MITIGATING MASTERPRINT VULNERABILITY

A SECURITY THREAT TO FINGERPRINT BIOMETRIC SYSTEMS

Ph.D. Thesis

By

MAHESH JOSHI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JULY 2022

MITIGATING MASTERPRINT VULNERABILITY

A SECURITY THREAT TO FINGERPRINT BIOMETRIC SYSTEMS

A THESIS

submitted to the

INDIAN INSTITUTE OF TECHNOLOGY INDORE

in partial fulfillment of the requirements for
the award of the degree of

DOCTOR OF PHILOSOPHY

by

MAHESH JOSHI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JULY 2022

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Mitigating Master-Print Vulnerability A Security Threat To Fingerprint Biometric Systems** in the partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy** and submitted in the **Department of Computer Science and Engineering, Indian Institute of Technology Indore,** is an authentic record of my own work carried out during the time period from July 2017 to July 2022 under the supervision of Dr Bodhisatwa Mazumdar, Associate Professor, and Dr Somnath Dey, Associate Professor, Indian Institute of Technology Indore, Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

mahesh
15 July 2022
Signature of the Student with Date

(Mahesh Joshi)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Bodhisatwa Mazumdar
15-07-2022

Somnath Dey
15-07-2022

Signature of Thesis Supervisors with Date

(Dr Bodhisatwa Mazumdar)

(Dr Somnath Dey (Officiating))

Mahesh Joshi has successfully given his Ph.D. Oral Examination held on 08-02-2023

Bodhisatwa Mazumdar
Signature of Thesis Supervisor #1

Date: 10-02-2023

Somnath Dey
Signature of Thesis Supervisor #2

Date: 10-02-2023

ACKNOWLEDGEMENTS

My journey from a small town to a research scholar at the prestigious Indian Institute of Technology (IIT) Indore was fascinating. I met several people whose motivation and belief in my capabilities inspired me to achieve more than I could have imagined. So, I wish to take this opportunity to thank each one of them as I complete my doctoral research work and submit the thesis.

I feel privileged to have **Dr. Bodhisatwa Mazumdar** and **Dr. Somnath Dey** as my supervisors. Dr. Bodhisatwa's attention to detail, the quest for excellence, and love for perfection has inspired me to give my best time and again. I am indebted to him for providing me with the research problem even when he has rarely worked in biometric security. His consistent and unending directions steered me on the right path and transformed me into a researcher who can work independently. The expertise and experience Dr. Somnath has in biometric system security have propelled my work resulting in quality research publications. His minute observations, accurate suggestions, and continuous support during the PhD gave me hope in tough times. I want to thank him for encouraging my research and allowing me to grow as a researcher. The confidence my supervisors have shown in me has fostered my motivation. They helped in defining my research and presentation skills, and I am deeply obliged to make the PhD experience memorable. I am grateful for their heartiest support in providing the necessary resources during the pandemic. I could not have imagined having better supervisors and mentors for my PhD study.

My profound thanks to **Prof. Suhas Joshi**, Director, IIT Indore, for providing a competitive research environment in the institute. Besides, I extend my gratitude to my PG Student's Progress Committee (PSPC) members, **Dr. Surya Prakash** and **Prof. Prabhat Kumar Upadhyay**, for their keen observation, valuable comments, insightful suggestions, encouragement, and validation of the research work. I also convey sincere thanks to **Dr. Neminath Hubbali**, DPGC Convener, Department of Computer Science and Engineering and **Prof. Abhishek Srivasatva** for a prolific conversation and motivation at every interaction. I also thank all the faculty members from the Department of Computer Science and Engineering for their helpful advice and constant encouragement. I am also grateful to the department technical staff members for their unfailing support and assistance. Special gratitude goes to the Indian Institute of Technology Indore for providing me with an opportunity to pursue a PhD under a state-of-the-art research environment and the Ministry of Education (MoE), formerly the Ministry of Human Resource Development (MHRD), Government of India, for granting financial aid to carry out this doctoral study successfully.

Friends played a significant role in bringing out the best in me and becoming a support system in taking brave decisions irrespective of the outcome. I wish to thank my fellow lab members **Dr. Rajat Saxena, Dr. Rudresh Dwivedi, Dr. Ram Prakash Sharma, Anuj Rai, Aparna Santra Biswas,**

and Priyanka Joshi. A short technical discussion with them would solve weeks worth of confusion and bring more clarity to work. The best part of this memorable journey had been the company of cheerful doctoral fellows. **Dr. Nikhil Tripathi, Dr. Mayank Swarnkar, Dr. Navneet Pratap Singh, Dr. Rajendra Choudhary, Dr. Aditya Shastri, Dr. Rohit Agrawal, Dr. Chandan Gautam, Dr. Piyush Joshi, Dr. Syed Sadaf Ali, Dr. Akhilesh Mohan Srivastava, Arun Kumar, Dr. Vikas Chauhan, and Ankit Jain** have always extended their feedback, encouragement, cooperation, and of course, friendship. Also, special thanks to **Hitendra Singh**, internship student, for cooperating in implementing an algorithm.

Teachers deserve the maximum credit for what I have achieved. My school teachers **Late Shri. Rade sir, Shri. Kala sir, Shri. Bhatkar sir** at Jawahar Navodaya Vidyalaya Washim (JNVW), constantly showered their immense love, care, and support while spending the unforgettable hostel days during my childhood. All the teachers believed in me during my intermediate schooling days and offered every possible help to excel in the board exam, including coaching at their homes. During my bachelor's course, **Prof. S V Dhopte**, my thesis advisor, often enquired about any problems in my studies and repeatedly tried his level best to convince me that I could be one of the best performers in the batch. It was during my master's degree that changed my perception of myself. **Prof. S R Sathe**, my master's thesis guide, undoubtedly believed in my potential and offered the latest infrastructure to carry out research. His friendly nature and humble gesture relieved the stress and gave me confidence in tough times. I bow down and feel blessed and grateful to the almighty for keeping me in the shadow of such beautiful souls.

Friends played a significant role in bringing out the best in me and becoming a support system in taking fearless decisions irrespective of the outcome. I sincerely thank all my friends for listening and tolerating me being a chatterbox. To start with JNVW days, I am grateful to my friends from the first batch for accompanying me for so many years. I especially thank **Dr. Srikant Wankhede** for reassuring me that research would be the best carrier choice. My bachelor's degree would not be joyful without sharing a room, lunchbox, bike, and canteen bills with **Amol, Nitin, Sidhu, Satish, Vijay, Dhiraj, Atul, Pankaj, Milind, and Sariputra**. A special thanks go to **Santosh Jadhav, Santo,** for being more than just a friend and cheering in every meeting. I want to thank **Mujeeb Rehaman, Bhavesh Ghode, Gajanan Bejgamwar, Satish More, Dr. Amit Khaparde, Sunil Jatti, Anurag Bangad, Sandip Nawale, Dr. Adepu Sridhar, and Dr. D V N Sivakumar** for spending a good time together.

Nothing would have been possible without the moral support of **my parents**, who have been the pillars of strength in all my endeavours. I am always deeply indebted to them for everything I have achieved. I wholeheartedly thank my wife, **Priyanka**, for her continuous support during the roller

coaster ride throughout the PhD course, her encouragement, understanding and sacrifice, without which it would have been impracticable for me to finish this work. I regret often being rude to my son, **Ishaan**, due to deadlines and work pressure. But I am grateful to him for providing me with the required space to carry out my incomplete tasks at home during the pandemic. I also thank **my in-laws** for looking after my son and taking care of our everyday needs when I needed them the most. My **maternal uncle** was always continuous support since childhood. At times he was more worried about my PhD completion than me and regularly enquired about the same. My **siblings and cousins** have been a stress buster at times. I feel blessed to have a loving, caring, and supportive family. I thus wish to thank every family member from this platform.

Mahesh Joshi
Indian Institute of Technology Indore
July 2022

Dedicated to My Parents & Teachers

ABSTRACT

Biometric-based user identification and authentication systems have gained popularity over the traditional password-based schemes wherein the user must remember the secret information. Moreover, fingerprint biometric systems have received wide acceptance in government offices, academic institutions, and industrial applications due to their affordable cost, user convenience, and high accuracy. However, these systems become vulnerable to various threats due to the diverse components involved in the system. The work in this thesis is an attempt to explore the security aspect of fingerprint biometric systems. Initially, a sixteen attack point-based threat model for a match-in-database fingerprint biometric system is proposed and compared with four existing models. The model depicted all probable threats to the system and its description provided the countermeasures for each attack scenario. The model is further employed to categorise the attacks into eight classes based on the vulnerable components. The main contribution of the dissertation focuses on the *MasterPrint* vulnerability. A MasterPrint is a partial fingerprint identifying at least 4% distinct subjects from the enrolled user's database. In simple words, a MasterPrint illustrates a non-unique user identification by the system. As latent fingerprints collected from crime spots are usually partial, we experimented on a latent fingerprint dataset to investigate the possibility of *Latent MasterPrint*. The results revealed that Latent MasterPrint do exist.

As the MasterPrint vulnerability has proved to be a significant threat to the fingerprint biometric system, our main contribution in this thesis aims to mitigate the vulnerability through novel minutiae-based feature extraction approaches. We considered MasterPrint vulnerability an identification problem and experimented with the proposed methods using open-set and closed-set identification scenarios. The work in this thesis presented two schemes and compared the results with six existing methods on five standard fingerprint datasets. The idea to disallow tolerance during feature comparison delivered satisfactory results for both approaches. The first method addressed the vulnerability by creating an eight-axes coordinate system around a reference minutia and extracting local features over binarized and thinned partial fingerprint images. Here, each minutia is represented as a seventeen-element integer-valued feature vector. A similarity score computation metric suitable for strict feature vector comparison is also introduced. The results showed about 92% accuracy while generating only 2.25% MasterPrints. In the second approach, three minutiae-based geometric constructs are formed to create a six-element integer-valued feature vector.

The method resulted in an identification accuracy of around 97% while generating merely 0.01% MasterPrints. Further, the second approach is investigated using 16 diverse combinations of thresholding and thinning approaches to study the impact of preprocessing techniques on its identification rate and percentage of MasterPrint generated using five benchmark fingerprint datasets. The observations proved that the thresholding and thinning methods substantially influence the accuracy and other crucial parameters. It also confirmed that every thresholding and thinning method from the literature may not be suitable for high-security applications, such as anonymous user identification using partial fingerprint.

