant variations in the feature value, thus making it more robust to the small changes or noise. However, as WTA only focuses on the rank of the feature, the adversary may obtain the order of the feature through ARM and reconstruct the order of the original template. Hence, motivated by this fact, a new template protection scheme named Random Binary Orthogonal Matrices Projection hashing (RBOMP) is proposed to overcome such limitation. In short, RBOMP transforms the voice feature from linear space to ordinal space via a binary orthogonal matrix. To further strengthen the security of the system, a user-specific random token is incorporated and a non-invertible function namely prime factorisation is used to conceal the returned index. In this work, the voice feature is represented in the form of a fixed-length real value representation, named *i*-vector (Dehak et al., 2011; Snyder et al., 2015).

The remainder of the chapter is organised as follows. In Section 3.2, the specification of RBOMP is discussed. Section 3.3 describes the experimental setup and provides performance analysis while Section 3.4 outlines the general security analysis. Section 3.5 presents the security of RBOMP hashing against attack-via-record multiplicity (ARM) in detail and lastly Section 3.6 gives a brief summary of this chapter.

### 3.2 The Specification of RBOMP

RBOMP is a hashing scheme consisting of  $k$  rounds of function  $h$  for  $k > 1$ . For ease of understanding, let  $i$  denotes the round number where  $i = 1$  to  $k$ . Each round function  $h_i$  takes an  $i$ -vector  $X$  that consists of  $n$  real numbers and a random positive integer  $Z_i$ , where  $1 \leq Z_i \leq 10000$ , as input and generates an index  $S_i$  as output. The concatenation of indexes  $S_i$  generated in each round function  $h_i$  is denoted as the hashed code  $S = S_1||S_2||\dots||S_k$ , where  $||$  denotes the concatenation. Mathematically, we have  $S = RBOMP(X, Z)$ , where  $Z = \{Z_1, Z_2, \dots, Z_k\}$  and  $S_i = h_i(X, Z_i)$ . More precisely,  $h_i$  consists of the following steps:

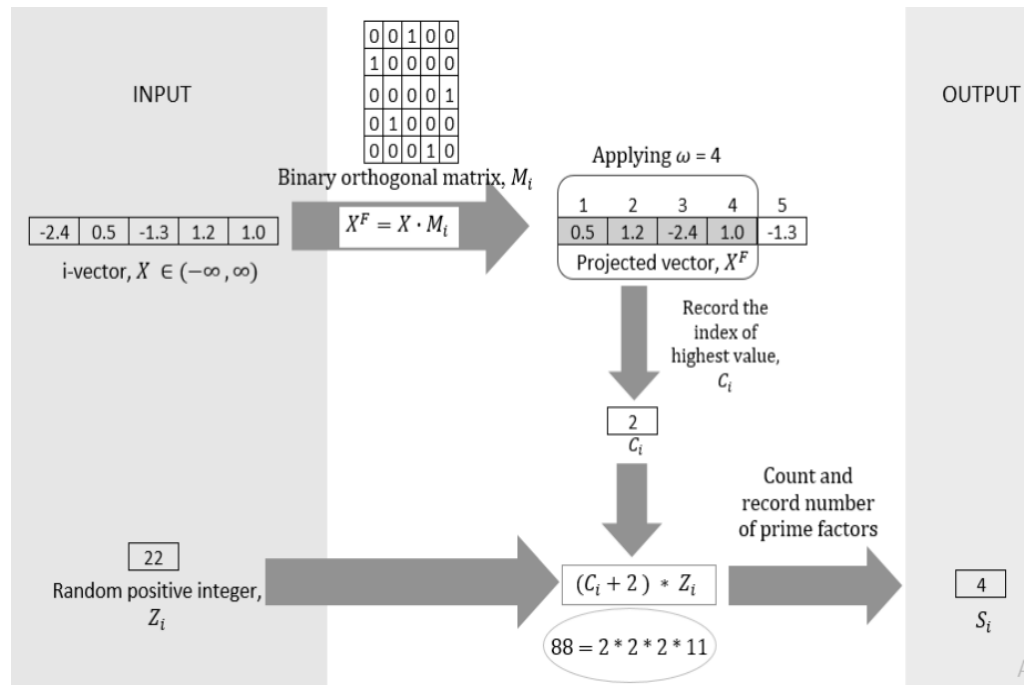
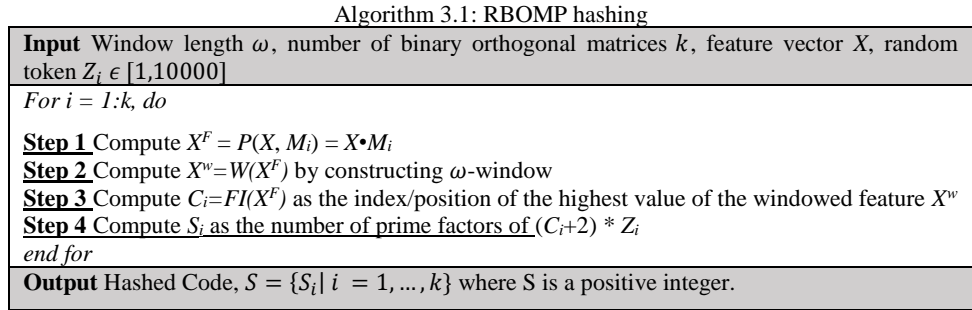
1. *Projection, P*: Given a random binary orthogonal matrix  $M_i$  with a dimension of  $n \times n$ , where  $n$  is the length of the  $i$ -vector, compute the feature vector  $X^F = P(X, M_i) = X \cdot M_i$ , where  $\cdot$  is the matrix multiplication.
2. *Window, W*: Given the feature vector  $X^F$  and the window length  $\omega$  (the exact range of  $\omega$  will be determined through experiments later), compute the windowed feature  $X^w = W(X^F)$  by taking first  $\omega$  real numbers of  $X^F$ . Since the length of the feature vector is reduced, certain information of the feature

vector is lost. The lost of information will help to increase the security of the system as it will limit the amount of information available to the adversary in the expense of the system recognition ability, hence there will be a trade-off between security and performance.

3. *Find Intermediate Index, FI*: Given the windowed feature  $X^w$ , compute the intermediate index  $C_i = FI(X^F)$  as the index/position of the highest value of the windowed feature  $X^w$ .
4. *Prime Factorisation, PF*: Given the intermediate index  $C_i$  and a positive integer  $Z_i$ , compute the index  $S_i = PF((C_i+2) * Z_i)$ , as the number of prime numbers of  $(C_i+2) * Z_i$ , where  $*$  is the integer multiplication. The addition of 2 with  $C_i$  is performed to lower the false acceptance rate (due to the fact that when  $C = 1$  and  $C = 2$  will yield the same factorisation).

Notice that different random binary orthogonal matrices  $M_i$  will be selected in different rounds of  $h$  function. In real world scenario, the selection of binary orthogonal matrices and random token (i.e.,  $Z$ ) is user-specific. In the event of the template being compromised, the user can revoke and reissue a new template by generating different binary orthogonal matrices and/or token to replace the compromised template. In this dissertation, the focus will be on lost-token scenario to evaluate the recognition performance and perform the security analysis for RBOMP hashing. In real-world scenario, all users will have different binary orthogonal matrices and random token, hence, to depict the worst-case scenario, the experiment will be carried out in lost-token scenario. In the lost-token scenario, the binary orthogonal matrices as well as the random token are assumed to be known to the adversary, therefore in the experiments,

all the users are assumed to share the same binary orthogonal matrix and the random token. The pseudocode of the proposed scheme is shown in Algorithm 3.1 while the one-round graphical implementation of RBOMP hashing is shown in Figure 3.1 for illustration purposes.



### 3.2.1 Determining the Range of $\omega$

As the range of value of the intermediate index,  $C_i$ , is closely related to the value of  $\omega$ , where  $1 \leq C_i \leq \omega$ . The range of  $\omega$  is set in such a way that there will be at least two mappings of distinct value  $C_i$  to  $PF(C_i + 2)$ . Here the  $Z$  is ignored as it is assumed to be known to the adversary. Hence, the range of  $\omega$  is set to be  $(2^{q-1} * 3) - 2 \leq \omega < 2^{q+1} - 2 < 500$ , where  $q$  is an integer in the range of  $[2, 8]$ . The motivation is to prevent the adversary from reconstructing the order of the  $i$ -vector through ARM practically.

### 3.2.2 Matching

Transforming the feature to ordinal space which is not sensitive to the value of the feature dimension shifts the focus to the implicit ordering implied by the values (Dean et al., 2013). As rank correlation refers to the measure of the degree of correlation between the ranks of the members within a set, the similarity measurement of the feature representation can be defined as the degree to which the rank of their feature dimension agrees (Dean et al., 2013). Let  $c$  refers to the maximum value of a given  $\omega$ -sized window. The similarity score is defined as the probability of both hashed code  $S$  and  $S'$  having  $c$  at the same position (i.e.  $S_x = S'_x$  for  $x = 1, \dots, k$ ). The higher the probability implies that the hashed code  $S$  and  $S'$  have a high similarity. In our experiments, similarity score will be calculated by counting the number of zeros after performing element-wise subtraction between two hashed codes.

The procedure of the similarity score calculation is described as follows.

- i. *Taking the difference of two hashed codes.* Given an enrolled hashed code,  $S_x$  and a query hashed code,  $S'_x$ , the difference of  $S_x$  and  $S'_x$  is computed by taking  $S_x - S'_x$ .
- ii. *Count the number of zeros.* The number of "0" is counted after taking the difference of  $S_x$  and  $S'_x$ . The "0" in this case indicates a match between the hashed codes and by counting the number of "0", the total matches of two hashed codes can be determined.
- iii. *Compute Similarity Score.* Similarity score is computed by taking the total number of "0" over the length of the hashed code.

Besides, Figure 3.2 shows an example of similarity score computation.

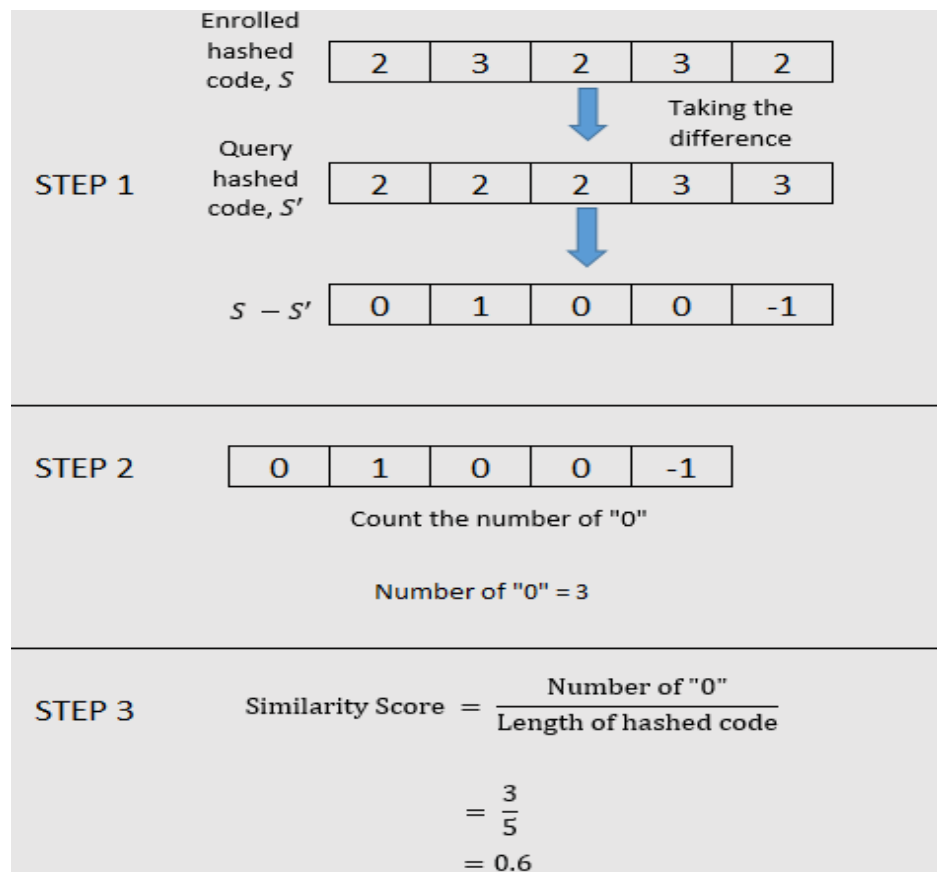


Figure 3.2: Example of similarity score computation for RBOMP hashing



### 3.3 Experimental Setup and Performance Analysis

Experiment was carried out using the database from NIST Speaker Recognition Evaluation 2004~2010, consisting 2001 speakers with 5 samples each. For intra-class comparison, there will be a total of 20,010 genuine matches whereas, for inter-class comparison there will be a total of 2,001,000 impostor matches. To avoid biasness from a single random binary orthogonal matrix and random token, the experiment was repeated for five times and the average equal error rate (EER) was obtained.

#### 3.3.1 Effect of $\omega$ and $k$ on the Recognition Performance of the Proposed Method

In this section, the effect of  $\omega$  and  $k$  on the EER is investigated. The number of binary orthogonal matrices,  $k$  is set to vary from 1000, 2000, 5000 and 10000 for different  $\omega$  settings (i.e.,  $\omega = 4, 5, 10, 11, 12$ ). Figure 3.3 shows the effect of different numbers of random binary orthogonal matrices and different lengths of window on the EER. The recognition performance of the system improves with the decrease in the value of  $\omega$  and the increase in the value of  $k$  due to more information available for the verification process to distinguish the speakers. Smaller values of  $\omega$  (i.e., lesser than 4) are not considered for security reasons to ensure that the adversary will not be able to obtain the order of the  $i$ -vector through attack-via-record multiplicity practically.

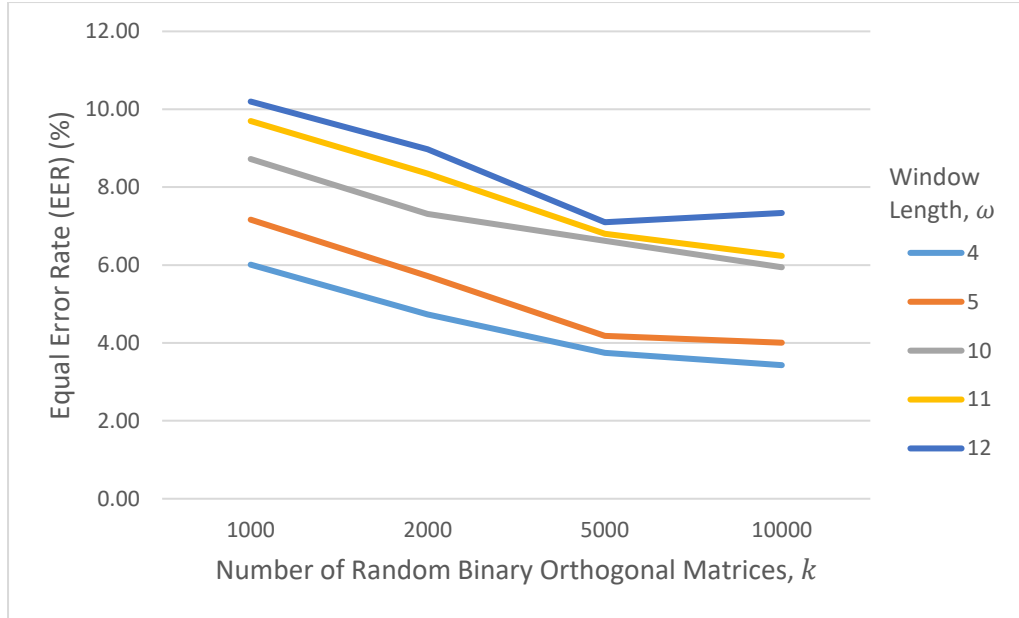


Figure 3.3: EER versus number of random binary orthogonal matrices for different lengths of window

### 3.3.2 Comparison of the Recognition Performances for Different Methods

The recognition performance of the proposed method is compared with other methods by selecting the lowest EER achieved from the experiments. Table 3.1 summarised the comparison results.

Table 3.1: Comparison of different speech template protection methods

Method	Database	Number of Speakers	Baseline EER (%) before template protection	Lowest EER (%) after template protection	Brute Force Attack Complexity (in bit)	Parallelizable	ARM Analysis
Universal Background Model	Text-Independent Digit Corpus	701	3.40	5.42	N.A.	×	×

Multi-bit Allocation	Text-Independent Digit Corpus	701	3.40	3.56	N.A.	✘	✘
Vaulted Voice Verification	MIT Mobile Device Speaker Verification Corpus	48	11.00	6.00	8 ~ 12 <sup>#</sup>	✘	✘
Random Projection + Fuzzy Vault	Mandarin Continuous Speech Recognition	40	2.22	2.22	O(1) <sup>†</sup>	✘	✘
<b>Proposed Method</b>	<b>NIST SRE 2004 ~ 2010</b>	<b>2001</b>	<b>1.67</b>	<b>3.43</b>	<b>40.24</b>	<b>✓</b>	<b>✓</b>

<sup>#</sup> Johnson *et al.* (2013) claimed that the attacker may gain access to the recognition system with a probability of  $2^{-8}$  to  $2^{-12}$  in addition to the  $n$ -bit of security offered by encryption. Under the lost key scenario, the encryption can be decrypted easily by the attacker and thus the security offered by encryption can be ignored.

<sup>†</sup> Zhu *et al.* (2012) claimed that the time complexity in launching a brute force attack to differentiate 32 genuine points out of total 332 points is  $\binom{332}{32} \approx 2^{148.11}$  assuming binary indices and key are kept secret. However, under the lost key scenario, the attacker can obtain the binary indices and key, thus the attacker can differentiate 32 genuine points out of total 332 points easily with negligible time complexity (i.e., O(1)).

Compared with other methods, the proposed method can offer strong security while preserving the recognition performance of the system with acceptable degradation (approximately 1.76% more in EER after template protection). The loss in accuracy is mainly due to the fact that there is less information used for verification as compared to the baseline system since

additional user-specific helper data is used in the baseline system. The lack of information has caused the loss of some discriminatory properties of the voice feature, hence degrades the recognition performance of the proposed method.

From Table 3.1, it shows that the proposed method is able to produce satisfactory recognition results. It is worth mentioning that the database used in this work consists of larger number of people, i.e., 2,001 people, as compared to other work. More importantly, NIST SRE series are widely used to measure the state-of-the-art speaker recognition systems. Different with other work, extensive analysis is conducted on the security of the proposed method against ARM (refer Section 3.5 for the details). Furthermore, inspired from the current trend of “parallelisable” (multi-core, pipeline, superscalar and vector), the proposed method can be parallelised given that each round function  $h$  is independent.

### **3.3.3 Comparison of the Recognition Performances for Different Databases**

To further justify the recognition ability of the proposed scheme, the experiment is carried out on two other datasets, namely Chinese Mandarin Speech Recognition Corpus – Digital String and Chinese Mandarin Speech Recognition Corpus – Conversation. Chinese Mandarin Speech Recognition Corpus – Digital String consists of 120 speakers with a total of 3,600 utterances while Chinese Mandarin Speech Recognition Corpus – Conversation consists of 943

speakers with a total of 5,780 utterances. Using the same parameter setting, where  $\omega = 4$  and  $k = 8000$ , the results of comparison is shown in Table 3.2.

Table 3.2: Comparison of recognition performance for different datasets

<b>Database</b>	<b>Baseline EER (%) before template protection</b>	<b>Lowest EER (%) after template protection</b>
NIST SRE 2004 ~ 2010	1.67	3.43
Chinese Mandarin Speech Recognition Corpus – Digital String	3.81	7.01
Chinese Mandarin Speech Recognition Corpus - Conversation	0.60	0.89

From Table 3.2, the results of the datasets suggest that the accuracy performance is well preserved for Chinese Mandarin Speech Recognition Corpus – Conversation and not large deterioration for Chinese Mandarin Speech Recognition Corpus – Digital String and NIST SRE 2004 ~ 2010. It is evident that the recognition performance of the proposed scheme is dependent on the quality of the voice feature extracted. Since Chinese Mandarin Speech Recognition Corpus is recorded from a clean and less noisy environment as compared to Chinese Mandarin Speech Recognition Corpus – Digital String, which consists of more noise, the former will have a better recognition performance (i.e., lower EER) as compared to the latter (i.e., higher EER).

### 3.4 Security Analysis

To ensure that the proposed RBOMP hashing fulfils the requirement of the biometric template protection, we further investigate the revocability and unlinkability properties of the RBOMP hashing as well as its security against brute force attack.

### **3.4.1 Revocability Analysis**

The revocability is evaluated by matching a particular hashed code with the other hashed codes generated from distinct random binary orthogonal matrices. A total of 100 hashed codes is derived from an  $i$ -vector with 100 different binary orthogonal matrices. The hashed codes are matched with respect to the first hashed code to compute the pseudo-impostor scores. The process is repeated using the same random token for different users to produce a total of  $99 \times 2 \times 2,001 = 396,198$  scores. The distribution of the genuine scores, impostor scores and the pseudo-impostor scores are computed using  $\omega = 4$  and  $k = 5000$  as shown in Figure 3.4. The difference in the number of scores computed for impostor and pseudo-impostor matching is because that in pseudo-impostor matching, we only focus on matching the first generated hashed code with other generated hashed code for each  $i$ -vector.

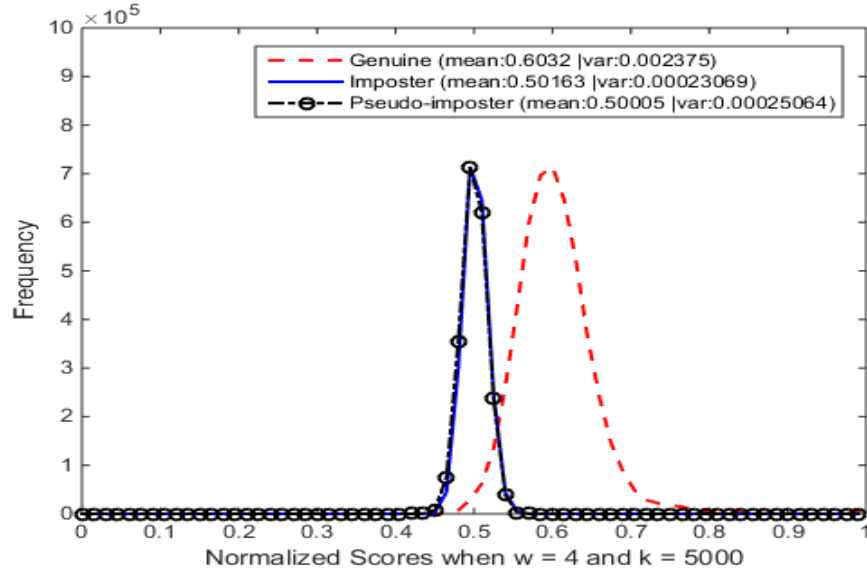


Figure 3.4: Distribution of genuine, imposter and pseudo-imposter scores

From Figure 3.4, we can notice that the pseudo-imposter scores distribution resembles the imposter scores distribution. This vindicates that the newly generated hashed codes are distinctive to each other although they are generated from the same  $i$ -vector. Since the newly generated hashed code is uncorrelated to the old hashed code, this justifies that RBOMP hashing has fulfilled the revocability criteria.

### 3.4.2 Unlinkability Analysis

The unlinkability is evaluated by introducing the pseudo-genuine scores. The pseudo-genuine score is computed by matching the hashed codes generated from different  $i$ -vector of the same user with different binary orthogonal projection matrices. Similar to the genuine matching, the pseudo-genuine match produces 20,010 scores. The overlapping of pseudo-imposter scores (from

Section 3.4.1) and pseudo-genuine scores will indicate whether the RBOMP hashed codes generated from whichever users are indistinctive. The hashed codes are considered to be *unlinkable* when it is difficult to differentiate them.

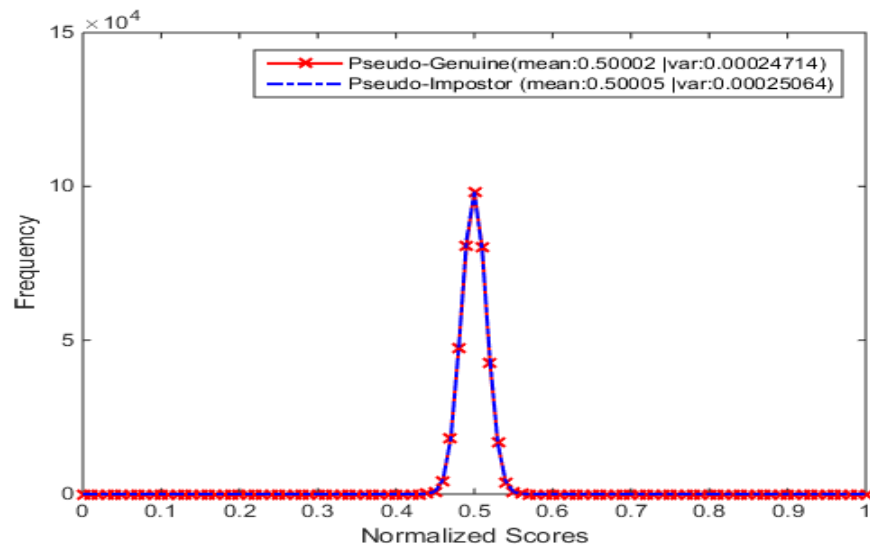


Figure 3.5: Distribution of pseudo-genuine scores and pseudo-impostor scores

From Figure 3.5, there is a large overlapping between the pseudo-genuine scores distribution and the pseudo-impostor scores distribution. Hence this suggests that the RBOMP hashed is able to fulfil the unlinkability property.

### 3.4.3 Brute Force Attack

As RBOMP projects the voice feature from linear space to ordinal space, it imposes strong non-invertible property to the system as it is computationally difficult for the adversary to recover the original feature value in linear space. From the distribution of the feature value ranging from a minimum of -7.0000



to a maximum of +7.0000, if the adversary wants to guess the correct value of an  $i$ -vector with a length of 500, the guessing complexity is of  $140000^{500} \approx 2^{8547}$  attempts. Even if the adversary wants to guess the rank of the biometric feature instead of the feature value itself, it would still require a guessing complexity of  $500! \approx 2^{3768}$  attempts.