PUBLICATIONS FROM THE THESIS

The following publications have evolved from this doctoral dissertation (as of February 2023):

In Refereed Journals

Published / Accepted

1. **Mahesh Joshi** and Bodhisatwa Mazumdar and Somnath Dey, "A comprehensive security analysis of match-in-database fingerprint biometric system". *Pattern Recognit. Lett.*, vol. 138, pp. $247-266$, 2020, https://doi.org/10.1016/j.patrec.2020.07.024, (SCI, IF: 4.757)

2. **Mahesh Joshi** and Bodhisatwa Mazumdar and Somnath Dey, "Mitigating MasterPrint vulnerability by employing minutiae geometry", *Journal of Electronic Imaging* 31(1), 013026(2022), https://doi.org/10.1117/1.JEI.31.1.013026, (SCIE, IF- 0.945)

3. **Mahesh Joshi** and Bodhisatwa Mazumdar and Somnath Dey, "Investigating the impact of thresholding and thinning methods on the performance of partial fingerprint identification systems: a review", *Journal of Electronic Imaging* 32(1), 010901 (2023), https://doi.org/10.1117/1.JEI.32.1.010901, (SCIE, IF- 0.945)

Under review

1. **Mahesh Joshi** and Bodhisatwa Mazumdar and Somnath Dey, "A Novel Minutiae-based Approach towards Mitigating MasterPrint Vulnerability for Partial Fingerprints", *Pattern Recognition* (SCI, IF- 8.518)
   (First submitted: November 2020**, Revised version submitted: January** 2022**)**

In Refereed Conference

Published

1. **Mahesh Joshi** and Bodhisatwa Mazumdar and Somnath Dey, "On the Prospects of Latent MasterPrints", *in the Proceedings of the* 6^{th} *IAPR International Conference on Computer Vision & Image Processing (CVIP)* $3-5^{th}$ *Dec, 2021 IIT Ropar, Punjab-140001, INDIA*, https://doi.org/10.1007/978-3-031-11346-8_49

Book chapter

1. **Mahesh Joshi**, Bodhisatwa Mazumdar, and Somnath Dey. 2021. *Biometric-based Secure Authentication for IoT-enabled Devices and Applications* . In (First) Security of Internet of Things Nodes Challenges, Attacks, and Countermeasures. Chapman, Boca Raton, Florida, 81–106. https://doi.org/10.1201/9781003127598

Poster presentation

1. **Mahesh Joshi** and Bodhisatwa Mazumdar and Somnath Dey, "Biometric Security and Authentication: Threats, Countermeasures and Template Protection Schemes", *Poster Session of* 10^{th} *International Conference on Security, Privacy and Applied Cryptographic Engineering (SPACE'*20*)* ($17^{th} - 20^{th}$ December 2020), IIT Kharagpur.

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations & Acronyms

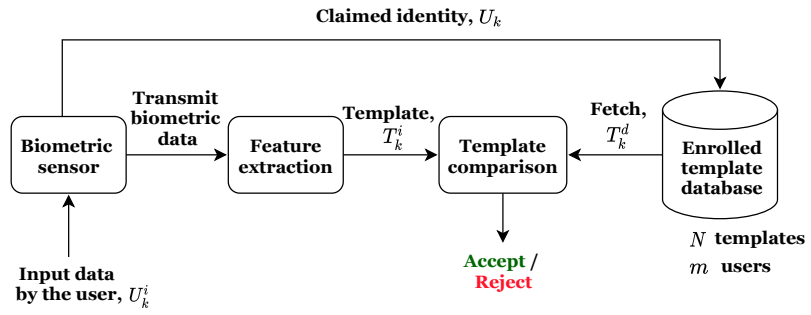
AFIS	Automatic Fingerprint Identification System
AP	Attack Points
ATM	Automated Teller Machine
BMM	Baseline Minutiae Matching algorithm
CDI	Correct Detect and Identify
CMC	Cumulative Matching Characteristic curve
CN	Crossing Number
CNN	Convolutional Neural Network
CNNAI	Combination of Nearest Neighbor Arrangement Indexing
DIR	Detect and Identification Rate
DDoS	Distributed Denial-of-Service attack
DoS	Denial-of-Service attack
EER	Equal Error Rate
FAR	False Alarm Rate
FBS	Fingerprint Biometric System
FVC	Fingerprint Verification Competition
GAN	Generative Adversarial Network
ICs	Integrated Circuits
IF	Intensity Factor
IS	Intensity Shift
IoT	Internet-of-Things
LFIQ	Latent Fingerprint Image Quality
LSA	Line Scan Algorithm
MCC	Minutia Cylinder-Code

MiD	Match-in-Database biometric systems
MoC	Match-on-Card biometric systems
MoD	Match-on-Device biometric systems
MOLF	Multi-sensor Optical, and Latent Fingerprint dataset
MoS	Match-on-Server biometric systems
MPC	MasterPrint count
MR	MasterPrint record
NBIS	NIST Biometric Image Software
NFIQ	NIST Fingerprint Image Quality
NIST	National Institute of Standard and Technology
ROC	Receiver Operating Characteristic curve
RR	Rejection rate
RRP	Representative Ridge Point
RSF	Ridge Shape Features
SFTA	Spaced Frequency Transformation Algorithm
SDK	Software Development Kit
SHA	Similarity Histogram Approach
SIFT	Scale-Invariant Feature Transform algorithm
SoC	System-on-Card biometric systems
SURF	Speeded-Up Robust Features
SVM	Support Vector Machine
TBS	Trusted Biometric System
ToC	Template-on-Card biometric systems
TTL	Time-to-live

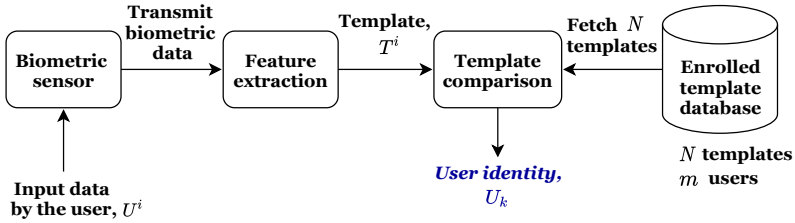
Chapter 1

Introduction

The term "biometrics" is derived from two Greek words *bios* meaning "life" and *metron* meaning "measurement" [5]. Thus, we can deduce that a biometric system collects and analyses biological data employing science and technology. The personal traits are challenging to forge, misplace, and share. Hence, the biometric systems are preferred over token-based security schemes, e.g., ID cards, keys, etc., and knowledge-based authentication methods, e.g., PIN, password, etc. [6]. These systems are widely utilised for individual recognition. The terms, *recognition and authentication*, are used in generic sense for *verification* and *identification* while referring to a biometric system. A typical scenario of verification and identification process is shown in Figure 1.1. The system administrator enrols at least three samples of the same finger of a user to improve the system accuracy. In this scenario, we assume that the template database contains N templates corresponding to m users, $m < N$. In the case of verification, the user, U_k, submits his biometric data, U_k^i, where $k \in \{1, 2, \ldots, m\}$ and i denotes the *input* fingerprint, to the sensing device and reveals his identity to the system. The system extracts features from the data to create an encrypted template, T_k^i. The template comparison module fetches the enrolled templates corresponding to the claimed user, T_k^d, where d denotes the *database* template, and compares it with T_k^i to decide the acceptance or rejection of the claim. However, the identification system accepts the input biometric data from an anonymous user and creates a template, T^i. The comparison module fetches every enrolled template from the database and compares it with T^i. As a result, the identity of a user, U_k, corresponding to the enrolled template possessing maximum similar features with T^i, becomes the unknown user's identity.



(a) Verification scenario



(b) Identification scenario

Figure 1.1: Biometric system application for (a) user verification, and (b) user identification.

Biometric systems accept various physiological traits, such as palmprint, fingerprint, face, iris, ear, etc., and behavioural characteristics, such as signature, gait, voice, stroke, etc., for individual recognition. However, fingerprint biometric systems (FBS) gained wide acceptance for commercial applications due to user convenience, high accuracy, and affordable cost. The fingerprint biometric systems are extensively used for access control, user authentication and authorisation in military, government, industry, and academic institutions. Hence, it becomes imperative to analyse these systems from security perspective. Therefore, the work in this thesis revolves around identifying security threats and vulnerabilities to FBS and uncovering possible ways to mitigate them. In the initial work from this thesis, a comprehensive threat model comprising sixteen vulnerable attack points to a match-in-database fingerprint biometric system is proposed. In September 2017, the investigations on the partial fingerprint identification system exposed the MasterPrint vulnerability [7]. A partial fingerprint fortuitously identifying at least 4% unique subjects enrolled with the system was termed a *MasterPrint*. Latent fingerprints are lifted from surfaces of objects that are inadvertently touched or handled by a person [8]. A latent MasterPrint is referred to a latent fingerprint that identifies at least 4% distinct subjects from a database storing fingerprints of past offenders.

Further work in this thesis is focused on mitigating the recently reported MasterPrint vulnerability, investigating the chances of generating Latent MasterPrint, and studying the impact of fingerprint preprocessing approaches towards addressing the MasterPrint vulnerability.

The organisation of this chapter is as follows. The necessity for investigating and addressing security concerns of FBS are discussed in Section 1.1. The motivation of the work in this thesis follows in Section 1.2. Subsequently, Section 1.3 states the objectives of the research carried out in this thesis. Section 1.4 highlights the major contributions made in this thesis. Finally, organisation of the thesis is given in Section 1.5.

1.1 Necessity of investigating FBS security

Fingerprint biometric systems are the most successful commercial products employed by various industries, universities, government, and military organisations worldwide. The other modalities, such as the face, iris, voice, palmprint, etc., have several drawbacks. The disadvantages of these systems include higher implementation costs, comparatively lower accuracy, and more user assistance. Moreover, the judicial systems consider latent fingerprints as supportive evidence. Therefore, academic researchers and forensic investigators rigorously study automatic fingerprint identification systems (AFIS).

Furthermore, other biometric modalities contain large-sized templates, so their response time is usually higher than the fingerprint biometric systems. Hence, other traits are less preferred for authentication involving large enrolled templates. Moreover, as fingerprint systems occupy the largest market share, a probable threat to these systems may incur a tremendous financial loss. We thus deduced that investigating various security aspects of fingerprint biometric systems will be more impactful, as it would affect a larger community.

1.2 Motivations

The research accomplished in this thesis work is motivated by the following observations from the literature:

- The latest threat model for a biometric system appeared in 2008. However, researchers presented several potential threats and attacks on these systems during the last decade, such as side-channel attacks, which remained missing from existing threat models. As the system has evolved, essential components, such as a template generation module, became an integral part of the system. However, they did not appear in the earlier threat models. Hence, we require a thorough analysis and examination of the fingerprint biometric system to pinpoint each vulnerable component and identify every probable threat to them.
- Recent investigations unveiled the MasterPrint vulnerability for a partial fingerprint identification system. The study proved that a MasterPrint could benefit an adversary to mount a wolf attack and disclose the identity of several enrolled users.
- As the latent fingerprints are mostly partial, there exist chances of observing a similar threat using latent fingerprint, i.e., Latent MasterPrint. The investigations in this direction could motivate the forensic agencies to cross-check their existing AFIS against robustness towards MasterPrint vulnerability.
- The biometric systems usually apply preprocessing techniques to the sensed data. Hence, it becomes imperative to study the impact of these methods on the system accuracy and towards addressing the MasterPrint vulnerability.

The above mentioned points motivated us to perform a comprehensive security analysis of the match-in-database fingerprint biometric system, investigate the possibility of Latent MasterPrint, examine the impact of preprocessing approaches on the overall system performance while addressing the MasterPrint vulnerability, and provide solutions to address the MasterPrint vulnerability.

1.3 Objectives

Based on the motivation stated above, the objectives pursued in this thesis is as follows.

- Performing an in-depth study of fingerprint biometric system security. The task involves introducing a threat model depicting every vulnerable component as a potential

attack point with its respective threats.

- Experimenting with Multi-sensor Optical And Latent Fingerprint (MOLF) Database [9] to study the possibility of Latent MasterPrint generation utilising a standard algorithm for feature extraction and matching.
- Introducing novel local feature-based schemes to identify partial fingerprints uniquely that simultaneously alleviates the MasterPrint generation. We will use minutiae-based features for partial fingerprint identification and addressing the vulnerability. Further, the performance of the proposed approaches will be evaluated using standard fingerprint datasets, and the results will be compared against existing approaches from the literature.
- Evaluating the impact of various thresholding and thinning approaches during the preprocessing stage towards MasterPrint generation and identification accuracy. The experimental setup will employ four well-known thresholding and thinning methods from the literature.

1.4 Contributions of the thesis

This section presents the significant contributions of the research work on the security assessment of fingerprint biometric systems and introducing novel methods to mitigate the MasterPrint vulnerability. A graphical overview of the contributions made in this thesis is shown in Figure 1.2. A brief description of these contributions is as below.