The similarity score is computed based on the number of matches between the query hashed code and the enrolled hashed code. If the similarity score exceeds the threshold, the user will be deemed as the legitimate user. The threshold needed for a user to gain access to the system is approximately 0.55. Hence, using  $k = 5000$ , the adversary will require a minimum of  $5000 \times 0.55 = 2750$  correct matches in order to gain illegitimate access to the system. Thus, it requires an average time complexity of  $2^{2750}$  attempts to gain access to the recognition system.

Consider the scenario that the adversary does not compromise any templates, since there are only two possible outcomes for  $S_i$ , and each binary matrix is independent and uniformly distributed, one can assume that  $S$  follows binomial distribution with probability of 0.5. Let  $X$  denotes the number of correct guesses, the probability of obtaining 2,750 or more correct guesses can be computed as follows:

$$\begin{aligned}
 P(X \geq 2750) &= P\left(z \geq \frac{2750 - 2500}{\sqrt{1250}}\right) \\
 &= P(z \geq 7.071) \\
 &= 2^{-40.24} \tag{3.1}
 \end{aligned}$$

From Equation 3.1, the number of guesses required is  $2^{40.24}$ . This can be referred as the average time complexity required to gain access to the system without compromising any templates. To further improve the security of the recognition system, one can limit the number of login attempts.

### **3.5 Security Analysis against Attack-via-Record Multiplicity (ARM)**

ARM refers to a privacy attack whereby the attacker uses multiple compromised templates with or without the associated information such as the parameters and algorithms to recover the original biometric template (Scheirer and Boulton, 2007). In this work, the main concern will be on whether the adversary is able to guess the rank of the biometric feature. This is mainly due to the fact that guessing the rank of the biometric feature is relatively easier as there are lesser possibilities as compared to recovering the original feature value which is in real number domain. If the random token  $Z_i$  and the hashed code  $S$  are compromised, the adversary might be able to obtain the intermediate index,  $C_i$ , by observing the number of prime factors in  $Z_i$  and  $S$ . This is because the value of  $S$  is computed by taking the sum of the prime factors of  $Z_i * (C_i + 2)$ . If one has the knowledge of  $C_i$ , he can reconstruct the order of the biometric feature. Hence for security purposes, it is important to set the range of the window length,  $\omega$ , in such a way that there will be at least two possible values of intermediate indices,  $C$  mapped to each  $S$ . For instance, setting  $\omega = 4$  will allow  $C$  to have 4 possible values, which are 1, 2, 3 and 4. Therefore, the number of prime factors of  $C + 2$  will be 1, 2, 1 and 2 respectively. In this case, it can be seen that if  $S$  has the value of 1 then the possible value of  $C$  would be either 1 or 3. Meanwhile, if  $S$  has the

value of 2 then the possible value of  $C$  would be either 2 or 4. Here the value of  $Z_i$  is not taken into account since the value of  $Z_i$  will not affect the analysis. A clear graphical representation on the mapping of  $C$  to  $S$  is shown in Figure 3.6 using different setting of  $\omega$ .

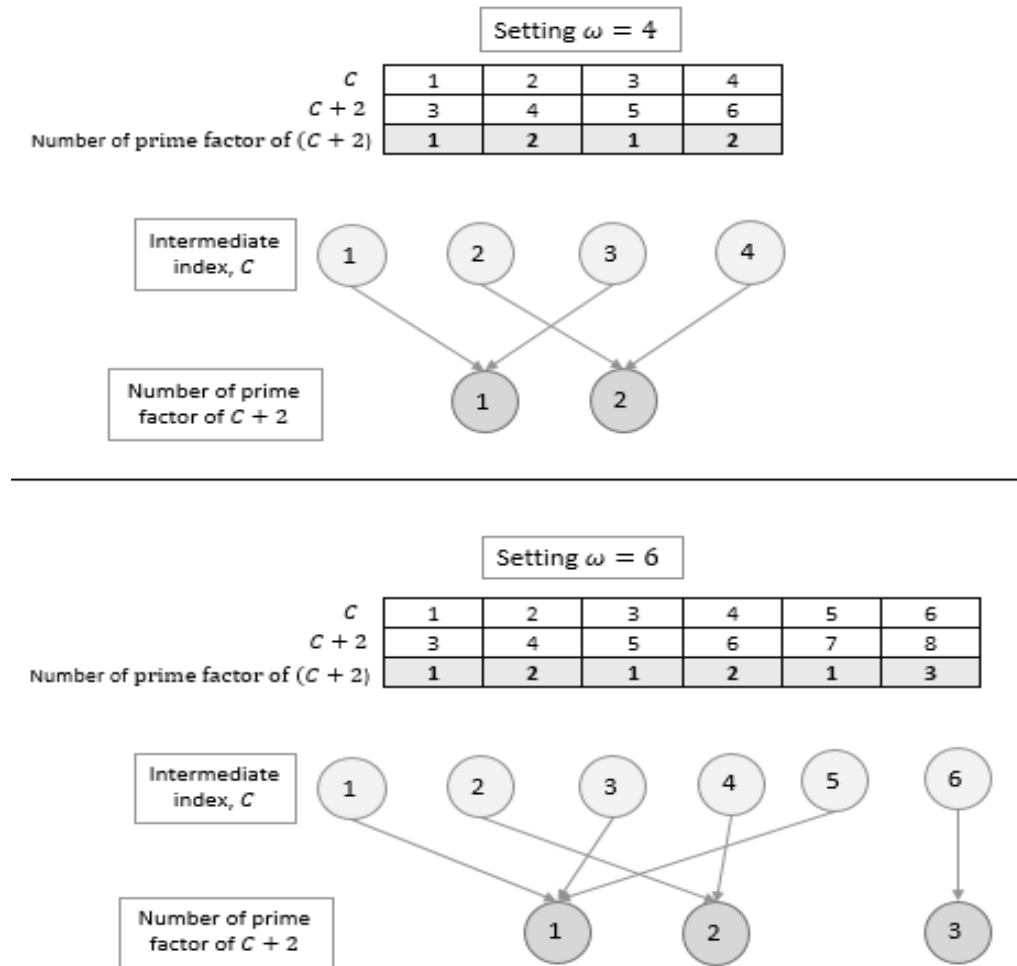


Figure 3.6: Mapping of intermediate index,  $C$ , at  $\omega = 4$  and  $\omega = 6$  respectively

As stated earlier, by observing the number of prime factors in  $Z_i$  and  $S$ , the adversary might be able to recover the value of  $C + 2$ . However, since there are two or more mappings to each value of  $S$ , (i.e., many-to-one mapping), there

will be an increase in complexity for the adversary to obtain the correct value of  $C$ . Figure 3.7 illustrates how the adversary might be able to recover the value of  $C$ .

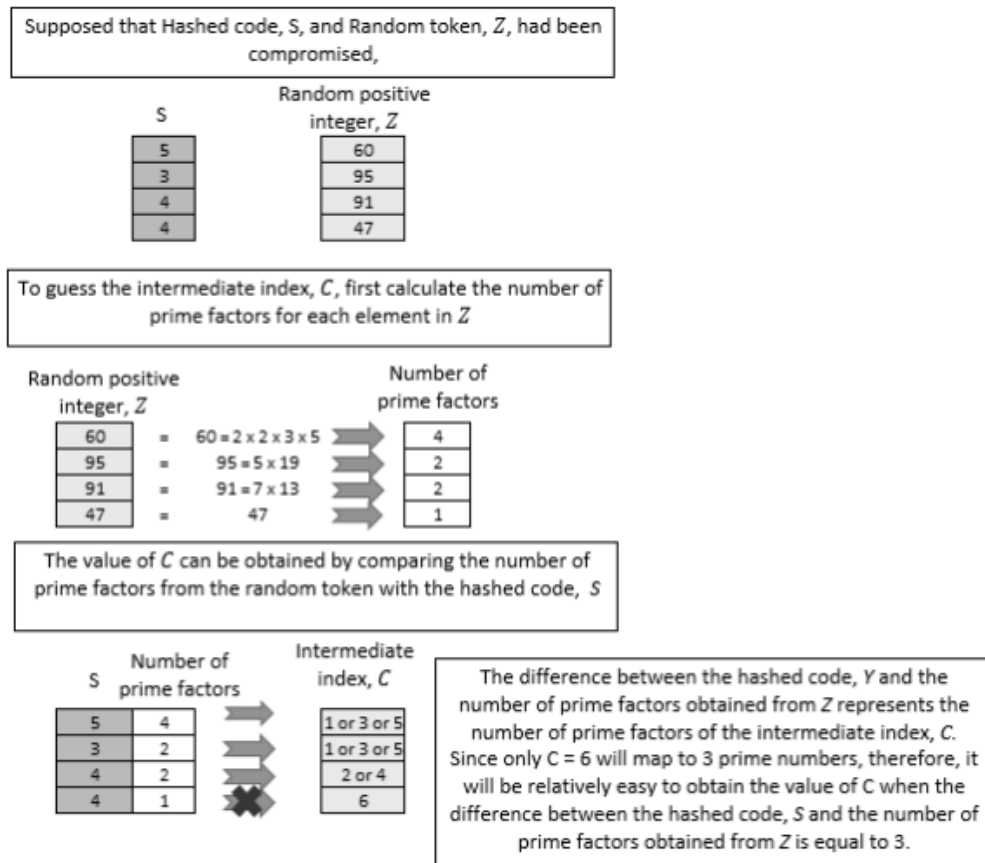


Figure 3.7: Guessing the index,  $C$ , at  $\omega = 6$  when both hashed code and random token are compromised

In this work, using  $\omega = 4$  and the minimum number of random binary matrices,  $k$  is set to 5000, the naïve approach for the adversary to recover the correct value of  $C$  would be  $2^{5000}$  as for each round, there will be two possibilities of  $C$ . However, in real world scenario, the adversary would require much lesser than  $2^{5000}$  attempts. Given the threshold of acceptance is 0.55 (as given in Section 3.4.3), using false accept attack, the adversary only needs to

guess 2750 bits correctly. Since the probability of guessing a bit correctly is 0.5, one would expect the adversary to obtain 2500 correct guesses (out of 5000 guesses) on average. In other words, the adversary would only require another 250 correct guesses to gain access to the system. Given that the  $S_i$  follows binomial distribution, the average time complexity to access the system will be  $2^{40.24}$  as stated in Section 3.4.3.

Given the worst case scenario that the adversary will always obtain the binary matrices he desired, the adversary may require lesser number of binary matrices to derive the order of the  $i$ -vector. Figure 3.8 shows how the adversary is able to obtain the order of  $i$ -vector using the desired binary matrices.

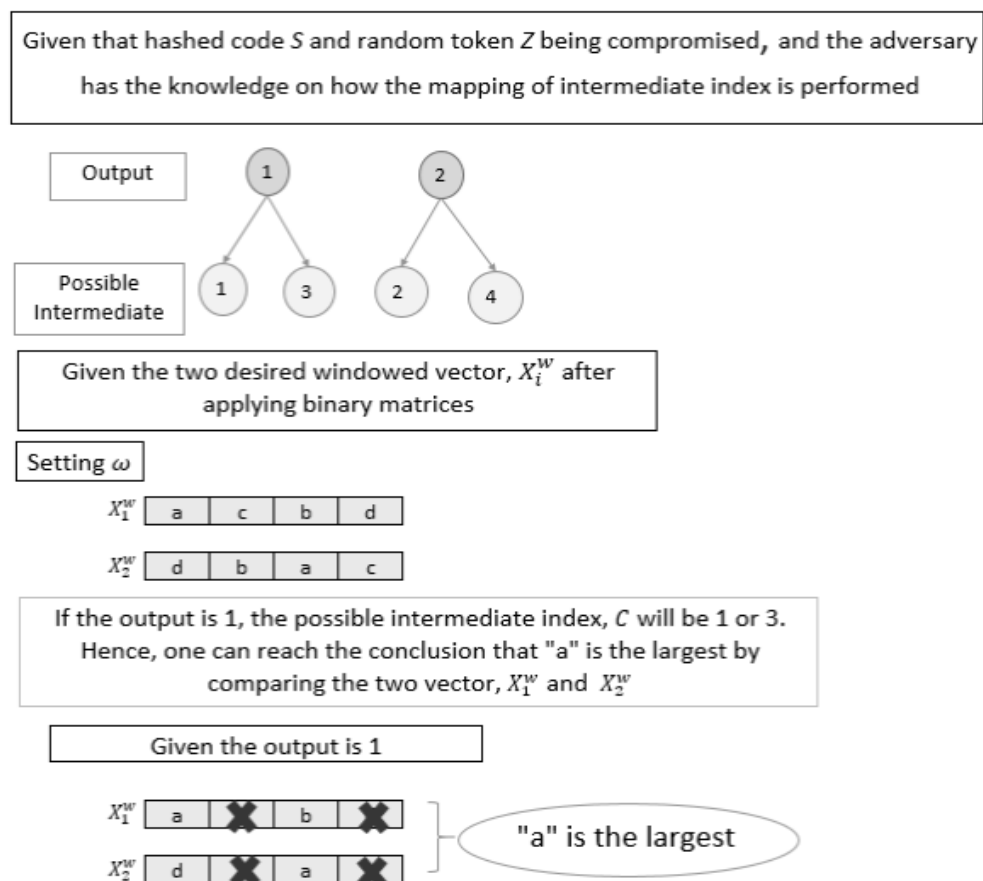


Figure 3.8: Deriving the order of  $i$ -vector by comparing two desired windowed vector

Using Figure 3.8 as an example, given that the adversary has compromised one binary matrix and the number of prime factor of  $(C + 2)$  is 1, the adversary will be able to recover the windowed vector,  $X_1^w$ , after applying the compromised binary matrix. By using the information of  $X_1^w$  and number of prime factor of  $(C + 2)$ , the adversary will conclude that the possible position for the largest value among the four elements in  $X_1^w$  will be at position 1 or 3. In other words, the elements in position 1 or 3 are the candidates for the largest value of among these four elements. Next, in order to determine the correct position of the largest value, he will need to compare with another windowed vector,  $X_2^w$ , where  $X_2^w$  can be obtained after applying a different binary matrix, or here we define as the desired binary matrix. The desired binary matrix is determined in such a way that it will result in  $X_2^w$  having the same four elements as  $X_1^w$  and there will be only one common element between the candidates for the largest value of  $X_1^w$  and  $X_2^w$  (i.e., the candidates for largest value of  $X_1^w$  are {"a", "b"} and the candidates for largest value of  $X_2^w$  are {"d", "a"}). The common element between this two sets is "a" and hence "a" can be determined as having largest value among these four elements). For any compromised binary matrix, there will be a total of 16 desired binary matrices that can be used to derive the largest value of the elements. As the length of  $i$ -vector is 500, there will be  $\binom{500}{4} \approx 2^{31.26}$  unique binary matrices. Under the situation where the adversary will always obtain the desired binary matrices, it will only require  $\text{him} \frac{2^{31.26}}{5000} \approx 2^{18.97}$  minimum number of templates to reconstruct the full order of the  $i$ -vector.

However, in real-world scenario the adversary will not always obtain the binary matrices that he wants. Hence if given the knowledge of all the possible binary matrices that can be generated, or here we refer as the distinct binary matrices, the adversary will be able to find a suitable pairing of the distinct binary matrices he needs and derives the order of the  $i$ -vector. Therefore, the focus of the issue will be on how many attempts are required for the adversary to obtain all the distinct binary matrices. This scenario can be reduced to the coupon collector's problem (Ferrante and Saltalamacchia, 2014). The coupon collector's problem is a probability problem where it describes the probability or the expected trials required to collect all different coupons from a finite set with replacement. In our case, the distinct binary matrices that provide useful information can be viewed as the coupon as the adversary are required to obtain all different useful binary matrices to reconstruct the ordering of  $i$ -vector. As each of the binary matrices are uniformly generated, one can say that each binary matrix is equally likely to be obtained at any time with a probability of  $\frac{1}{m}$ , where  $m$  is the total number of distinct matrices,  $2^{31.26}$ . Let  $X_i$  be the random variable for the number of trials required to complete the order of  $i$ -vector, and the probability of obtaining a new distinct binary matrix will be  $\frac{m-i+1}{m}$ , where  $i$  is in the range from  $[1, m]$ . By the assumption of independence,  $X_j, j \in \{1, m\}$ , is independent to each other and it follows a geometric distribution with the parameter,  $p = \frac{m-j+1}{m}$ . The expected number of trials of a particular distinct matrix,  $E(X_j)$ , can be computed by using the formula  $\frac{1}{p}$  and hence taking the

sum of all the number of expected trials of distinct matrices, the total expected number of trials,  $E(X)$ , needed to obtain  $m$  distinct matrices will be as follows,

$$X = X_1 + X_2 + X_3 + \dots + X_{m-1} + X_m \quad (3.2)$$

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + E(X_3) \dots + E(X_{m-1}) + E(X_m) \\ &= \frac{m}{m} + \frac{m}{m-1} + \frac{m}{m-2} + \dots + \frac{m}{2} + m \\ &= m \sum_{i=1}^m \frac{1}{i} \end{aligned} \quad (3.3)$$

Using approximation formula,

$$E(X) = m(\log m + \gamma + \frac{1}{2m} + o(\frac{1}{m^2})) \quad (3.4)$$

where  $\gamma \approx 0.5772156649$  is the Euler-Mascheroni constant. From Equation 3.4, given  $m = 2^{31.26}$ , the expected number of trials needed to obtain  $m$  distinct binary matrices is around  $2^{35.70}$ . Figure 3.9 shows the expected number of trials needed to collect  $m$  distinct binary matrices.



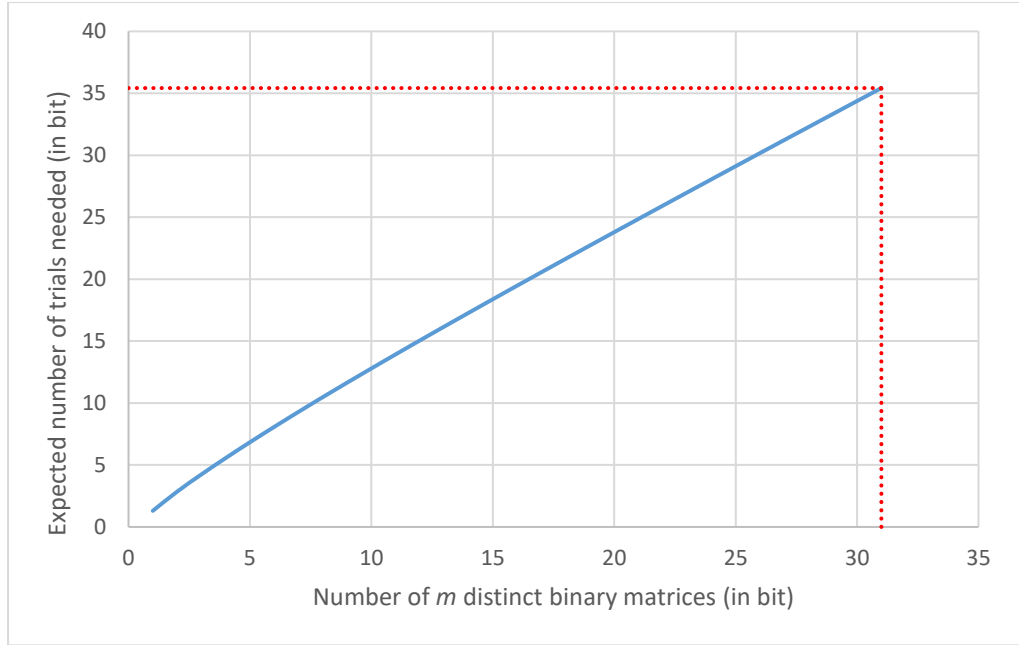


Figure 3.9: Expected number of trials needed to collect  $m$  distinct binary matrices

Hence, the expected number of templates required for the adversary to obtain all the distinct binary matrices will be  $\frac{2^{35.70}}{5000} \approx 2^{23.41}$ . However, it is impractical for an adversary to compromise  $2^{23.41}$  templates. This can be referred as the data complexity required to gain access to the system with probability of one where the time complexity is negligible.

To relax the data complexity and the time complexity, we consider the scenario where the adversary only requires to compromise a small number of templates to gain access to the system with smaller time complexity. Given that the adversary is able to compromise some number of templates, the adversary will be able to deduce some correct number of  $S_i$  from the compromised templates and guess the remaining number of  $S_i$  to gain access to the system. Thus, the adversary can launch the false accept attack (i.e. described in Section

3.4.3) with lesser time complexity. Table 3.3 shows the relationship of the number of intermediate indices from the templates with the corresponding time complexity required to access to the system.

Table 3.3: Comparison between number of intermediate indices compromised versus time complexity required to access the system

Number of Intermediate Indices Compromised from Template	Remaining Number of Intermediate Indices Required to Guess	Remaining Number of Intermediate Indices Required to Access the System	Mean	Std dev	Time Complexity (in bit)
100	4900	2650	2450	35.00	27.4357487
200	4800	2550	2400	34.64	17.034102
300	4700	2450	2350	34.28	9.145804742
400	4600	2350	2300	33.91	3.832707225
500	4500	2250	2250	33.54	1
600	4400	2150	2200	33.17	Negligible
700	4300	2050	2150	32.79	Negligible
800	4200	1950	2100	32.40	Negligible
900	4100	1850	2050	32.02	Negligible

In our work, the intermediate indices,  $C$ , used is 5000 and the threshold to access the system is 0.55. In other words, the adversary will require 2750 correct guesses to access the system. Hence the remaining number of intermediate indices,  $C$ , required to access the system will be referring to how many guesses are needed in order to exceed the threshold (2750 correct guesses) after certain number of  $C$ s have been compromised. Assuming the remaining number of intermediate indices required to access the system follows binomial distribution with the parameter,  $p = 0.5$ , the remaining number of intermediate indices required to guess is denoted as  $N$  and the number of correct guesses is denoted with  $X$ . The time complexity is calculated by taking the reciprocal of

the probability of  $X$  greater than or equal to the remaining number of intermediate indices,  $C$ , required to access the system, with mean of  $N*p$  and standard deviation of square root of  $N*p*(1-p)$ . From Table 3.3, it can be seen that the time complexity is negligible if the number of intermediate indices,  $C$ , compromised is at least 500. Hence one can expect the adversary will gain access to the system with probability of one once he/she has compromised at least 500 intermediate indices.

In order to obtain the minimum number of templates required to compromise at least 500 intermediate indices (as the time complexity is negligible after 500 intermediate indices are compromised), an experiment is carried out to determine the minimum number of templates required to compromise for different data size. Experiment is performed on a system consisting of 48GB RAM running on Ubuntu OS. The procedure of the experiment is as follows:

1. An empty array is created consisting of  $m$  rows and 24 columns (as there are a total of 24 possible binary matrices that will result in the windowed vector consisting the same four elements), where  $m$  is the data size.
2. Mersenne Twister pseudorandom number generator (Matsumoto and Nishimura, 1998) is used to generate a pseudorandom number and the occurrence of the random number is marked on the empty array.
3. The 24 columns of the empty array are divided into three groups consisting of eight columns each, namely  $P$ ,  $Q$  and  $R$  in such a way that a collision is only considered when there is at least a pairing of

occurrences of different groups of the same row (i.e., elements in group  $Q$  and  $R$  can be referred as the desired binary matrices for the elements in group  $P$ ). As explained previously, for any compromised binary matrix, there will be a total of 16 desired matrices that can be used to derive the largest value of the four elements. Here, the pseudorandom number symbolizes that a particular binary matrix is compromised and along with a desired matrix (another random number in a different group), the adversary will be able to derive the largest value (referred as collision in this experiment).

4. Steps 2 to 4 are repeated and number of runs are recorded when the number of collisions reached 500.
5. The experiment is repeated for five times to obtain the average number of runs needed to obtain 500 collisions.
6. The minimum number of templates required to be compromised is computed by dividing the number of runs with 5000 (as a template consists of 5000 runs).
7. The experiment is repeated for different data sizes,  $m$ , and the minimum number of templates needed is recorded.

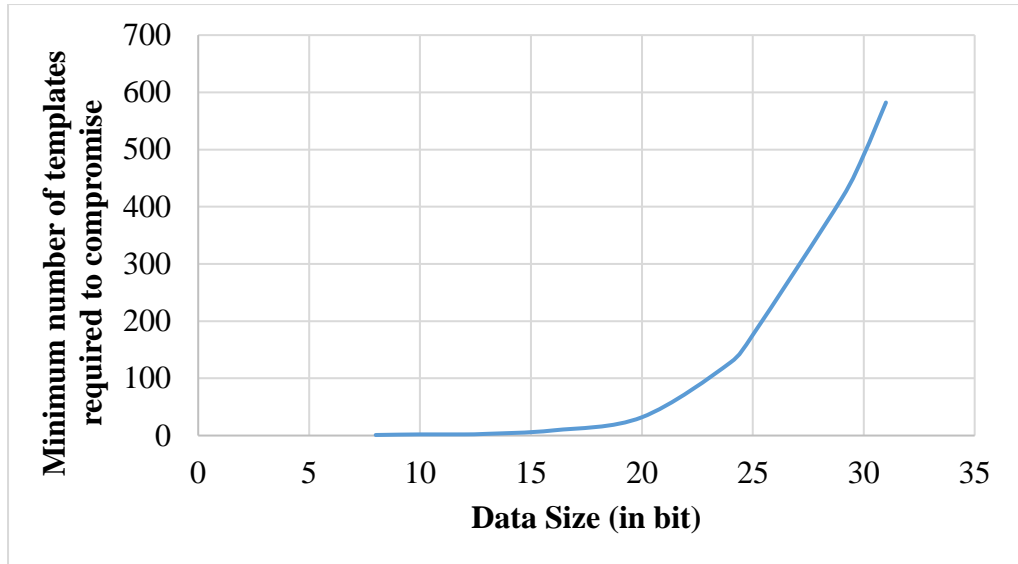


Figure 3.10: Minimum number of templates required to compromise to access the system for different data size

As we have  $m = 231.26$  unique matrices (i.e. obtained from  $\binom{500}{4}$ ), the minimum number of templates required to be compromised is approximately 600 as shown in Figure 3.10. However, for any person, he or she may not store up to 100 biometric templates across different applications in his or her lifetime, let alone 600 templates. Thus, it is impractical to compromise 600 templates from different database systems in the real world.