1.4.1 Threat modeling of fingerprint biometric system

A threat model represents the system from an adversary's perspective to locate the structural vulnerabilities and probable threats to break the system security. We introduce a comprehensive threat model for the match-in-database fingerprint biometric system comprising sixteen attack points (AP). The proposed model highlights the existing and potential threats to six components and ten communication links between them. The model includes the template protection techniques module and the biometric system controlled application which were

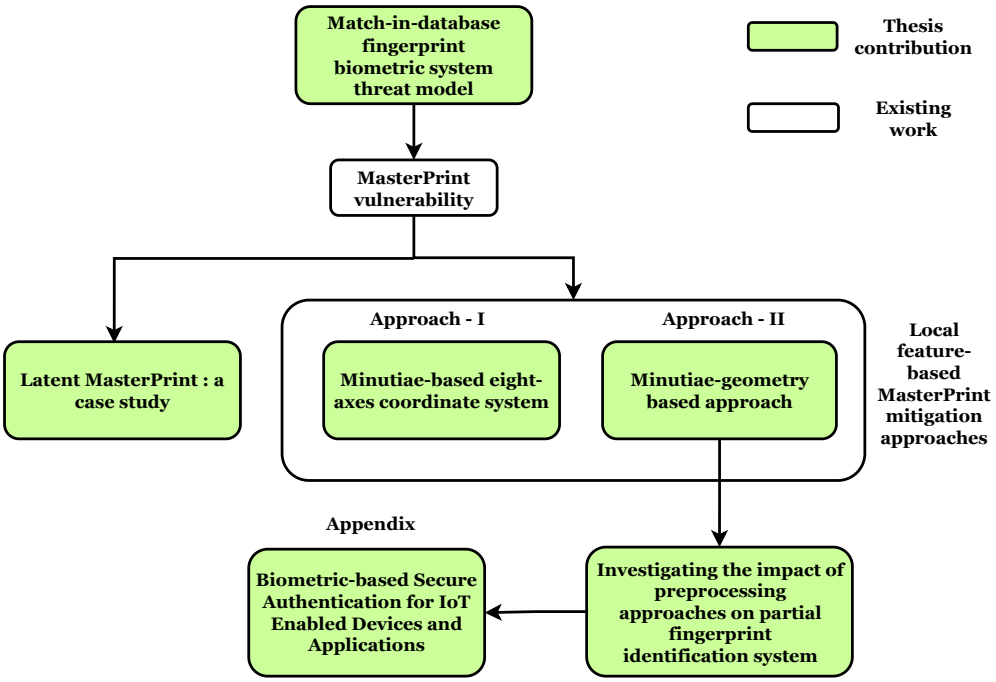


Figure 1.2: Graphical overview of the contributions made in this thesis

missing in the existing models. The description of the model includes the countermeasures to strengthen the security of the biometric system. The extension of the threat model comprises classifying the system threats into eight classes to facilitate securing multiple APs with a fewer mitigation techniques.

1.4.2 Latent MasterPrint: a case study

The investigation to study the possibility of Latent MasterPrint is still an open problem. Hence, in our subsequent work, we employ a latent fingerprint dataset under a closed-set identification environment to detect minutiae using the MINDTCT algorithm and perform matching using the BOZORTH3 algorithm. The experiments target assessing Latent MasterPrint generation and identification accuracy at various thresholds.

1.4.3 MasterPrint mitigation using minutiae-based coordinate system

In the first attempt towards MasterPrint mitigation, we create an eight-axes coordinate system around a reference minutia. The feature extraction in this case consists of the *intensity-shift*

over each axis in a *binarized* fingerprint image, and the *intensity-factor* over each axis in a *thinned* fingerprint image. The proposed method outperforms existing approaches and proved robust against MasterPrint vulnerability.

1.4.4 MasterPrint mitigation using minutiae geometry

In another attempt to further improve the identification rate and reduce MasterPrint generation compared to the first approach, we have presented a novel method to address the MasterPrint vulnerability. In this case, we employ geometric constructs involving minutiae as its vertices to create an integer-valued feature vector. The approach delivered better identification accuracy and significantly reduced MasterPrints.

1.4.5 Investigating the impact of preprocessing approaches on the performance of partial fingerprint identification systems

As fingerprint identification systems comprises several independent components having dedicated functions, it will be unfair to label the comparison module as the only component responsible for generating MasterPrints. We have experimented with various thresholding and thinning approaches in our subsequent task to analyse their effect on the system accuracy and percentage of MasterPrints generated in minutiae geometry-based approach.

1.4.6 Biometric-based secure authentication for IoT-enabled devices and applications

This work provided a broad review of the current scenario of Internet-of-Things (IoT) technology. It covered various types of IoT devices, and services and presented differences between them in terms of their functionalities. This work also mentioned the security issues in these devices and suggested various corrective and mitigation steps against any probable threats to such system. The study also addressed the need for a highly secure authentication mechanism like a biometric system to authorise and authenticate an individual. It emphasised

the benefits of employing a fingerprint-based biometric system in an IoT infrastructure.

1.5 Organisation of the thesis

This thesis comprises seven chapters including this introductory chapter. The flow diagram of the chapters is depicted in Figure 1.3. The description of the thesis organisation is as follows.

Chapter 2: Background

This chapter briefly introduces existing threat models and presents the proposed threat model comprising sixteen attack points. It further includes classifying various system threats into eight classes to facilitate the system designer with countermeasures for similar threats at different locations within the system.

Chapter 3: Threat Analysis of Fingerprint Biometric system

The third chapter introduces the recently researched MasterPrint vulnerability as an alarming threat to the partial fingerprint biometric system. As the latent fingerprints are mostly partial, we further investigate the possibility of Latent MasterPrint.

Chapter 4: MasterPrint Mitigation Using Minutiae-based Coordinate System

The fourth chapter presents a minutiae-based approach for addressing the MasterPrint vulnerability. The experimental setup consists of partial datasets cropped from five standard fingerprint datasets. The results of the proposed method are compared with existing five schemes.

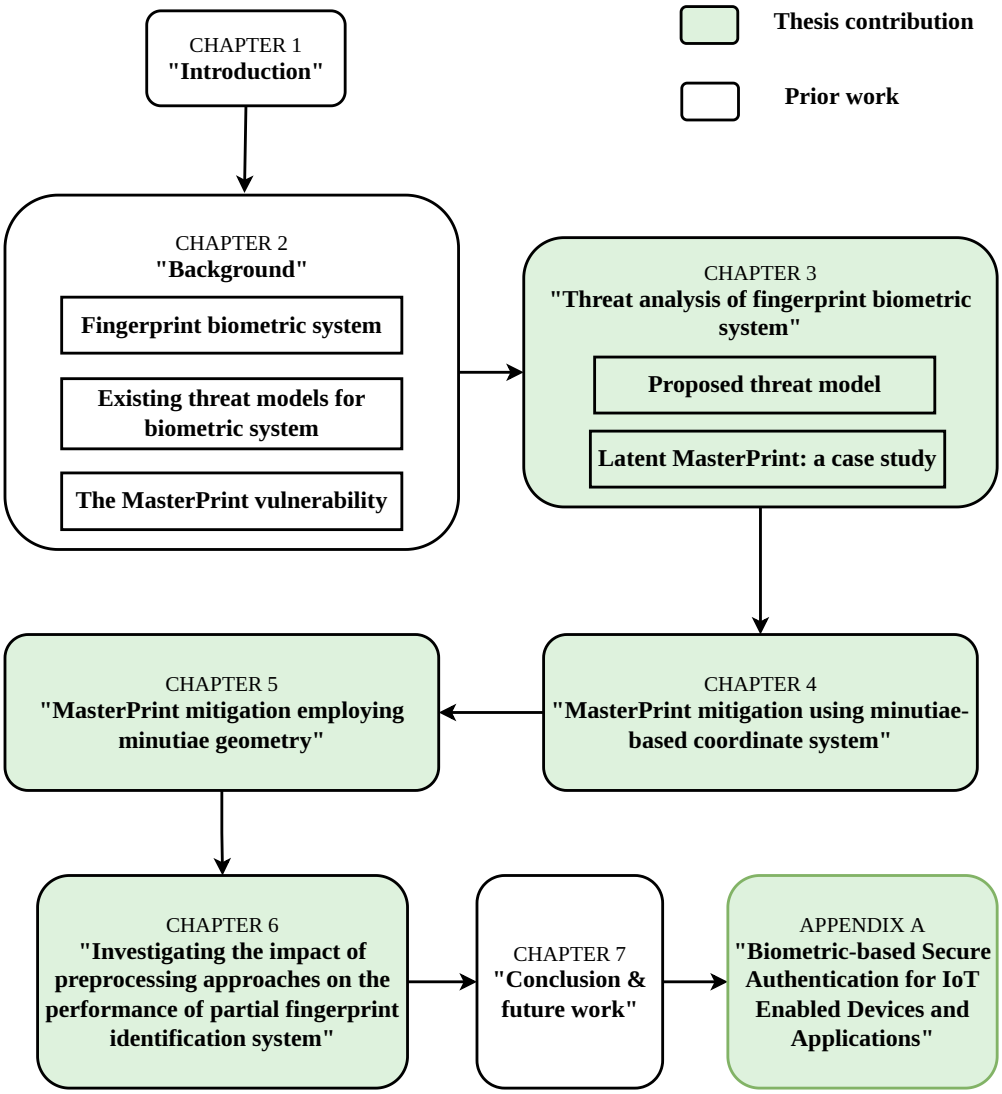


Figure 1.3: Flow diagram of the thesis

Chapter 5: MasterPrint Mitigation Employing Minutiae Geometry

The content of this chapter is another attempt to thwart the MasterPrint vulnerability. We have used minutiae geometry to create a novel feature vector which solved our primary aim of addressing the MasterPrint vulnerability and delivered high identification accuracy. The results of the proposed approach are compared with existing six methods.

Chapter 6: Impact of Preprocessing Approaches on the Performance of Partial Fingerprint Identification System

This chapter includes the experiments conducted using the method introduced in the fifth chapter to assess the impact of various combinations of thresholding and thinning approaches towards evaluating the identification accuracy and addressing the MasterPrint vulnerability.

Chapter 7: Conclusions and Future Work

This chapter concludes our work on the security aspects of the fingerprint biometric systems and the MasterPrint vulnerability. We further provide future research directions to extend the work presented in this thesis.

Appendix A: Biometric-based Secure Authentication for IoT Enabled Devices and Applications

This work mentioned the security issues in IoT-enabled devices and applications and suggested various corrective and mitigation steps against any probable threats to such system. The study also addressed the need for a highly secure authentication mechanism, such as a biometric system to authorise and authenticate an individual. It emphasised the benefits of employing a fingerprint-based biometric system in an IoT infrastructure.