### 3.6 Summary

The proposed RBOMP hashing is able to achieve a reasonable degradation with an EER of 3.43% as compared to 1.67% without template protection, and it has also been shown that the performance of the proposed system is dependent on the underlying quality of the voice recording. In the security aspect, it has been shown that the proposed RBOMP hashing has a security complexity of 40.24 bits and it has been demonstrated that the hashing is able to survive against

ARM practically, with a requirement of 600 different templates across different applications, to reconstruct the order of the biometric feature, which is practically impossible. Furthermore, it has also been justified that the proposed RBOMP hashing is able to satisfy the biometric template protection criteria, such as unlinkability and revocability.

## CHAPTER 4

### TWO-DIMENSIONAL WINNER-TAKES-ALL HASHING (2DWTA)

#### 4.1 Introduction

The RBOMP hashing has shown to be able to preserve the performance of the system as well as survive against ARM. However, in the perspective of information security, the security strength of 40.24 bits in RBOMP hashing may not be sufficient. In addition to that, the drawback of RBOMP hashing is that it is only suitable for one-dimensional input vector. Hence, to solve the limitations of RBOMP hashing, we take one step further to propose a two-dimensional winner-takes-all hashing (2DWTA). Similar to RBOMP hashing, 2DWTA hashing also transforms the feature from continuous value to discrete value. However, different from RBOMP hashing, 2DWTA is intended for two-dimensional feature input (i.e., matrix form). To further enhance the security of the system, an additional biometric modality, namely fingerprint, is incorporated and is fused with voice modality at feature level.

The remainder of the chapter is organised as follows. In Section 4.2, the specification of 2DWTA is presented. Section 4.3 describes the experimental setup and provides performance analysis of 2DWTA hashing. Section 4.4 outlines the security analysis of 2DWTA hashing. Lastly, Section 4.5 gives a brief summary of this chapter.

## **4.2 The Specification of 2DWT**

### **4.2.1 Architecture of the Multimodal Biometric System**

A general structure of the proposed scheme is illustrated in Figure 4.1. Firstly, at the enrolment stage, the fingerprint and voice sample of the user are both captured and go through different feature extraction processes. Feature level fusion is then performed by combining the extracted fingerprint feature and voice feature. Next, 2DWT is applied on the fusion of fingerprint feature and voice feature. Finally, the output is stored as the template into the database. During verification, similar procedure is conducted to generate the query template. The matching module will then match the query template with the enrolled template stored in the database. Authentication is successful when the similarity score obtained from the enrolled template with the query template exceeds the predefined threshold.



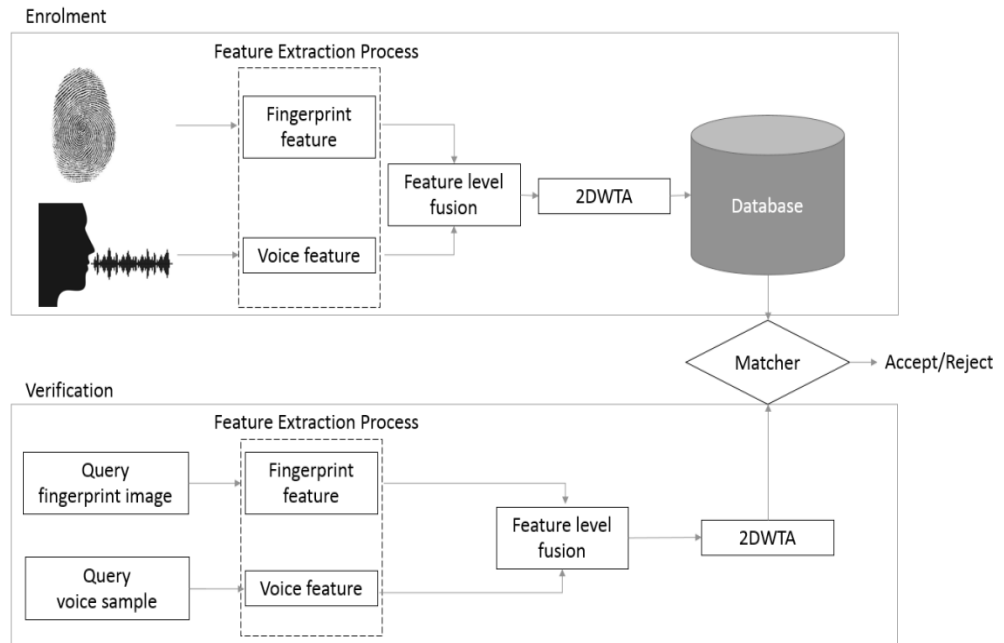


Figure 4.1: Overview of the proposed feature level fusion template protection method

#### 4.2.2 Fusion of Fingerprint and Voice Templates

The fingerprint feature and voice feature are first extracted and fused together via matrix multiplication. Given that the fingerprint feature,  $f$ , represented in the form of a row vector with a dimension of 299 and the voice feature (i.e.,  $i$ vector),  $v$ , represented in the form of a column vector with a dimension of 500, the matrix multiplication of these two features can be written as  $M = v \cdot f$ , where  $M$  denotes the matrix produced after fusion and  $\cdot$  is the matrix multiplication. We refer it as the fusion matrix with dimension of  $500 \times 299$ . As both of the biometric features are normally distributed (i.e., fingerprint features have a mean of 0 and standard deviation of 0.02, while voice features have a mean of 0.04 and standard deviation of 1.89), the biometric features will not dominate each other and hence no normalisation technique is required. Figure 4.2 shows

the fusion process of voice feature and fingerprint feature via matrix multiplication. The fusion of the biometric features will produce a fusion matrix,  $M$ , with high dimensionality. Hence, to overcome the “curse of dimensionality” issue, two-dimensional Winner-Takes-All hashing (2DWTA) is proposed. As stated by Daum and Huang (2003), the increase in the volume of space will lead to sparse useful information. Therefore, it was suggested by Yagnik et al. (2011) that the precise value of the feature is not needed when dealing with data with high dimensionality. Motivated by the fact that the precise value of feature is less important in classification problem, 2DWTA hashing defines an ordinal embedding with an associated rank-correlation measure. Transformation to ranking space offers a degree of invariance with respect to the perturbation of the numeric values (Dean et al., 2013). In other words, 2DWTA focuses on the implicit ordering of the feature instead of the absolute feature value of the features. Hence, this makes the fusion matrix resilient to insignificant changes in the feature value itself. In addition, similar to ranking-based hashing (Yagnik et al., 2011), 2DWTA ranks the random permutations of the input feature and uses the index of the maximum feature value to encode the compact representation of the input feature. The index of the maximal feature value is used since feature with higher ranking in a list consists of more significant information as compared to those lower rankings in the list (Yagnik et al., 2011).

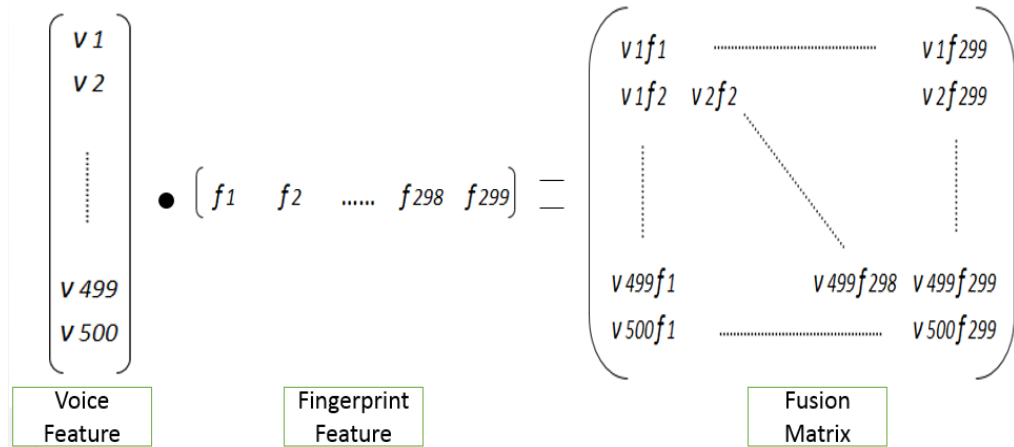


Figure 4.2: Fusion of voice template and fingerprint template at feature level via matrix multiplication

### 4.2.3 Two-dimensional Winner-Takes-All Hashing (2DWTA)

The limitation of WTA and RBOMP is that they cannot be applied directly to data that are in the form of matrix as WTA and RBOMP were initially intended for data that are in the form of vector representation. Hence, to address this issue, a modified version of WTA coined as 2DWTA is proposed. The procedures of 2DWTA are listed as follow:

- i) **Random Permutations.** The row dimension of the fusion matrix,  $\mathbf{M}$ , is first randomly permuted followed by the column dimension to produce  $\mathbf{M}^P$ .
- ii) **Selection of the first  $k$  items from both row and column dimensions.** The first  $k$  items are selected from both row and column dimensions of  $\mathbf{M}^P$  for  $2 \leq k \leq n - 1$  and let  $n = \min(p, q)$ , where  $p$  and  $q$  denote the length of fingerprint vector,  $\mathbf{f}$  and voice vector,  $\mathbf{v}$  respectively. This step reduces the dimensionality of the matrix and hence there will be information loss

during this stage. At the end of this step, a square matrix with dimension of  $k \times k$  will be produced.

- iii) Record index of the highest value. The index of the highest value from the  $k \times k$  dimension square matrix is recorded in the form of an ordered pair (i.e., (row index, column index)).
- iv) Repeat Step i to Step iii using  $h$  different permutation sequences for both row and column dimensions, where  $\theta^r_h$  and  $\theta^c_h$  represent the permutation sequences for row dimension and column dimension respectively. A series of ordered pairs,  $C_i$  will be generated, where  $i \in [1, h]$  and let  $S$  be the set consisting the generated  $C_i$  (i.e.,  $S = \{C_1, C_2, \dots, C_h\}$ ).

Notice that different permutation seeds are selected for different rounds. In real world scenario, the selection of permutation seeds for both row and column dimensions are user specific. In the event that both of the permutation seeds are compromised, the user is able to revoke and reissue a new template by generating different permutation seeds to replace the old template. Figure 4.3 illustrates the process of one-round 2DWTA.

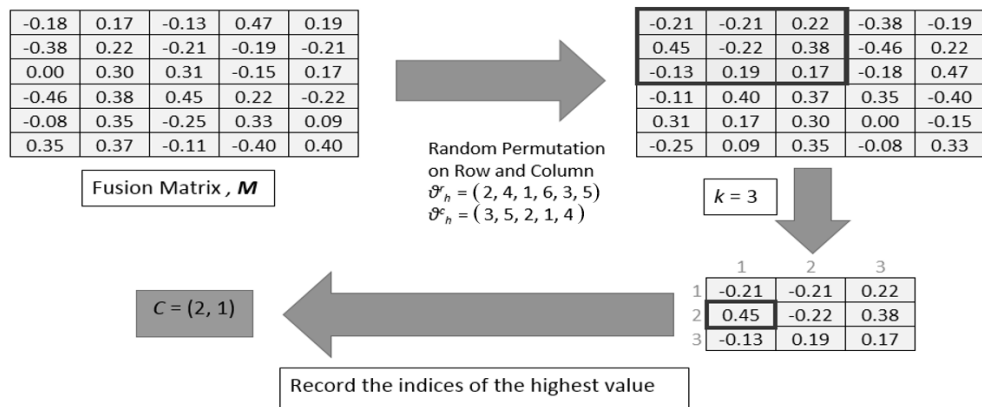


Figure 4.3: Process of one round 2DWTA

#### 4.2.4 Matching

Similar to the matching procedure in Section 3.2.2, the similarity score is produced by counting the number of zeros after performing elemental-wise subtraction between two hashed codes. Note that for 2DWTA, the similarity scores are computed by taking the total number of “0” over the total number of elements in the hashed code (an ordered pair of  $C_i$  will have two elements which are the row index and column index respectively) instead of the length of hashed code. Figure 4.4 illustrates the matching procedure of 2DWTA.