Chapter 2

Background

Biometric systems utilise personal traits, such as fingerprint, face, iris, palmprint etc., of an individual for authorisation and person identification. The system stores the features extracted from these traits in encrypted format as a *template*. These systems typically comprises a sensor to capture the personal traits, a feature extractor, a matcher module, and a template storage device. The biometric systems are broadly categorised as *match-in-database* (MiD) or *match-on-server* (MoS) and *match-on-device* (MoD) systems. The main difference between them is the location where templates are being stored. In the case of MoS systems, a remote server contains the template database. However, MoD systems require a dedicated chip on a smart card to store the user's biometric template. These systems can be further categorised as *template-on-card* (ToC), *match-on-card* (MoC), and *system-on-card* (SoC). In case of ToC, the smart card contains only the encrypted template. The MoC system contains user template and the matching component on the smart card. However, SoC is a complete biometric system on the smart card, wherein the encrypted template, matching component, and the biometric sensor for collecting the live fingerprint are present on the smart card itself. As MiD biometric systems are widely used due to user convenience compared to MoD systems this chapter briefly describes MiD systems, and various existing threat models to them. Furthermore, it discusses the MasterPrint vulnerability associated with the partial fingerprint identification systems

The organisation of this chapter is as follows. The chapter can be roughly divided into three parts. The first part provides a detailed description of MiD fingerprint biometric system in Section 2.1. The second part presents existing threat models to generic biometric systems

and discusses the MasterPrint vulnerability in the third part. The Ratha et al. model appears in Section 2.2.1. The Fishbone model follows in Section 2.2.2. A description of the Nagar et al. model features in Section 2.2.3. Further, Section 2.2.4 explains the Bartlow and Cukic framework. The details about the MasterPrint vulnerability comes under Section 2.3. Finally, Section 2.4 provides the chapter summary.

2.1 Fingerprint Biometric System

A fingerprint biometric system utilises digital imaging techniques for fingerprint acquisition, storage, and user verification or identification requiring minimal human efforts. These systems use unique, obscure, and permanent patterns from the biometric data to confirm the user identity. The primary motive behind administering these systems is to provide non-repudiable authentication [10]. Over the past two decades, biometric systems have become the first choice application as a trusted mechanism for various organisations, including the financial and health sector, education, government offices, military and border security applications, etc. [6]. The consumer industry has witnessed a significant increase in implementing biometric authentication to replace traditional password-based schemes for electronic equipment, such as, smart doors, smartphones, access control measures, tablets, etc. These systems also enable controlling frauds in government schemes, reminding vaccination schedules, tracking missing children, etc. [11].

Typical components of a fingerprint biometric system are depicted in Figure 2.1. The fingerprint image is captured when the user touches the biometric sensor with his finger. The quality of the sensed finger depends on various environmental and physiological factors, including the pressure exerted by the user, sweat or dryness on the finger, cut or injured finger portion, etc. Hence, usually, an image enhancement module is employed to improve the fingerprint quality and facilitate accurate feature extraction. The binarization process involves converting the grayscale fingerprint into a black and white image. A sample binarized fingerprint image is shown in Figure 2.2. The dark lines form the *ridges*, whereas the intermediate white portion are termed as *valleys*. The point where two ridges emerge or terminate is the most commonly employed and accepted feature known as a *minutia* [6]. The fingerprint

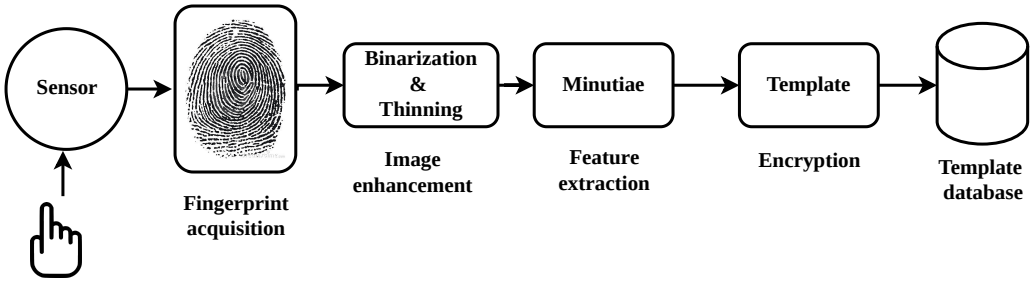


Figure 2.1: Components of a typical fingerprint biometric system

Figure 2.2: A typical fingerprint pattern showing ridges, valleys, and minutia.

also possesses other features such as *core*, *delta*, an *island*, a *lake*, etc. These features are categorised as local and global. A typical fingerprint often exhibits the local features in large quantity, such as minutiae. The global fingerprint features comprise *singular point*, *core*, and *delta*. The *NIST Fingerprint Image Quality* (NFIQ) level is a standard metric to measure the quality of a fingerprint image [12]. The "NFIQ value" ranges from $1-5$, where a fingerprint with an NFIQ value of 5 indicates a poor-quality image.

Typically a biometric template possesses encrypted information essential for comparison with an input fingerprint template, i.e., it does not include details about the original fingerprint. The ISO/IEC JTC1/SC37 standardisation has provided three formats for fingerprint templates, namely *minutiae-template*, *pattern-template*, and *image-template* [2]. The minutiae coordinates, orientation, and types, such as *ridge bifurcation* and *ridge ending*, are stored in a *minutiae template*. The encrypted information about the ridge patterns, such as *loop*, *arc*, *whorl*, etc., comprises the contents of a *pattern template*. However, the unadapted output of a sensing device, such as, the fingerprint impression in its scrambled format is stored as an *image template*. The adversary must be refrained from predicting or reverse-engineering

the original biometric features, ridge patterns or their close replica through the encrypted templates [13]. Thus, it is presumed that the *template generation* module operates as a "one-way" scheme, such that, retrieving actual fingerprint details from the encrypted templates is practically infeasible [14].

The properties to be satisfied by a protected and secure biometric template are as follows: [10]:

1. It ensures that any attempt to produce the actual biometric data is practically infeasible,
2. It safeguards user privacy,
3. It guarantees marginal variation in similarity score due to acquisition fault or environmental factors,
4. It prohibits the reuse of a secured template at several biometric controlled applications without the user's consent.

Typically, the fingerprint biometric system operates in two phases, namely, *enrollment* and *authentication*. It can function as a verification or identification system during the authentication phase. A generic procedure employed during the enrollment and authentication phase of a match-in-database fingerprint biometric system is depicted in Figure 2.3. The enrollment procedure is usually carried out under the supervision of a system administrator. The biometric sensor captures the finger impression. The system then applies image enhancement techniques to the sensed fingerprint pattern to facilitate accurate feature extraction in the subsequent stage. The binarization technique creates a black and white ridge pattern, whereas the thinning procedure reduces these patterns to single-pixel width to initiate minutiae detection. The template generation module utilises minutiae-based features to construct a secured and scrambled template. The system administrator completes the enrollment process by assigning a unique user identity against the stored template. These templates are stored in a remote database server and retrieved during the verification and identification phase.

The template database in Figure 2.3 stores p templates for n subjects. In general, the system administrator enrols a finger two to four times for user convenience during the authentication phase. There are s templates per subject stored in the database. Thus, the database

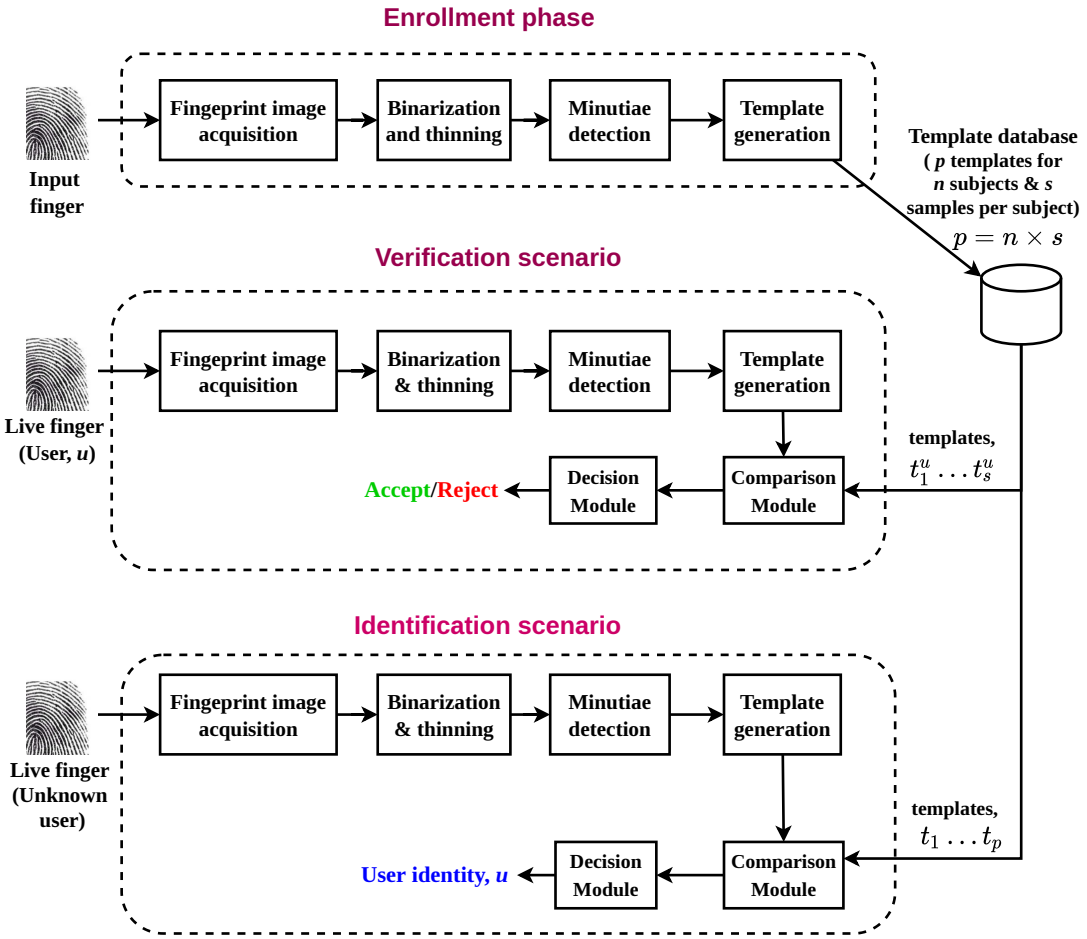


Figure 2.3: A typical minutiae feature-based match-in-database fingerprint biometric system with enrollment phase, authentication scenario, and identification scenario. The system is assumed to possess p templates corresponding to n subject and s samples per subject are collected during enrollment phase, i.e., $p = s \times n$.

comprises a total of $p = s \times n$ templates. During the verification phase, the user claims his identity, u, while submitting the finger to the sensor. The system retrieves s templates belonging to the claimed identity, i.e., $t_1^u, t_2^u, \ldots, t_s^u$, to verify the claim. The similarity score between these templates is used to decide if the claim is true or false. However, the identification system accepts the fingerprint of an unknown user to determine his identity based on the comparison with each stored template, i.e., $t_1, t_2, \ldots, t_p$. The system assigns the identity of the stored template to the unknown user that generates the highest similarity score with the live fingerprint.

The fingerprint samples of a finger acquired at multiple instances are not always same. Hence, the comparison module allows tolerance while matching the templates. However, the system may not consistently deliver accurate results due to intrinsic limitations, human

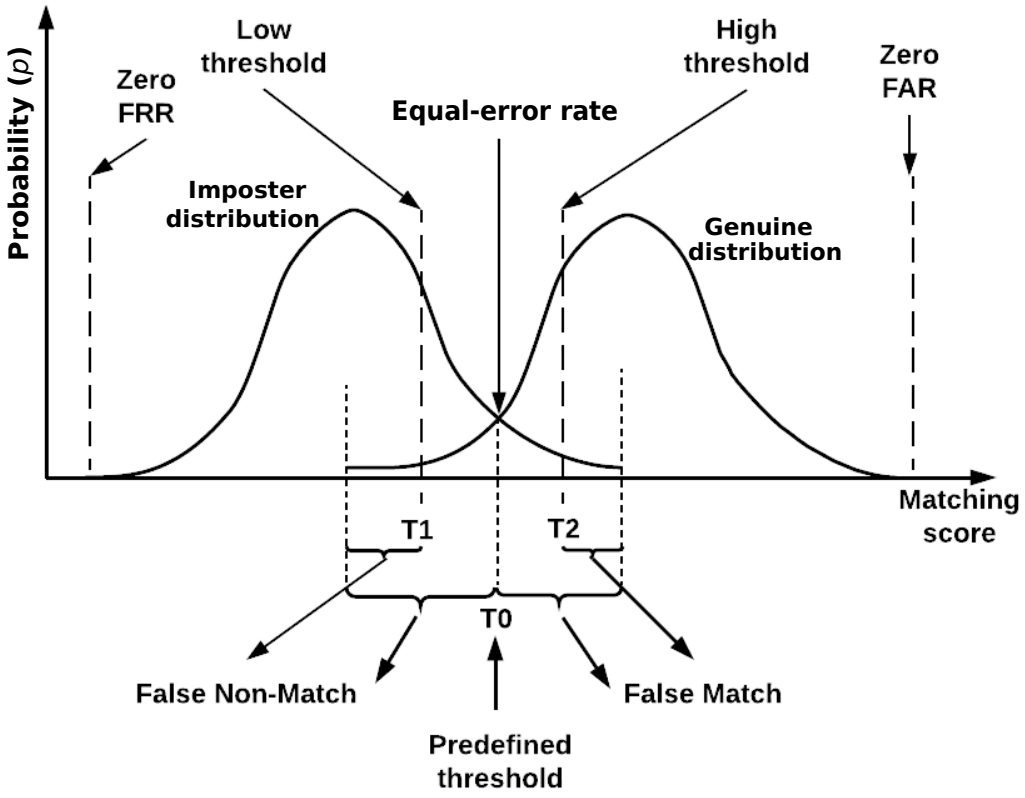


Figure 2.4: Error distribution curves depicting genuine and imposter distributions for a biometric system

error, or environmental factors [15]. The system is thus prone to several errors that reduce its accuracy. A false match or false acceptance error occurs when the system confirms the identity of a non-enrolled individual. A *false match rate* (FMR) or *false accept rate* (FAR) is defined as the percentage of instances when the system accepted an intruder. A false non-match or false reject instance involves rejecting the claim of an enrolled user. A *false non-match rate* (FNMR) or *false rejection rate* (FRR) is defined as the percentage of instances when the system rejected a registered user. The *system threshold* determines the FAR and FRR. Typically, an organisation requiring a biometric system for high-security applications will set a high threshold so that the false acceptance would be a rare instance. However, an administrator at a customer service centre or a staff attendance application would lower the threshold for the convenience of enrolled users. The typical range of FAR and FRR applicable to high-security fingerprint recognition systems is 10^{-6} and 10^{-4}, respectively [16].

Figure 2.4 shows the FAR and FRR as a function of the system threshold. Impostor distribution represents the spread of matching scores for non-enrolled users. Genuine distri-

bution defines the spread of matching scores for enrolled users. The point where the curves for these distributions intersect is termed as *crossover error rate* (CER) or *equal-error rate* (EER). It reflects the accuracy of a biometric system in general. The figure showed that FAR and FRR arise due to system limitations, such as false acceptance and false rejection. If the *system threshold* is increased from EER, the FMR reduces whereas the FRR increases. On the other hand, if the system threshold is decreased from EER, FRR reduces and FMR increases.

2.1.1 Security Vulnerabilities in fingerprint biometric system

Biometric systems are extensively employed in academia, industry, government, and military applications due to their reliability over pass-code and password-based security mechanisms. However, each component in a biometric system is vulnerable to multiple security threats of varying intensity. Li et al. [17] used reconstructed fingerprints for demonstrating a successful presentation attack. The authors utilised the minutiae points to reconstruct the full fingerprints. Initially, a binary ridge pattern having a ridge flow similar to the original fingerprint was created. The reconstruction of the continuous phase was carried out by eliminating the spirals from the phase image derived from the ridge pattern. Finally, a phase refinement process was introduced to diminish the artefacts observed due to intervals in the recreated phase image. The demonstration also showed that the biometric systems employing *amplitude* and *frequency* (AM-FM) modulated fingerprint models are susceptible to the presentation attack through reconstructed fingerprints. Rozsa et al. [18] illustrated the synthesis of fingerprint minutiae templates using a genetic algorithm attack. The authors analysed the impact of synthesised minutiae templates on the security and privacy aspects of specific fingerprint authentication schemes by studying the patterns of high similarity scores. The system was designed to attack through the communication channel. It successfully mounted an offline attack on *Minutia Cylinder-Code* (MCC) [19] and *Protected Minutia Cylinder-Code* (PMCC) [20] systems through MCC SDK v1.4 [21].

Pashalidis [22] presented simulated annealing attack that produced synthesised, gummy, ISO/IEC standard agreeable minutiae templates. These templates resemble and match with

specific unknown target fingerprints. Vicinity-based matching algorithms initially divide the fingerprints into several regions and compute the similarity score over these regions. The paper demonstrated the construction of a minutiae template which are expected to be most effective against systems that employ vicinity-based matching algorithms, such as *Protected Minutiae Cylinder Code* (PMCC) scheme. The fuzzy vault scheme uses cryptographic operations to secure the biometric templates [23]. However, investigations confirmed that a template created by employing the fuzzy vault scheme could be unlocked using the correlation attack [24]. The adversary in the attack scenario collects two fuzzy vaults assuming that they are encrypted using the same biometric data. The correlation analysis of these two templates reveal the protected information from the vaults. In another investigation, the vaults implemented by employing three distinct schemes were found vulnerable to brute-force attack [25].

In September 2017, the investigations on the partial fingerprint identification system exposed the MasterPrint vulnerability [7]. A partial fingerprint fortuitously identifying at least 4% unique subjects enrolled with the system was termed a *MasterPrint*. It was demonstrated that the vulnerability facilitates mounting a wolf attack on these systems. In this attack, the identity of a large percentage of enrolled users can be revealed by employing a dictionary of the top five MasterPrints that identify maximum number of distinct subjects. The work presented in this thesis provides mitigation schemes to address the MasterPrint vulnerability.

2.2 Existing threat models for biometric system

Threat modelling involves locating vulnerable components within a system and identifying potential structural defects and threats, or prioritising the needfulness of desirable security precautions. The literature on the biometric system shows that the threats and vulnerabilities to these systems were extensively studied until 2008. The literature comprises four threat models presented for a generic biometric system. The following subsections describe these models in detail. The models depict the vulnerable components of the system assuming that the adversary is well-equipped to access these components for mounting specific attacks.

2.2.1 Ratha et al. model

Ratha et al. [1] proposed the first model in 2001 locating eight vulnerable information access points as shown in Figure 2.5. The model provides a generic view of the system with single attack at each vulnerable point. The adversary exploits these sensitive system components to mount a specific attack, characterised as Type 1 to Type 8.

- *Type 1 attack* : According to the model, sensing devices possess Type 1 threats through *spoofing*, *false biometric submission*, and *residual attacks*. In the case of residual attack, the adversary accesses the system's main memory to collect the temporarily stored template and recreates the original fingerprint to mount a presentation attack at the sensor. The potential attack includes presenting fake biometrics to the sensor. The adversary acquires the biometrics of an enrolled user to access the system using his dummy physical finger.

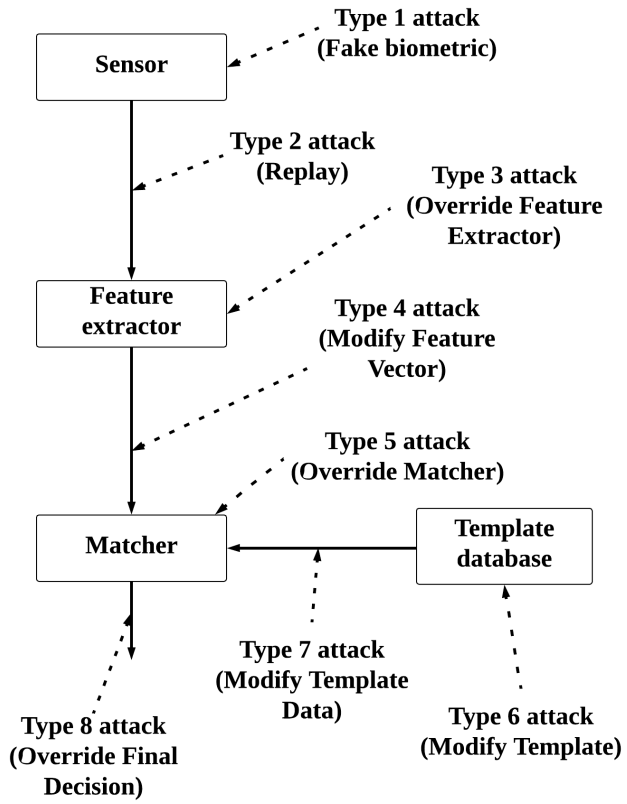


Figure 2.5: Ratha et al. model [1]: The dashed arrows indicate the access point of information for mounting a specific type of attack

- *Type 2 attack* : The possibility of replaying a previously captured biometric data through the communication link connecting the sensing device and the feature extraction component is designated as Type 2 attack. The attacker uses the previously collected digitised biometric data bypassing the sensor to mount such an attack. Masquerading is an example of a Type 2 attack.

- *Type 3 attack* : The model shows the possibility of overriding the output of the feature extraction module as Type 3 attack. A Trojan horse executing in the feature extraction component is exercised to mount such attacks. The adversary replaces the feature extraction module output with his chosen feature set. The adversary targets the software or firmware of the biometric system to execute a Trojan horse attack [26]. A Trojan horse is a malicious code that stays hidden or idle once implanted and can be triggered to activate its malicious behaviour at a given time in the future. Once triggered, the Trojan can delete, copy, or modify the data from the targeted system component.

- *Type 4 attack* : The adversary targeting the communication channel connecting the feature extraction and the matcher module to modify the feature vector is termed a Type 4 attack. The probability of success for a Type 4 attack is marginal as these two modules are commonly implemented as a single component. The attacker can use the channel to mount brute-force and hill-climbing attack as a Type 4 threat to the system.

- *Type 5 attack* : The model shows an attempt of overriding the output of the matcher module as Type 5 attack. The similarity score computed by the matcher module can be manipulated using a Trojan Horse executing on it.

- *Type 6 attack* : A Type 6 attack includes unauthorised access of the template database records. The attacker attempts unauthorised access to the remote template database through an external compromised system to fraudulently steal, delete, modify, substitute, and reconstitute the stored templates. The adversary exploits a bug in the database software, log information on the disk, information in cache, or primary memory to mount a direct or indirect attack on the database. Further, the adversary explores the possibility of template database leakage, gathers the leaked templates, and replays them to get authorised.

- *Type 7 attack* : The threat of modifying the template data in transit over the communication link connecting the template database and the matcher module is labelled as Type 7 attack. An attacker eavesdropping on the channel manages to collect the templates under transmission. These templates or their altered copies are utilised in mounting a replay attack on the matcher module.

- *Type 8 attack* : In the case of a Type 8 attack, the attacker tries to override the system decision. The intruder circumvents all system components to influence the decision while mounting such attacks. Hence, a successful Type 8 attack proves that there is no effectiveness in providing excellent countermeasures to secure earlier system components.

Ratha et al. [1] model is a generic and simplistic model that specified eight components and a few threats to these components. However, it did not mention the link between the feature extraction module and the template database as a part of the enrollment process. The feature extractor follows the template generation module wherein the encrypted templates are produced. The template generation module is also missing in the model.

2.2.2 The fishbone model

In 2008, Jain et al. [2] presented a cause and effect based fishbone model. The model depicting various causes leading to the vulnerabilities is shown in Figure 2.6. The oval shaped boxes in the figure represent the causes of system faults. The arrows denote the effects, i.e., possible attacks due to these causes.