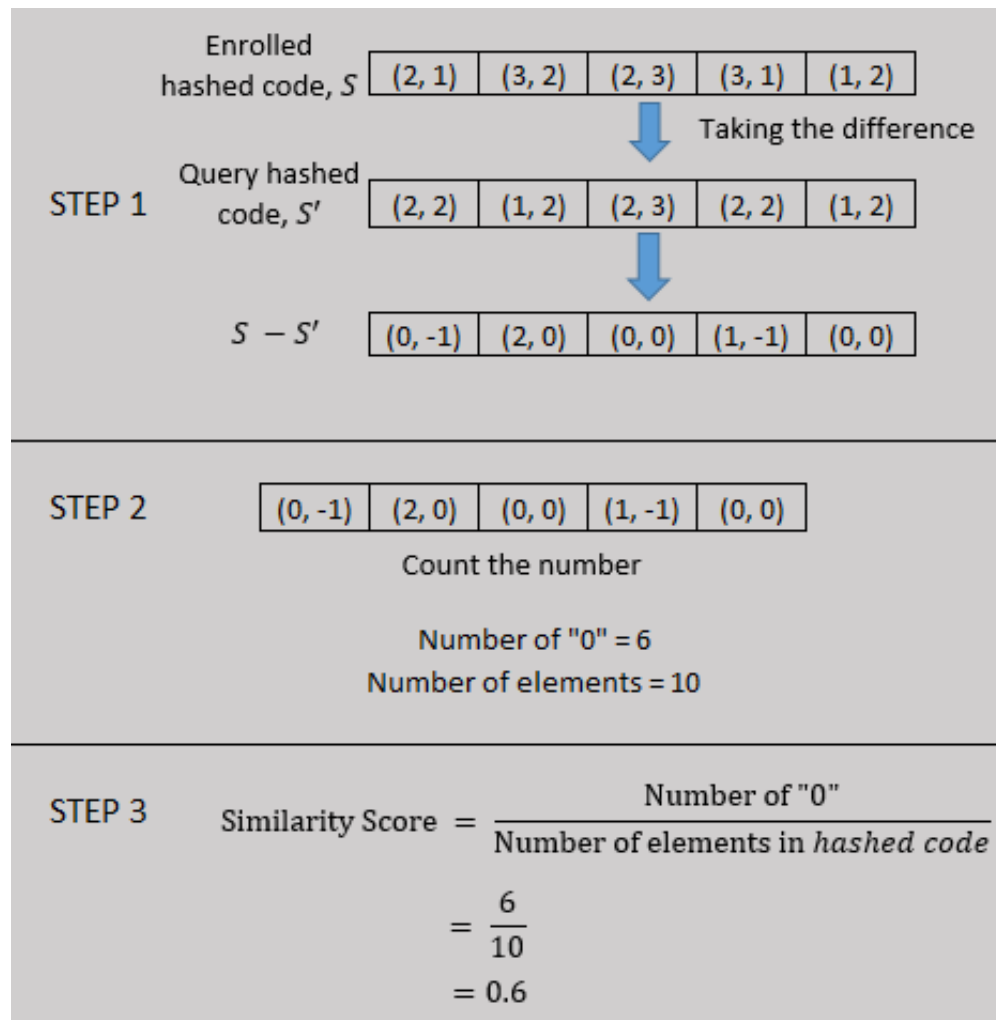


Figure 4.4: Example of similarity score computation for 2DWTA

### 4.3 Experimental Setup and Performance Analysis

In the experiment, there are a total of three fingerprint datasets and one voice dataset being used. The fingerprint datasets are FVC2002 DB1, FVC2002 DB2 and FVC2002 DB3 while the voice dataset is NIST Speaker Recognition Evaluation (SRE) 2004 ~ 2010.

The details and the denotations of the database used are as follow:

- V*: Voice dataset NIST Speaker Recognition (SRE) 2004 ~ 2010 consists of 2001 speakers and each with five samples.
- F1*: Fingerprint dataset FVC2002 DB1 consists of 100 users and each with eight samples.
- F2*: Fingerprint dataset FVC2002 DB2 consists of 100 users and each with eight samples.
- F3*: Fingerprint dataset FVC2002 DB3 consists of 100 users and each with eight samples.

The fingerprint features from 100 users are selected and assigned to each speaker to result in 100 different individuals with different voice and fingerprint features. As a summary, the proposed scheme is evaluated with these dataset combinations:  $\{V, F1\}$ ,  $\{V, F2\}$  and  $\{V, F3\}$ . In real-world or ideal scenario, all individuals will have different sets of permutation seeds (both row and column

dimensions). However, throughout this experiment, we assume the worst case scenario where the permutation seeds have been stolen by the adversary. To simulate the worst case scenario or stolen token scenario, all of the individuals are assumed to share the same permutation seeds. The performance and security evaluation are then conducted under this worst case scenario.

For intra-class comparison, there are a total of 1000 genuine matches while for inter-class comparison there are a total of 4950 impostor matches. In addition, to avoid any biasness of the results obtained from a single set of permutation seeds, the experiment for each parameter is repeated for five times and the average equal error rate (EER) is recorded.

#### 4.3.1 Effect of $k$ and $h$ on the Recognition Performance of the Proposed Method

This section evaluates the effect of  $k$  and  $h$  on the performance of the system. The value of  $k$  is set to be varied from  $k = 2, 5, 10, 20, 50$  and  $100$  for different  $h$  setting (i.e.  $h = 1000, 2000, 5000$  and  $8000$ ). The results of the parameters are presented in Table 4.1, Table 4.2 and Table 4.3 for the respective datasets.

Table 4.1: Performance of different parameters on dataset combination of  $V$  and  $F1$  in EER (%)

	$k = 2$	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 100$
$h = 1000$	8.00	2.82	1.46	1.26	<b>0.93</b>	1.77
$h = 2000$	4.81	2.24	1.73	1.22	1.02	1.57

$h = 5000$	2.53	1.82	1.53	1.33	0.94	1.65
$h = 8000$	2.17	1.87	1.63	1.13	0.94	1.84

Table 4.2: Performance of different parameters on dataset combination of  $V$  and  $F2$  in EER (%)

	$k = 2$	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 100$
$h = 1000$	8.44	4.36	4.12	2.28	1.90	2.26
$h = 2000$	7.29	5.41	4.01	2.32	<b>1.82</b>	2.28
$h = 5000$	3.84	4.47	4.30	2.25	<b>1.82</b>	2.28
$h = 8000$	3.26	4.51	3.37	2.65	1.86	2.23

Table 4.3: Performance of different parameters on dataset combination of  $V$  and  $F3$  in EER (%)

	$k = 2$	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 100$
$h = 1000$	15.19	9.13	8.70	7.36	6.23	6.02
$h = 2000$	11.79	8.29	8.05	7.08	6.34	6.53
$h = 5000$	5.99	7.12	7.94	7.22	6.36	6.46
$h = 8000$	<b>4.54</b>	7.42	7.82	7.22	6.33	6.35

Notice that, in general, the increase in the number of rounds,  $h$ , will improve the performance of the system (i.e., lower EER). This is due to the fact that more information will be available for the system to distinguish different users. As the size of the window dimension,  $k$ , increases, there will be no significant changes in the EER thereafter. From the results obtained, the optimal parameter for the multimodal biometric system to achieve the best performance



is when  $k = 50$  and  $h = 8000$ . For security concerns, we do not consider small value of  $k$ , although the datasets achieve better performance at smaller value of  $k$  (i.e., in the case of dataset combination of  $V$  and  $F3$ ). This is mainly to ensure that the adversary will not be able to guess and reconstruct the order of the biometric features easily.

### 4.3.2 Comparison of Results

In this section, the performance of the proposed scheme is validated against the unimodal biometric system. From Table 4.4, it can be seen that the feature level fusion of biometric is able to preserve the performance of the unimodal biometric system. This is mainly due to the fact that the loss of the discriminative power of the biometric features is minimal as matrix multiplication between the two biometric features is just the scaling of one of the biometric feature vector to the element of the feature vector of another biometric modality (i.e., referring to Figure 4.2, the first row of the fused matrix is actually the scaling of the fingerprint vector to the factor of  $v_1$ ). Besides that, the fusion of these two biometric modalities also provides added security to the recognition system as the adversary will not be able to access to the system without prior knowledge on both of the biometric features.

Table 4.4: Performance comparison between unimodal biometric and WTA in stolen token scenario

Datasets	Fused with	EER(%)
$V$	-	4.50

<i>V</i>	<i>F1</i>	0.94
<i>V</i>	<i>F2</i>	1.86
<i>V</i>	<i>F3</i>	6.33
<i>F1</i>	-	0.93
<i>F1</i>	<i>V</i>	0.94
<i>F2</i>	-	1.66
<i>F2</i>	<i>V</i>	1.86
<i>F3</i>	-	4.11
<i>F3</i>	<i>V</i>	6.33 <sup>1</sup>

---

#### 4.4 Security Analysis

Similarly, to ensure that the proposed 2DWTA hashing fulfils the requirement of the biometric template protection, the security as well as revocability and unlinkability properties of 2DWTA are investigated.

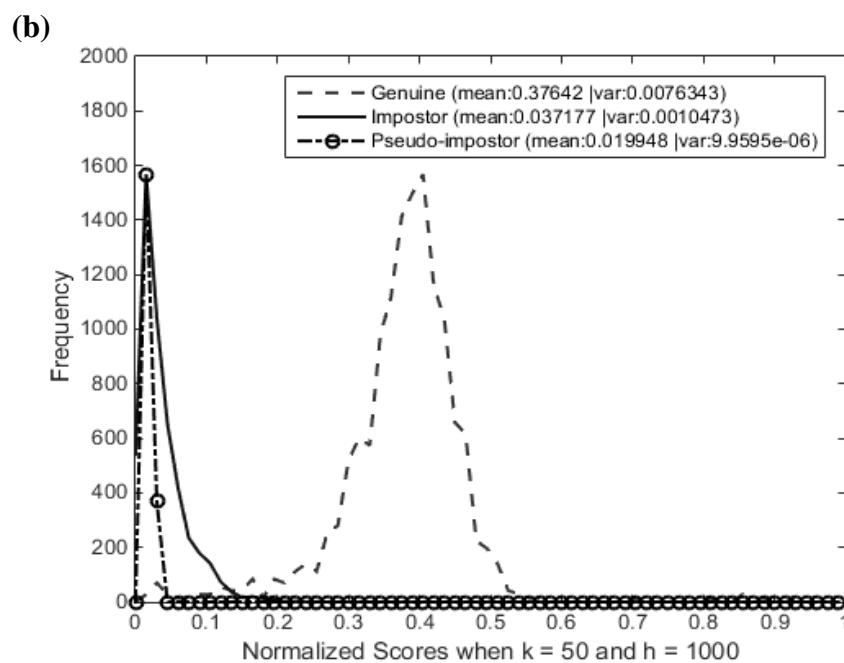
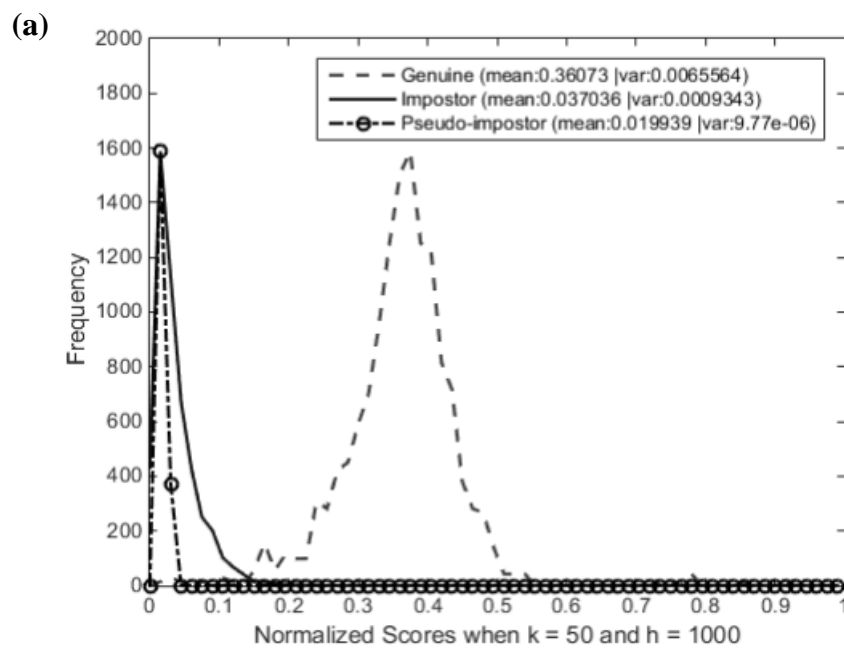
##### 4.4.1 Revocability Analysis

In the event of template being compromised by the adversary, the user should be able to generate a new template derived from the same biometric trait to replace the compromised template. To evaluate this criterion, a total of 100 hashed codes derived from the same fusion matrix,  $M$ , are generated and the first generated hashed code is matched with other generated hashed code to

---

<sup>1</sup> The best performance achieve is 4.54% when  $k = 2$  and  $h = 8000$ . However, due to security concerns, we do not consider the smaller value of  $k$ .

produce the pseudo-impostor scores. The 100 hashed code are generated using 100 different permutation seeds and these permutation seeds will be used for different users (i.e., same 100 different permutation seeds are applied to all users) to produce a total of  $99 \times 5 \times 100 = 49500$  pseudo-impostor scores for each datasets. The distribution of the genuine scores, impostor scores and pseudo-impostor scores for the respective combination of datasets are presented as shown in Figure 4.5.