- *Intrinsic failure* comprises the system errors accountable for an individual's false acceptance or rejection. A user placing an inappropriate or partial finger at the sensor produces variations in intra-user feature extraction. Hence, significant deviations in the registered user's input and stored features stimulate false rejection. The matcher limitations include false acceptance and false rejection. The sensor may capture partial fingerprint due to environmental factors and improper finger placement at the sensing device leading to intrinsic failures in the biometric systems.

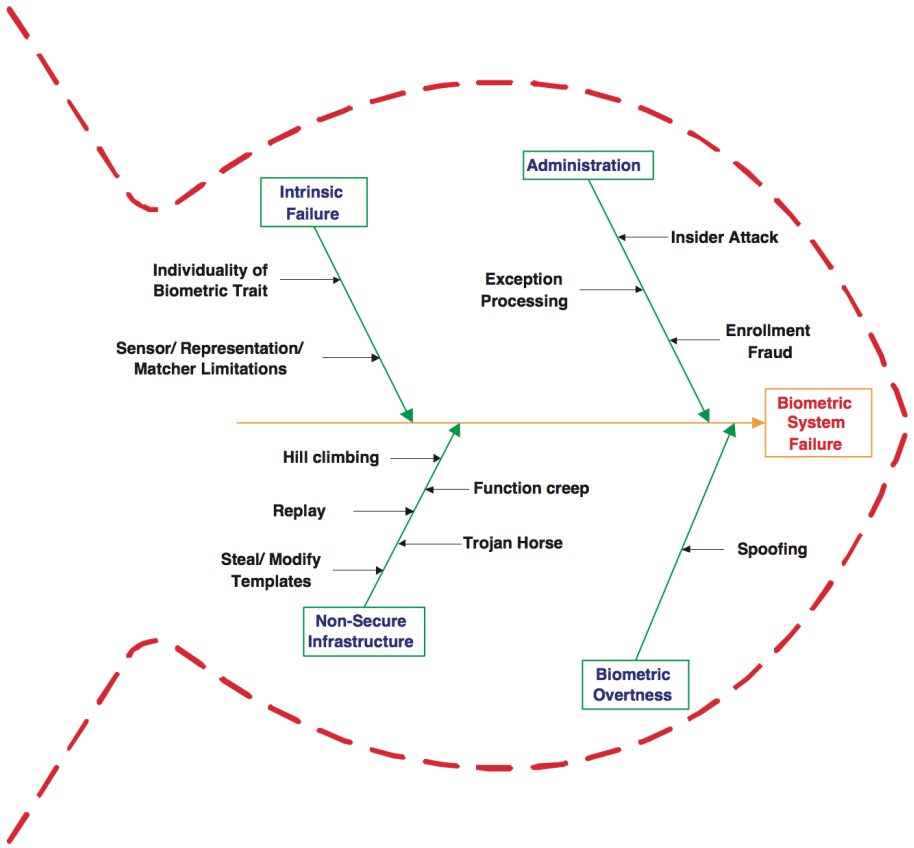


Figure 2.6: The fishbone model [2]. The blocks denote the processing phases of a biometric system that are vulnerable targets in this model. The possible attacks are mentioned on the arrows.

- *Administration cause* represents a dishonest staff, such as a system administrator, acting on behalf of an adversary. A traitorous system administrator may cooperate with an intruder to perform exception abuse or an enrollment fraud. Exception abuse is an emergency service provided to the enrolled users to facilitate system access in the case of a denial-of-service attack or repeated rejection due to system limitations. The vengeful administrator may enrol the adversary with the system to carry out illicit activities for financial benefit. The insider attacks can cause severe financial damage to an organisation since a system administrator knows the system internals.

- The system vulnerabilities arising as consequences of authentication or an insecure enrollment process comprise *insecure processing causes*. In the case of a function creep scenario, the adversary retrieves the enrolled user's template through database leakage or intercepting over the communication channel and performs a cross-comparison of their biometric features. Additionally, such illegally accessed templates can be reused

for other biometric controlled applications such as banking and passport services. The adversary mounts a hill-climbing attack to create a fake fingerprint template that generates a similarity score more than the system threshold. He uses these templates for replay attack to access the system under the identity of an enrolled user.

- The biometric template secrecy relates to the *biometric overtness*. The adversary utilises the biometrics of an enrolled user collected from public places to create an artificial or a gummy finger mould. Usually, clay, plastic, wood glue, or gelatin materials are used for such practices [27]. These moulds are exercised to mount a spoofing attack. The spoofed fingerprints are presented to the biometric sensor to access the system with the identity of an existing enrolled user.

2.2.3 Nagar et al. model

Nagar et al. model [2] is an extension of the fishbone model as shown in Figure 2.7. The model depicts the factors giving rise to threat opportunities for the adversary. The factors are specified near the dotted ovals, and their vulnerabilities leading to possible attacks are displayed inside the ovals. The rectangles in the figure refer to primary constituents of the biometric system under consideration.

- The adversary exploits the insecure infrastructure, such as the communication channel, to access sensitive data. The non-secure processing and the infrastructure factors in the fishbone model are merged in this category. The threats arising from non-secure infrastructure include replay attack, hill-climbing attack, and Trojan horse attack. The feature extraction and comparison module are vulnerable to Trojan horse attack. The adversary replaces the feature extraction module output with his chosen feature set and targets the software or firmware of the biometric system to execute a Trojan horse attack [26]. A Trojan horse is a malicious code that stays hidden or idle once implanted and can be triggered to activate its malicious behaviour at a given time in the future. Once triggered, the Trojan can delete, copy, or modify the data from the targeted system component.

- The threat involving the *overtness of biometrics* includes utilising publicly available

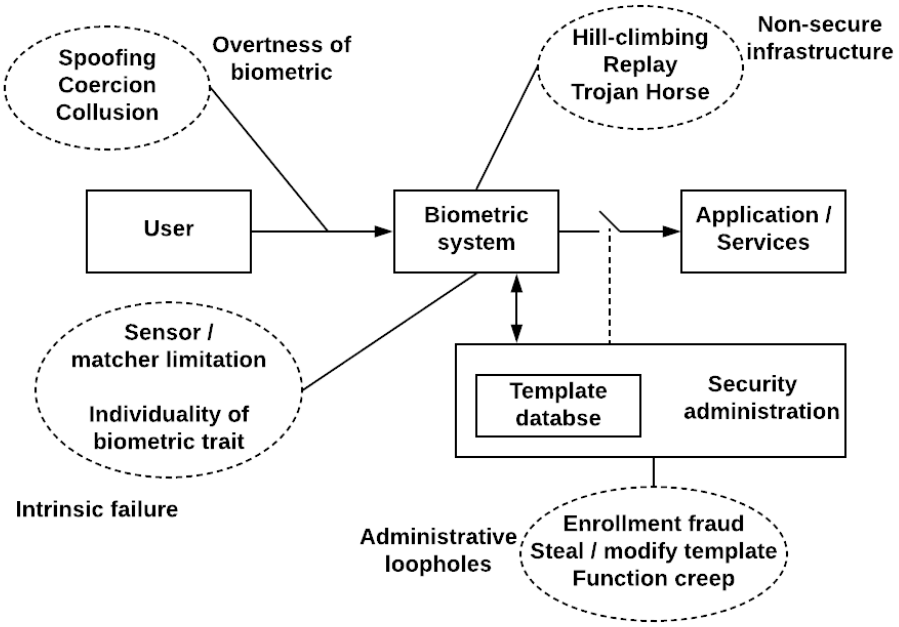


Figure 2.7: Nagar et al. model [2]

information about an individual's biometrics, such as, face image cropped from photos on the Internet, finger impressions left on various surfaces, voice captured during a telephonic conversation, etc. The adversary targets the input device to mount a spoofing attack or coercion as a potential threat under the overtness of biometrics.

- *Intrinsic failure* comprises the system constraints authorising a non-registered user by false acceptance. The system threshold primarily decides the authenticity of every user trying to access the system. The system administrator of an organisation with thousands of employees may set the system threshold to a low value and provide an opportunity to an outsider having marginally similar features with an enrolled user's fingerprint template to enter the premises.

- A disloyal employee can exploit *administrative privileges* to harm the organisation. The administrator cooperates with an adversary in collusion, enrollment fraud, and mounting a spoofing attack.

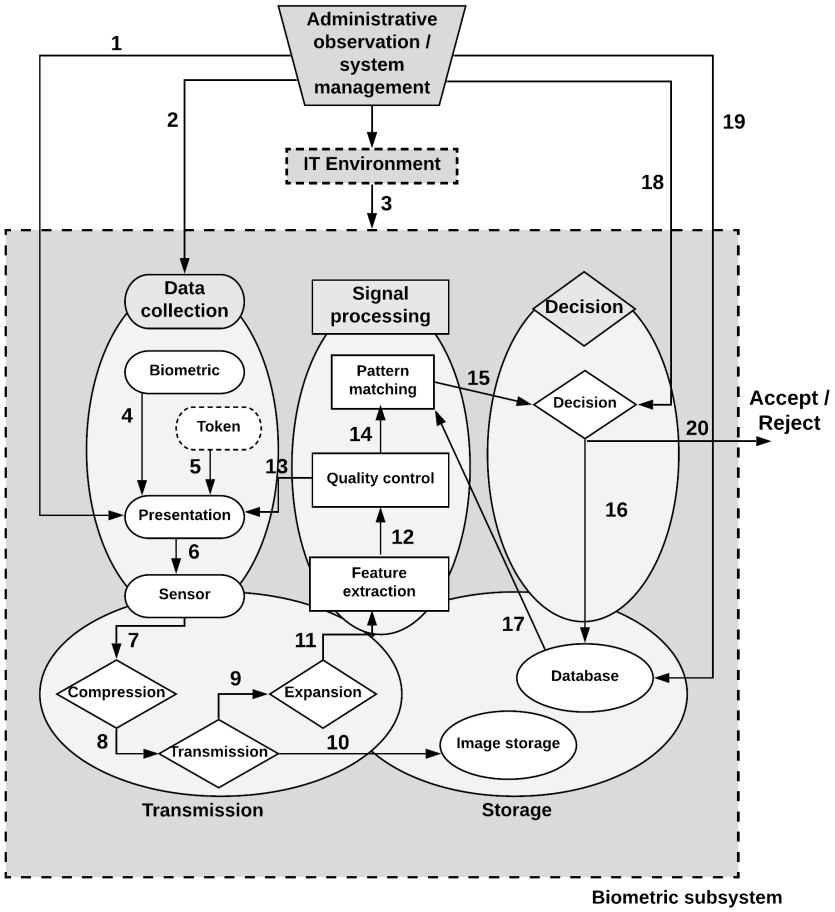
2.2.4 Bartlow and Cukic framework

In 1996, Wayman suggested classifying a biometric system as a composition of five basic subsystems according to their application [28]. These subsystems include data collection, transmission, signal processing, storage, and decision. Figure 2.8 depicts the Bartlow and Cukic framework [29] [26] specifically considering the technical testing of biometric devices as an extended version of Ratha et al. model and Wayman's subsystem architecture. The framework consists of three modules shown in dark grey colour and identified more than twenty vulnerable potential attack points. According to Wayman's suggestion, it further categorises the system management module into five subsystems shown in large oval boxes. The arrows indicate the vulnerable points within the subsystems and modules. The vulnerabilities across the system management module can be described as follows.

- The *administrative observation/system management module* consists of a deceitful system administrator. The arrows originating from this module represent the potential attack points targeted at different components by the administrator. An adversary seeking cooperation from an administrator to attack the data collection module is designated as attack points 1 and 2. However, the attacker manipulates the decision through attack points 18 and 20. An attempt by the corrupt administrator to carry out false enrollment is denoted as an attack point 19.

- The *IT environment subsystem* comprises operating systems and database management systems interacting with the biometric system. A potential threat from these applications is designated as attack point 3. An intruder can execute malicious code or illegally invade the system through these system components.

- The *biometric subsystem* comprises the data transmission between the internal components and their vulnerabilities. The *data collection* module is responsible for collecting and presenting user biometric data. The input device capturing the biometric data is accountable for attack points 4, 5, and 6. The adversary can mount replay, masquerade, and spoofing attacks on the data collection module. The transmission module compresses and transmits the biometric data. This module is susceptible to *parallel*



1 - Bad admin; 2 - Bad admin, fail secure, power, bad user, undetect bypass; 3 - Bad admin, fail secure, power, bad user, undetected, bypass, corrupt, degrade, tamper, residual, crypt attack; 4 - Casual, artifact, regeneration, mimic, evil twin; 5 - Bypass, replay, fake template; 6 - Tamper, replay noise; 7 - Tamper, residual, crypt attack, replay, noise; 8 - Crypt attack; 9 - Crypt attack, replay; 10 - Regeneration, replay; 11 - Crypt attack; 12 - Tamper, replay, noise; 13 - Poor image; 14 - Replay, noise; 15 - Tamper, replay; 16 - Casual, regeneration; 17 - Casual, regeneration, Weak ID; 18 - Bad admin, casual, evil twin; 19 - Bad admin, evil twin; 20 - Bad admin.

Figure 2.8: Bartlow and Cukic framework. The numbers $(1 - 20)$ represent potential attack points. The biometric system is divided into three modules, namely, biometric subsystem module, IT environment module, and administrative observation/system management module. The biometric subsystem module is divided into five subsystem, namely, data collection, signal processing, decision, transmission, and storage subsystem.

sessions, *masquerade attacks*, and *replay attacks* through attack points 7, 8, 9, 10, and 11. The attacker eavesdropping on the communication channel gathers the data for mounting a *brute-force attack*, *hill-climbing attack*, and *replay attack*.

Attack points 12 and 14 are vulnerable to Trojan horse attacks targeting the quality control and feature extraction modules, respectively. The adversary can use a bad quality image

to execute a *hill-climbing* and *brute-force* attack at attack point 13. The similarity scores collected by mounting a *hill-climbing* attack at the comparison module facilitate the construction of the fingerprint image. The adversary may generate several enhanced fingerprint images by applying image enhancement techniques on a poor quality image and employ them in a brute-force attack.

The threats at attack point 15 represent the data interception during transmission between the decision and the comparison module. The attack points 16 and 17 report acquiring template data from an insecure database and replaying on the comparison module. An attacker targeting to override the outcome of a decision module with the assistance of a system administrator is depicted as attack point 20. The model presented potential threats as attack points at various components within the biometric system. However, it did not suggest the strategies for implementing security schemes for these systems [30]. Therefore, the model appears as a benchmark for validating and testing the effectiveness of security techniques.

2.3 The MasterPrint vulnerability

The system administrator usually seeks at least ten partial fingerprint samples from a user to improve the system accuracy and ensure unique user identification. However, the system responds with a decisive match to every comparison between the input and stored template that generates the similarity score above the system threshold. Moreover, the uniqueness of a partial fingerprint has not been affirmed yet as in the case of the full fingerprint. Hence, the system replies with positive decisions for several stored templates belonging to diverse subjects. The scenario thus triggers a non-unique user identification through a partial fingerprint as a potential system vulnerability.

Roy et al. [7] investigated the partial fingerprint identification system and discovered that, as expected, the system does not perform unique user identification. The authors coined the term *MasterPrint* for a partial fingerprint that can identify at least 4% distinct subjects enrolled with the system. A MasterPrint was defined in correlation to a random guess of a 4-digit PIN. The term implies that a MasterPrint should possess similarity with at least 4% enrolled users, where 4% denotes the average chance of guessing a random 4-digit PIN.

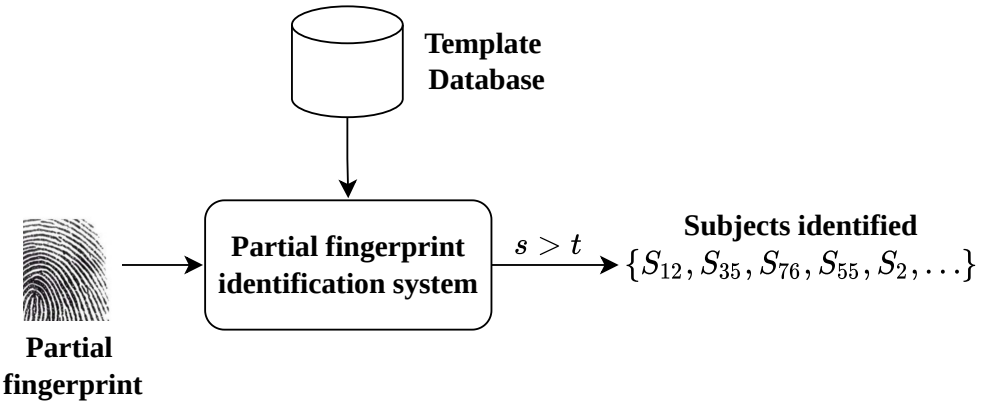


Figure 2.9: A MasterPrint scenario: Assuming that the system is enrolled with 100 subjects, S_1, S_2, ..., S_{100}. Here, s is the similarity score between input partial fingerprint and database template, t is the system threshold.

A sample scenario depicting a probable MasterPrint is shown in Figure 2.9. The system is enrolled with 100 subjects, S_1, S_2, ..., S_{100}. The given input partial fingerprint shows a similarity score, s, above the system threshold, t, for more than 4% of the enrolled subjects. Hence, it becomes a *MasterPrint* for this identification system.

2.3.1 MasterPrint Existence Hypothesis

The MasterPrint existence hypothesis states that the probability of observing MasterPrint in the partial fingerprint dataset is higher than in the full fingerprint dataset. Suppose there are N subjects whose J fingers are to be enrolled such that K samples per finger should be registered for high accuracy. The full fingerprint dataset, F^i_{jk}, is represented as $\mathcal{F} = \{F^i_{jk} | i \in \{1, \ldots, N\}, j \in \{1, \ldots, J\}, k \in \{1, \ldots, K\}\}$. Each fingerprint from F^i_{jk} has dimension $W \times H$. Let N_G represent the number of full fingerprints per subject, and N_T denote the total full fingerprints. Thus, $N_G = J \times K$ and $N_T = N \times N_G$.

Biometric devices with small sensors often capture partial fingerprints of an individual. Suppose each full fingerprint is trimmed into L partial fingerprints with dimension $w \times h$ such that $w < W$ and $h < H$. Thus, the partial fingerprint dataset, PF^i_{jkl}, is increased L times and represented as $\mathcal{F}' = \{PF^i_{jkl} | i \in \{1, \ldots, N\}, j \in \{1, \ldots, J\}, k \in \{1, \ldots, K\}, l \in \{1, \ldots, L\}\}$. However, the L fold increase in the dataset size adds the count of imposter and genuine scores for every input fingerprint under the all-versus-all comparison. Con-

sequently, the likelihood of observing chameleons showing high imposter similarity scores may increase significantly, leading to false matches.

2.3.1.1 Hypothesis

Let $P(MP \subset \mathcal{F})$ and $P(MP \subset \mathcal{F}')$ be the likelihood of observing MasterPrint (MP) in $\mathcal{F}$ and $\mathcal{F}'$, respectively. The null hypothesis is stated as:

$$H_0 : P(MP \subset \mathcal{F}) \geq P(MP \subset \mathcal{F}') \tag{2.1}$$

Thus, the hypothesis holds only if H_0 is rejected.

2.3.1.2 Hypothesis test

The hypothesis test was conducted on FVC 2002 DB1_A dataset comprising 800 full fingerprints belonging to 100 users with eight samples per subject. A partial fingerprint dataset of dimension $w \times h$ was created by cropping the full fingerprint dataset. A 50% overlap between adjacent partial fingerprints was ensured while cropping. Moreover, the window size was set similar to the Apple Touch ID, i.e., 150×150. On average, the resultant dataset comprised ten partial fingerprints for each full fingerprint.

The full and partial fingerprints possessing a minimum of ten minutiae were exercised during comparison. The partial fingerprint matching was carried out using the VeriFinger 6.1 SDK. The full fingerprint dataset results showed 0.1% average *false match rate* (FMR) and 99.18% average *true match rate* (TMR) while generating 0.125% MasterPrints. However, 14.63% MasterPrints were observed for the partial fingerprint dataset. The t-test was applied to these results to strengthen the hypothesis further. The observed p-value was less than 2.2×10^{-16} when the expected confidence level was 0.05. The empirical p-value was $\ll 0.5$. So, H_0 can be rejected. The results thus confirm the alternative hypothesis that the probability of MasterPrints in partial fingerprint datasets is higher than in the full fingerprint dataset.

2.3.2 MasterPrint generation

Roy et al. [7] explored two methods of MasterPrint generation. The methods were termed as *Sampled MasterPrints* (SAMPs) and *Synthetic MasterPrints* (SYMPs). These methods were designed for fingerprint authentication systems employing minutiae-oriented features.

2.3.2.1 Sampled MasterPrint generation

The SAMPs were selected from real fingerprint datasets. In this case, the MasterPrints were sampled from a fixed-sized training dataset. *Imposter Match Rate* (IMR) represents the count of false matches while a fingerprint is compared with other finger images, i.e., imposters. The IMR is computed for every candidate fingerprint. Let $S((\chi),(i,j,k,l))$ denote the *similarity score* between χ and F^i_{jkl} and θ be the system threshold. The $IMR(\chi)$ is defined as,

$$IMR(\chi) = \frac{1}{(N-1)\cdot L\cdot N_G}\sum_{\forall i,j,k,l}\phi((\chi),(i,j,k,l))$$

where,

$$\phi((\chi),(i,j,k,l)) = \begin{cases} 1 & \text{if } S((\chi),(i,j,k,l)) > \theta \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

The fingerprints producing maximum IMRs were selected as SAMPs. However, the IMRs of these SAMPs may be affected when applied to other fingerprint datasets due to variations in fingerprints captured from various sensors, noise, and quality of images. The existence of SAMPs in a given dataset is not always possible. Hence, the feasibility of constructing synthetic MasterPrints from SAMPs was explored.

2.3.2.2 Synthetic MasterPrint generation

The goal behind creating SYMPs was to maximise the IMR of SAMPs that generates better MasterPrints artificially from training over a fingerprint dataset. Let Σ denote all SAMPs, and ω represent a candidate MasterPrint such that $\omega \in \Sigma$. The aim is to obtain SYMP, ω', from Σ that maximises the objective function given below,

$$\omega' = \underset{\omega \in \Sigma}{\operatorname{argmax}} \{IMR(\omega)\} \tag{2.3}$$

As probing over the entire search space is challenging, a local search is preferred to find the solution. Synthetic fingerprints generated using a hill-climbing attack have a high probability of being falsely accepted by the fingerprint matcher [31]. Hence, the hill-climbing attack was utilised for SYMP generation using SAMPs as the initial seed. In a hill-climbing attack scenario, a synthetic minutiae template is generated randomly and submitted to the fingerprint matcher [32]. The similarity score is collected to alter the template iteratively until a specific condition is accomplished. During the SYMP generation process, the SAMPs are modified such that the minutiae location and orientation are changed, or the minutiae count is decreased or increased. The IMR value drives this process as stated in equation 2.3.