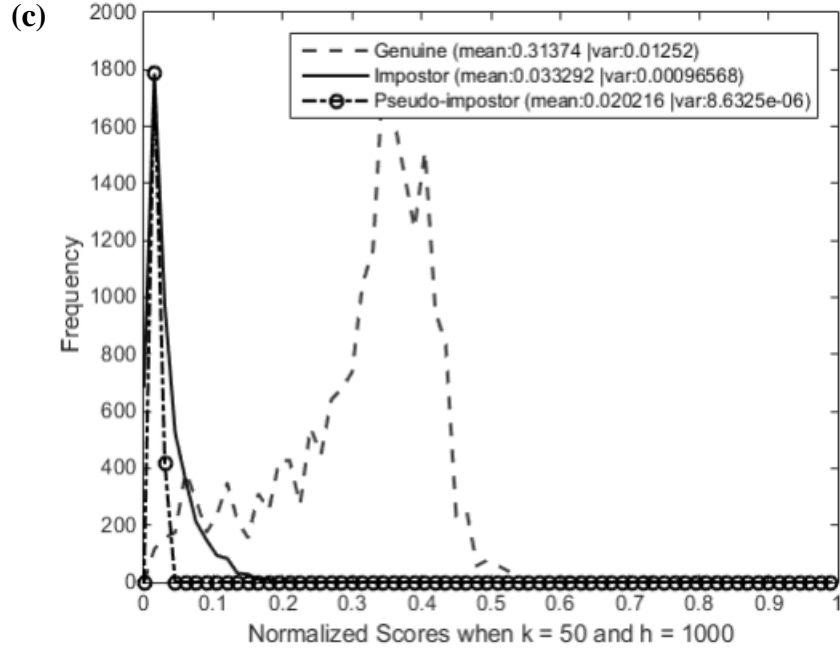


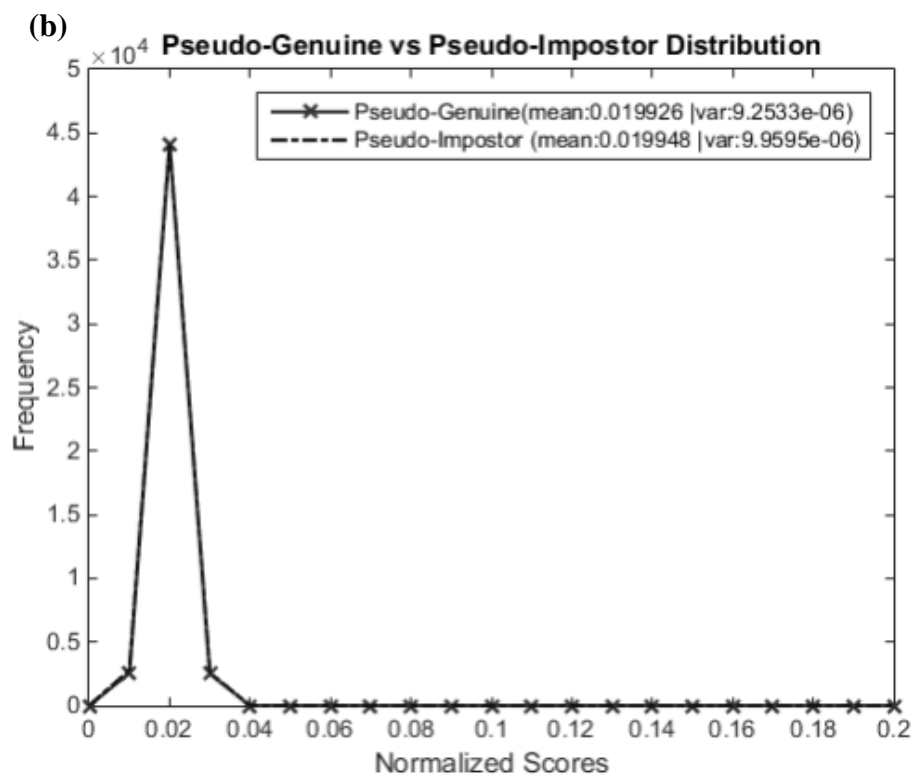
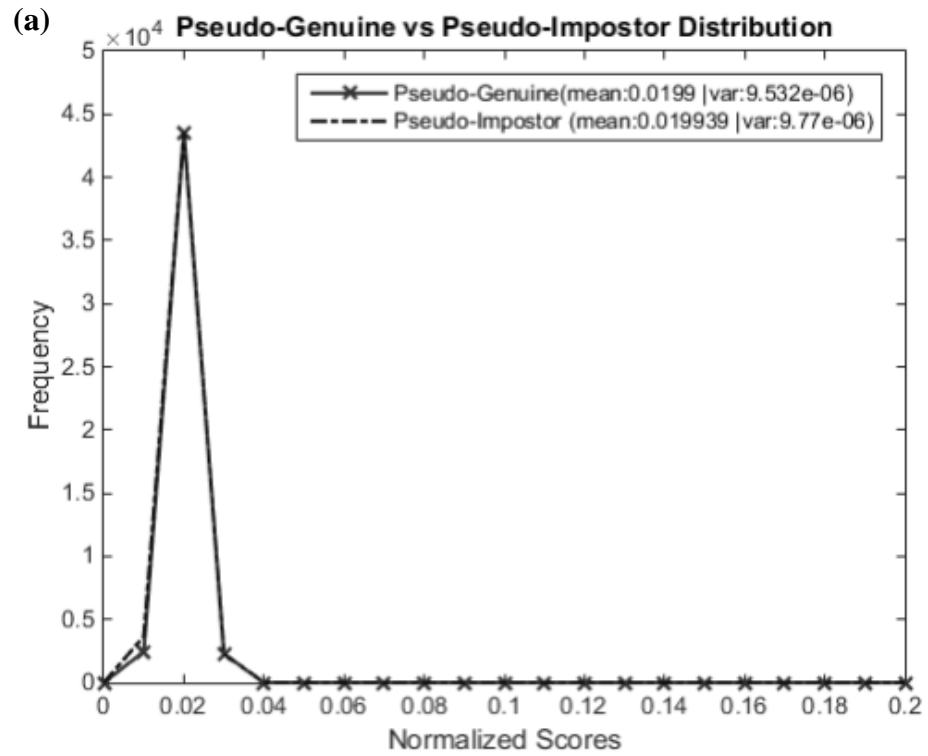
Figure 4.5: The genuine, impostor and pseudo-impostor distribution for database combination: (a)  $V$  and  $F1$ , (b)  $V$  and  $F2$  and (c)  $V$  and  $F3$

From Figure 4.5, it can be seen that the pseudo-impostor distribution resembles the impostor distribution. Hence, this vindicates that the newly generated hashed codes are distinctive to each other although they are generated from the same fusion matrix. Thus, the revocability criterion has been satisfied as the newly generated hashed codes are uncorrelated to the old hashed code.

#### 4.4.2 Unlinkability Analysis

To evaluate the unlinkability property of the proposed scheme, pseudo-genuine score is introduced. Pseudo-genuine scores are computed by matching the hashed code generated from different fusion matrices from the same user using different permutation seeds. Similar to genuine matching, there are a total of 1000 pseudo-genuine matches. The distribution of the pseudo-genuine

distribution against pseudo-impostor distribution for each dataset combination is plotted as shown in Figure 4.6.



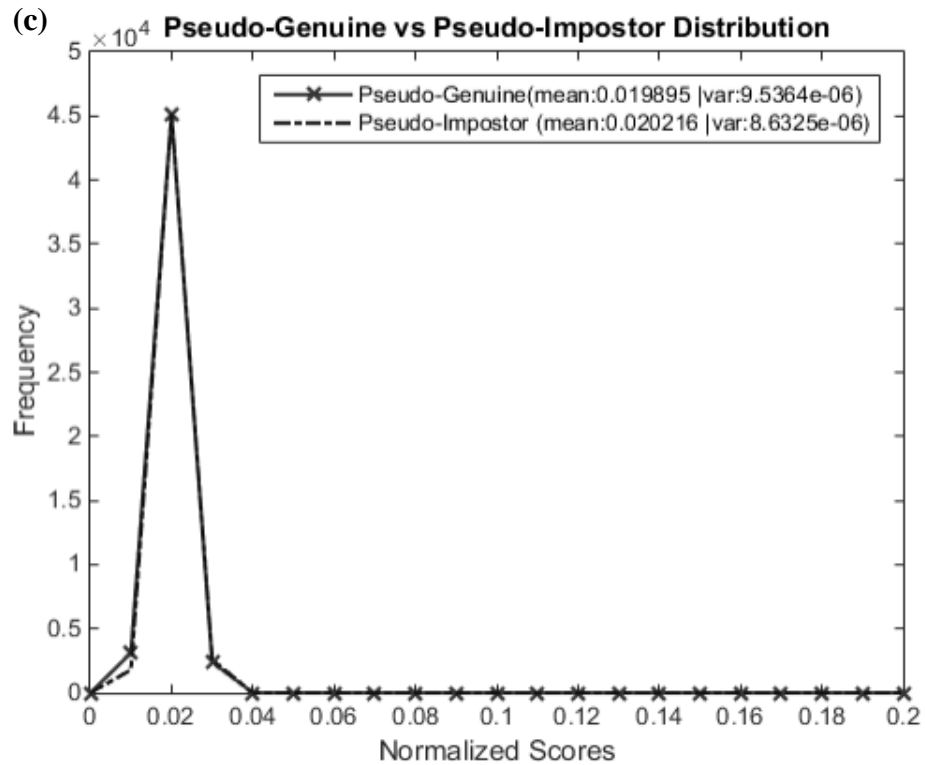


Figure 4.6: The pseudo-genuine and pseudo-impostor distribution for database combination: (a)  $V$  and  $F1$ , (b)  $V$  and  $F2$  and (c)  $V$  and  $F3$

The overlapping of pseudo-genuine distribution and pseudo-impostor distribution vindicates that the hashed codes generated are indistinguishable from each another. Therefore, this suggests that the generated hashed codes are unlinkable as it is difficult to differentiate them. Hence, the unlinkability property of the proposed scheme is justified.

#### 4.4.3 Brute Force Attack

The projection of the fused template to ordinal space imposes strong non-invertible property to the recognition system as it is computationally difficult for the adversary to derive the original feature value in linear space. Under the worst case scenario where the adversary is able to obtain the permutation seeds

as well as the hashed code, the adversary will still not be able to recover the feature value from index value to continuous value.

From the distribution of the feature value of the fusion matrix,  $M$ , ranging from -0.4000 to +0.4000, if the adversary conducts brute force attack to recover the original feature value, approximately  $8001^{500 \times 299} \approx 2^{1938411.79}$  attempts are required (notice that  $M$  has a dimension of  $500 \times 299$ ). As 2DWTA also focuses on the ordering instead of the feature value itself, hence if the adversary wants to guess the ranking of the feature value,  $299! \approx 2^{2033}$  attempts and  $500! \approx 2^{3768}$  attempts are still needed for fingerprint and voice respectively.

The similarity score is computed based on the number of matches between the query hashed code and the enrolled hashed code. If the similarity score exceeds the predefined threshold, the user will be deemed as legitimate user. Hence, the adversary might exploit the threshold to obtain the sufficient number of matches in order to access the system. In this experiment, the threshold is approximately 0.15. Setting  $h = 8000$  and each round will produce two indices (i.e., row index and column index), the total correct matches required to access the system will be  $8000 \times 0.15 \times 2 = 2400$  matches. Thus, a time complexity of  $50^{2400} \approx 2^{13545.25}$  attempts is needed to gain access to the system.

Consider the scenario that the adversary does not compromise any templates, as the setting of  $k = 50$ , meaning that there would be a maximum of 50 possible outcomes for a particular index, the probability that the adversary

obtain the correct index will be  $1/50 = 0.02$ . Understanding the fact that there will only be two possible outcomes, either the adversary obtains the correct index or incorrect index for each round regardless of row index or column index are independent and uniformly distributed, one can assume that the hashed code,  $S$ , follows binomial distribution with the probability,  $p = 0.02$ , and the total number of indices,  $n = 8000 \times 2 = 16000$ . The mean and the variance of the binomial distribution will be  $n \times p = 16000 \times 0.02 = 320$  and  $n \times p \times (1-p) = 16000 \times 0.02 \times (1-0.02) = 313.6$  respectively. Let  $X$  defines the number of correct guesses of indices, the probability of obtaining at least 2400 correct guesses can be computed as follows:

$$\begin{aligned}
P(X \geq 2400) &= P\left(z \geq \frac{2400 - 320}{\sqrt{313.6}}\right) \\
&= P(z \geq 117.45) \\
&= 2^{-9959.85}
\end{aligned} \tag{4.1}$$

From Equation 4.1, the average time complexity required or the number of guesses needed to guess the order of the elements in the feature matrix is  $2^{9959.85}$ .

#### 4.4.4 Attack-via-Record Multiplicity (ARM)

As the computation of the similarity score is dependent on the hashed code, which is derived from the highest indices of the  $k$ -dimension square window matrix, the main concern is whether the adversary can reconstruct the order or the ranking of the elements in the fusion matrix. Here, we focus on the ability



of the adversary to recover the order of the elements instead of the value of the elements as it is relatively easier for the adversary to guess the order since it has a lesser possibilities as compared to the feature value which is continuous. Figure 4.7 depicts the information obtained by the adversary under ARM scenario.

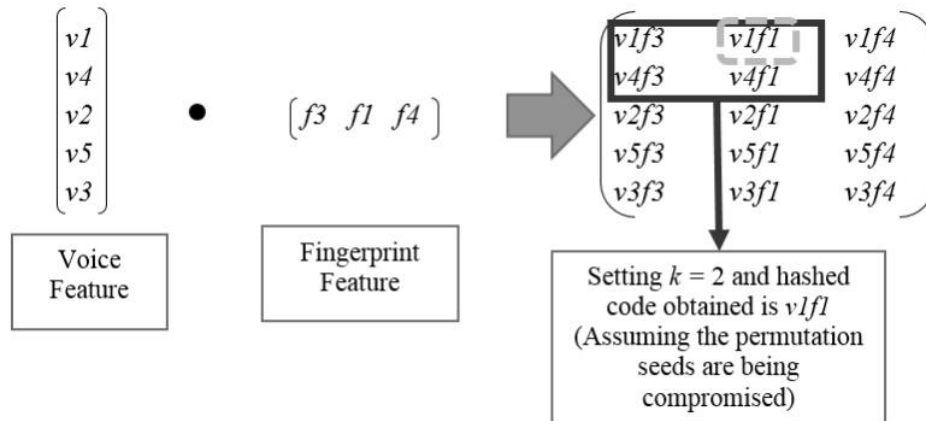


Figure 4.7: Example of hashed code,  $k$  and permutation seeds being compromised by the adversary

From Figure 4.7, given that the output from 2DWTA is  $v1f1$ , using inequality approach, it can be reasoned that  $v1 > v4$  and  $f1 > f3$ . By repeating inequality approach on the remaining elements, one can deduce the order of the biometric features. However, this assumption is only valid if and only if the values of the biometric features are of the same sign (i.e., the values of the biometric features are all positive or negative). For instance, if  $v1$  and  $f1$  are both negative and  $v4$  and  $f3$  are both positive, it can be deduced that  $v1f1 > v1f3$  and  $v1f1 > v4f1$ . Under the scenario where  $|v1f1| > |v4f3|$ ,  $v1f1$  will still have the highest indices and thus this will violate the earlier assumption where in this scenario,  $v1 < v4$  and  $f1 < f3$ . Therefore, this violation implies that it is insufficient for the

adversary to derive the entire order of the biometric feature using inequality approach.

#### **4.5 Summary**

Similar to RBOMP hashing, the proposed 2DWTA is able to preserve the performance of the recognition system. In the aspect of security, 2DWTA is more resistant towards major security attacks (i.e., ARM and stolen token attacks) as compared to RBOMP hashing with a complexity of approximately 10000 bits. It has also been shown that using inequality approach is insufficient for the adversary to reconstruct the order of the feature matrix. Furthermore, it is also justified that the 2DWTA is able to satisfy the unlinkability and revocability properties of biometric template protection requirement.

## CHAPTER 5

### CONCLUSION

#### 5.1 Conclusion

We have proposed two voice template protection schemes based on Winner-Takes-All hashing (WTA), namely RBOMP hashing and 2DWTA hashing, where the former is intended for one-dimensional input (i.e., vector) and the latter is designed for two-dimensional input (i.e., matrix). Both schemes project the biometric feature to ordinal space that is resilient to insignificant intra-class variation while inducing strong non-invertible property. With the addition of factor of authentication, the security of the proposed schemes have been enhanced as the use these additional factors will further increase the security complexity of the system. Furthermore, the use of two or more factors of authentication are able to reduce the risk of identity thefts as the adversary will need to compromise both the biometric data as well as the additional token or passcode. Experimental analysis have shown that both of the schemes are able to preserve the performance of the recognition system while offer certain degree of security against major attacks, such as stolen token attack and ARM. Additional analyses have been carried out for both schemes. Analyses have justified that both schemes are able to satisfy all the evaluation criteria of the

biometric protection requirements (i.e., revocability, unlinkability, non-invertibility and performance).

## 5.2 Future Work

There are many interesting directions that have been left for the future due to the time constrain. As this work mainly focused on voice template protection, it would be great if the following ideas could be tested:

- It would be interesting to apply the biometric protection schemes not only on voice feature but also other biometric modalities with real-value representation. As the proposed biometric protection schemes are able to protect the original biometric features by indexing the value, the similar approach may be applicable to other biometric modality such as fingerprint or face modalities which are also in real-value domain.
- It would be interesting to improve the performance of the proposed schemes by exploring the possibility of employing deep learning in the template protection. Due to the emergence of the internet of things, deep learning has been at the centre of the focus among the researchers and we also like to further investigate the possibility of employing the methodology or techniques of deep learning in our schemes. As deep learning has shown significant improvement in performance over identification and classification problem, hence it would be feasible for us to believe that such approach will help us in enhancing the performance of our proposed schemes.

## REFERENCE

Badiul, A., Zhe, J., Yap, W.S. and Goi, B.M., 2018. An alignment-free cancelable fingerprint template for bio-cryptosystems. *Journal of Network and Computer Applications*, 115, pp. 20-32.

Billeb, S., Busch, C., Reininger, H., Kasper, K. and Rathgeb, C., 2015. Biometric template protection for speaker recognition based on universal background models. *IET Biometrics*, 4 (2), pp. 116-126.

Chandra, E. and Kanagalakshmi, K., 2011. Cancelable biometric template generation and protection schemes: A review. *International Conference on Electronics Computer Technology (ICECT)*, pp. 15-20.

Chang, E., Shen, R. and Teo, F.W., 2006. Finding the original point set hidden among chaff. *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security (ASIACCS), 2006*, pp. 182-188.

Chong, L.Y. and Teoh, A.J., 2007. Probabilistic random projections and speaker verification. *International Conference on Biometrics (ICB), 2007*, pp. 445-454.

Daum, F. and Huang, J., 2003. Curse of dimensionality and particle filters. *2003 IEEE Aerospace Conference Proceedings (Cat. No03TH8652)*, 4, pp. 1979-1993.

Dean, T. et al., 2013. Fast, accurate detection of 100,000 object classes on a single machine. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1814-1821.

Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P. and Ouellet, P., 2011. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19 (4), pp. 788-798.

Dharavath, K., Talukdar, F.A. and Laskar, R.H., 2013. Study on biometric authentication systems, challenges and future trends: A review. *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1-7.

Ferrante, S. and Saltalamacchia, M., 2014. The coupon collector's problem. *Mater. Math.*, 2014, (2), pp. 1-35.

Fontaine, C. and Galand, F., 2007. A survey of homomorphic encryption for nonspecialists. *EURASIP Journal on Information Security*, 2007 (1), pp. 1–15.

García-Romero, D. and Espy-Wilson, C.Y., 2011. Analysis of i-vector length normalization in speaker recognition systems. *Interspeech 2011*, pp. 249-252.

Gomez-Barrero, M., Maiorana, E., Galbally, J., Campisi, P. and Fierrez, J., 2017. Multi-biometric template protection based on Homomorphic Encryption. *Pattern Recognition*, 67, pp. 149-163.

Inthavisas, K. and Lopresti, D.P., 2011. Speech cryptographic key regeneration based on password. *2011 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1-7.

Jain, A.K., Nandakumar, K. and Nagar, A., 2008. Biometric template security. *Journal on Advances in Signal Processing*, pp. 1-17.

Jain, A.K. and Nandakumar, K., 2012. Biometric Authentication: System Security and User Privacy. *IEEE Computer Society*, 45 (11), pp. 87-92.

Jain, A.K., Nandakumar, K. and Ross, A., 2016. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79, pp. 80-105.

Johnson, R.C., Scheirer W.J. and Boulton, T.E., 2013. Secure voice-based authentication for mobile devices: Vaulted voice verification. *Proc. SPIE*,



*Biometric and Surveillance Technology for Human and Activity Identification X*; doi: 10.1117/12.2015649.

Juels, A. and Sudan M., 2006. A fuzzy vault scheme. *Design, Codes and Cryptography*, 38 (2), pp. 237-257.

Juels, A. and Wattenberg, M., 1999. A fuzzy commitment scheme. *Proceedings of the 6th ACM Conference on Computer and Communications Security (CCS)*, pp. 28-36.

Kenny, P., 2010. Bayesian Speaker Verification with Heavy-Tailed Priors. *Odyssey 2010*: 14.

Labati, R.D., Piuri, V. and Scotti, F., 2012. Biometric privacy protection: Guidelines and technologies. *E-Business and Telecommunications Communications in Computer and Information Science*, pp. 3-19.

Lei, Y., Scheffer, N., Ferrer, L. and McLaren, M., 2014. A novel scheme for speaker recognition using a phonetically-aware deep neural network. *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference*, pp. 1695-1699.

Li, M., Liu, L., Cai, W. and Liu, W., 2016. Generalized i-vector representation with phonetic tokenizations and tandem features for both text

independent and text dependent speaker verification. *Journal of Signal Processing Systems*, 82 (2), pp. 207-215.

Matějka, P. et al., 2011. Full-covariance UBM and heavy-tailed PLDA in i-vector speaker verification. *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference*, pp. 4828-4831.

Matejka, P. et al., 2014. Neural network bottleneck features for language identification. *Proc. IEEE Odyssey*, pp. 299-304.

Matsumoto, M. and Nishimura, T., 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random generator. *ACM Transactions on Modeling and Computer Simulation*, 8 (1), pp. 3-30

Morgan, S.C., 2016. Cyber crime costs projected to reach \$2 Trillion by 2019. Retrieved July 24, 2016, from <http://www.forbes.com/sites/stevemorgan/2016/01/17/cyber-crime-costs-projected-to-reach-2-trillion-by-2019/>

Morgan, S.C., 2016. The cybersecurity market report q2 2016. Retrieved July 24, 2016, from <http://cybersecurityventures.com/cybersecurity-market-report/>

Nandakumar, K. and Jain, A. K., 2015. Biometric template protection: Bridging the performance gap between theory and practice. *IEEE Signal Processing Magazine*, 32 (5), pp. 88-100.

Paillier, P., 1999. Public-key cryptosystems based on composite residuosity classes. *Proc. EUROCRYPT, 1999*, pp. 223–238

Pandey, R., Zhou, Y., Kota, B.U. and Govindaraju, V., 2016. Deep secure encoding for face template protection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016*, pp. 77-83.

Paulini, M. et al., 2016. Multi-bit allocation: Preparing voice biometrics for template protection. *The Speaker and Language Recognition Workshop (Odyssey)*, pp. 291-296.

Ponemon Institute LLC, 2015. 2015 Cost of cyber crime study: Global. North Traverse City, Michigan: Ponemon Institute LLC. Available at: [http://www.cnmeonline.com/myresources/hpe/docs/HPE\\_SIEM\\_Analyst\\_Report\\_-\\_2015\\_Cost\\_of\\_Cyber\\_Crime\\_Study\\_-\\_Global.pdf](http://www.cnmeonline.com/myresources/hpe/docs/HPE_SIEM_Analyst_Report_-_2015_Cost_of_Cyber_Crime_Study_-_Global.pdf) (Accessed: 20 April 2016).

Raju, S.V., Vidyasree, P. and Madhavi, G., 2014. Enhancing security of stored biometric template in cloud computing using fec. *International*

*Journal of Advanced Computational Engineering and Networking*, 2 (2), pp. 35-39.

Ratha, N.K., Connell, J.H. and Bolle, R.M., 2001. An Analysis of Minutiae Matching Strength. *Proceedings of Third International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp. 223–228.

Ratha, N.K., Cikkerur, S., Connell, J.H. and Bolle, R.M., 2007. Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 (4), pp. 561-572.

Scheirer, W.J. and Boulton, T.E., 2007. Cracking fuzzy vaults and biometric encryption. *Proceedings of the Biometrics Symposium*, pp. 1-6.

Snyder, D., Garcia-Romero, D. and Povey, D., 2015. Time delay deep neural network-based universal background models for speaker recognition. *Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop*, pp. 92-97.

Teoh, A.B.J., Yip, W.K. and Lee, S.Y., 2008. Cancellable biometrics and annotations on biohash. *Elsevier - Pattern Recognition*, 41 (6), pp. 2034-2044.

Wang, S., Yang, W. and Hu, J., 2017. Design of alignment-free cancelable fingerprint templates with zoned minutia pairs. *Pattern Recognition*, 66, pp. 295-301.

Wang, Y. and Plataniotis, K.N., 2010. An analysis of random projection for changeable and privacy-preserving biometric verification. *IEEE Trans. Systems, Man, and Cybernetics, Part B*, 40 (5), pp. 1280-1293.

Xu, W. and Cheng, M., 2008. Cancelable voiceprint template based on chaff-points-mixture method. *2008 International Conference on Computational Intelligence and Security (CIS)*, 2008, pp. 263-266.

Yagnik, J., Strelow, D., Ross, D.A. and Lin, R., 2011. The power of comparative reasoning. *2011 International Conference on Computer Vision*, pp. 2431-2438.

Yang, W., Wang, S., Hu, J., Zheng, G. and Valli, C., 2018. A fingerprint and finger-vein based cancelable multi-biometric system. *Pattern Recognition*, 78, pp. 242-251.

Zhu, H., He, Q. and Li, Y., 2012. A two -step hybrid approach for voiceprint-biometric template protection. *2012 International Conference on Machine Learning and Cybernetics*.

## LIST OF PUBLICATION

Chee, K.Y. et al., 2018. Cancellable speech template via random binary orthogonal matrices projection hashing. *Pattern Recognition*, 76, pp. 273-287.

Chee, K.Y., Jin, Z., Yap, W.S., Goi, B.M., 2017. Two-dimensional winner-takes-all hashing in template protection based on fingerprint and voice feature level fusion. *APSIPA ASC 2017*.