The average count of minutiae in a partial fingerprint is often low, i.e., 20 minutiae on average [33]. Hence, to ensure that the minutiae count in SYMPs is not lowered further, a lower bound is set on the number of minutiae in SYMPs. The modified minutiae template replaces the current template only if its IMR is improved. Thus, during the "hill-climbing" process, the IMR increases gradually. The SAMP undergoes modification until the predetermined maximum IMR is achieved or it completes all the iterations.

2.3.3 Experimental results

The experimentation was performed in two scenarios: *image-level comparisons* and *finger-level comparisons*. In the case of image-level comparisons, the best SAMPs were employed for "all-versus-all" matching. However, during finger-level comparisons, selected SAMPs

and their corresponding SYMPs were used in attacking the enrolled subjects from the test set. The datasets employed in the investigations comprise the fingerprint of a single finger per subject. Hence, the terms "finger" and "subject" are used interchangeably in this chapter. The value of IMR defines the accuracy of an attack using a MasterPrint.

Consider f distinct fingers and i samples collected per finger. The dataset size d is defined as $d = f \times i$. Suppose a given MasterPrint is matched with p images that belong to q fingers. The image-level comparison considers the count of images matched by a MasterPrint. However, the finger-level comparison checks the count of fingers successfully matched by a MasterPrint. Thus, the image-level and finger-level comparison results for the mentioned example will be p and q, respectively.

2.3.3.1 Image-level comparison

An adversary mounting a dictionary attack using MasterPrint will try a set of predetermined fingerprints. Hence, a fingerprint dictionary comprising MasterPrint that ensures a gradual increase in the chances of matching a large number of target fingerprints is required. The dictionary of MasterPrint formed by *independent SAMP selection* consists of the five sampled MasterPrints corresponding to the partial fingerprint having the highest IMRs. *Sequential SAMP selection* is another approach for creating the dictionary. Initially, the partial fingerprint showing the highest IMR was chosen. The selected MasterPrint and all the partial fingerprints identified by it were omitted from the dataset to create a new training set. The IMR of each partial fingerprint in the new dataset was computed, and the fingerprint with the highest IMR was chosen as the second element of the dictionary. A similar approach was repeated until five partial fingerprints were selected for the dictionary.

2.3.3.2 Finger-level comparison

The *finger-level comparison* represents the scenario of replicating a dictionary attack where the adversary is unaware of the fingerprint dataset. In this case, a MasterPrint is compared with multiple templates of partial or full fingerprints, and a match is declared for every instance when the imposter match score is higher than the predefined threshold.

2.3.4 Result analysis

Roy et al. [7] studied the partial fingerprint identification by employing a 150×150 pixel partial fingerprint dataset cropped from standard FVC2002 DB1_A and FingerPass DB7 capacitive partial dataset. A partial fingerprint possessing at least ten minutiae was considered during the research. The commercial fingerprint identification software, *VeriFinger* 6.1 *SDK*, was used for user identification from partial fingerprint. As an unexpected result, the SDK produced 14.63% MasterPrints from the FVC dataset. The study further revealed a dictionary of the top five MasterPrints identified 65.2% users from the FVC dataset. Hence, a probable wolf attack through a dictionary of MasterPrints was another alarming concern exposed during the research. The authors also proved that a hill-climbing attack could facilitate the synthetic MasterPrint generation. The important outcome from this work is summarised as follows:

1. The results confirmed the possibility of a dictionary attack involving meticulously selected MasterPrint with high accuracy. The MasterPrints can be sampled from an existing dataset or may be developed synthetically by employing a hill-climbing attack. The likelihood of MasterPrints from the partial fingerprint dataset and their attack accuracy is significantly higher than in the full fingerprint datasets.

2. Consider a scenario involving a dictionary of five MasterPrints from a partial fingerprint dataset and allowing at most five authentication attempts. In this process, 26.46% enrolled users in the FingerPass DB7 dataset and 65.20% enrolled users in the FVC dataset were attacked successfully. However, the attack accuracy deviated considerably with the change in FMR value and the count of samples per finger.

3. The investigation demonstrated that the attack accuracy involving sampled MasterPrints could be improved using synthetic MasterPrints created by applying a hill-climbing approach. The observation showed that the average increase in the accuracy on the FingerPass dataset was $\approx 4\%$, whereas on the FVC dataset was $\approx 3\%$.

4. The study on minutiae distribution of chosen MasterPrints indicated that upper delta points within the fingerprints form the high minutiae activity area. Cao et al. [34]

reported that such minutiae often possess low discriminative power leading to a high probability of matching with an imposter.

5. The results showed that the number of distinct subjects identified by a MasterPrint could be high even if it matches a small percentage of partial fingerprints. The risk increases as multiple samples are enrolled per finger for each subject. This observation implies that the type and number of partial fingerprint templates must be selected with prudence to lower the chances of matching an imposter [32].

2.3.5 Potential threats due to MasterPrints

An enrolled user may hurriedly submit a fingertip or incomplete portion of the enrolled finger at the biometric input sensor. The input device exposed to the external environment may not be able to capture a full fingerprint impression due to small sensor size. In such cases, a partial fingerprint is captured and compared with the enrolled templates. An adversary discovering a biometric system vulnerable to MasterPrints can exploit the system controlled application in several ways. As the MasterPrint acts like a master key, the adversary conducts illegal activities for personal benefit or harm others. A robber may use a MasterPrint to unlock a smart home security application. The *automated teller machines* (ATMs) accept user fingerprints and PINs for improved security while authorising a transaction. However, an innocent person may unknowingly share his partial impression with a criminal who successfully gathered other credentials for his stolen credit or debit card. As a result, the person's partial fingerprint acts as a MasterPrint to empower an illegal financial transaction to withdraw money from the victim's bank account. Thus, a presentation attack employing a MasterPrint has a high success probability at a biometric-based access control application.

Bontrager et al. [35] presented two approaches for synthetic MasterPrint generation employing a Generative Adversarial Network (GAN). The approach created image-level MasterPrints, which produced complete images capable of mounting a successful dictionary attack on the biometric system. Further, Roy et al. [36] demonstrated using evolutionary search algorithms for synthetic MasterPrint generation with minimum parameter tuning. The au-

thors were able to generate highly accurate MasterPrints capable of attacking unknown subjects. A dictionary of the top five such MasterPrints successfully attacked 84% of users in the FVC dataset. Thus, creating a fake, synthetic MasterPrint is feasible.

Smartphone vendors usually save manufacturing costs or reduce design complexity by offering small sensing areas to capture the user's fingerprint as an authentication mechanism and provide access to the device. However, as these small-sized sensors are expected to capture the partial fingerprint of a user, the MasterPrint vulnerability poses a severe security threat to such devices. Synthetic MasterPrint further increases the chances of being a victim of a presentation attack through these artificially generated and highly precise MasterPrints. A vengeful individual may leave behind a synthesised MasterPrint that matches an innocent person at a crime site to escape from being caught or entrap the innocent person. The judiciary accepts forensic reports on fingerprint matching as supportive evidence instead of any conclusive evidence from planting MasterPrint. However, a MasterPrint can be exploited by a criminal to strengthen the evidence against an innocent. Moreover, IoT devices have been in high demand recently. The smart devices vendors provide security features in IoT infrastructure by employing a small sensor-based fingerprint biometric authentication procedure. However, a MasterPrint can be effectively exploited in breaking these smart devices. In a nutshell, any device providing partial fingerprint-based identification and potentially vulnerable to MasterPrints can be targeted by the adversary. Hence, research efforts to alleviate the MasterPrint vulnerability has become essential to ensure consumers trust these applications.

2.3.6 Future research direction on MasterPrint vulnerability

The investigations on MasterPrint vulnerability confirmed that such MasterPrints could be generated by employing feature and image-level synthesis [36]. Further, Bontrager et al. [35] termed *DeepMasterPrint* to those MasterPrints generated by employing a complete image-level synthesis method. The authors utilised the latent variable evolution technique to create DeepMasterPrints from partial fingerprint images. The work by Roy et al. [7] opened up several research directions which can be summarised as below,

1. exploring the effect of the distribution of partial fingerprints on the accuracy of attacks

using MasterPrints,

2. improving the methods of synthetic MasterPrint generation,

3. constructing digital artefacts by applying image-level modifications to create physical artefacts for utilising in mounting a spoofing attack [37].

2.3.6.1 Addressing MasterPrint vulnerability

The authors advised viewing the MasterPrint vulnerability from a broader perspective for developing an accurate and reliable user identification system employing partial fingerprints. Some of the suggestions in this direction are summarised as below,

1. designing and developing robust anti-spoofing techniques [37],

2. precisely choosing the samples and characteristics of partial fingerprints during enrollment [38],

3. improving small-sized sensor resolution to promote extracting highly discriminative features [39],

4. developing novel feature extractor and matcher modules that employ minutiae and ridge texture features [40],

5. developing robust and efficient fusion techniques that integrate the information from several partial fingerprints of a user [41], [42].

Roy et al. [43] introduced a template selection scheme to alleviate MasterPrint vulnerability. The approach applies the maximum coverage (MC-K) algorithm to represent a finger by choosing appropriate partial fingerprint templates with minimal parameter tuning. The authors evaluated the performance of the approach on FVC 2002 DB1_A and the FingerPass DB7 datasets utilising a commercial fingerprint matcher. The results proved the resilience of the scheme towards MasterPrint generation. The method reduced the MasterPrint generation by 8% and 6% in the FingerPass DB7 and FVC dataset at 0.1% false match rate (FMR), respectively.

2.4 Summary of the chapter

The chapter initially provided the details about the fingerprint biometric system, discussed the identification and verification scenarios and types of errors, and reported recently investigated attacks on these systems. The chapter also presented existing threat models to a biometric system. The MasterPrint vulnerability and potential threats to a partial fingerprint identification system due to the existence of MasterPrints were discussed in detail. The chapter thus provided the background to the work presented in the subsequent chapters of the thesis.

Chapter 3

Threat Analysis of Fingerprint Biometric System

In the previous chapter, mechanisms about existing threat models to generic biometric systems were presented. As these models were proposed more than a decade ago, they do not comprise recently investigated threats, such as the MasterPrint vulnerability exploited to mount a presentation attack. Moreover, several other attacks targeted against different components of these systems have been demonstrated and investigated since the year, 2018, when the recent threat model was presented. The study and analysis of these threats and vulnerabilities have not been reported. The details about every attack at a vulnerable point are missing in the description of the existing models. Moreover, the literature on these models do not comprise countermeasures and does not illustrate the internal components of the models. In this chapter, we propose a comprehensive threat model uncovering vulnerable system components and potential attacks on a match-in-database fingerprint biometric system. Subsequently, we conduct a case study to investigate the possibility of Latent MasterPrint.

The organisation of the chapter is as follow. Section 3.1 presents the proposed threat model. We discuss the attack scenarios at sixteen attack points for the match-in-database fingerprint biometric system and present countermeasures to thwart such attacks in Section 3.1.1. Subsequently, we perform the threat analysis in Section 3.1.2. Section 3.2 presents the case study on investigating the possibility of Latent MasterPrint. Section 3.3 summarises the chapter.

3.1 Proposed threat model for MiD FBS

We can attribute the probable attacks to the design perspectives of the system that do not consider the potential threats during product development. We introduce our model in Figure 3.1, which demonstrate sixteen *attack points* (AP). The system components are depicted as closed shapes. The arrow connecting the components shows the output of one component as input to the next component over a communication medium. The dashed lines show the potential attacks at the respective vulnerable points. The numbers denote a possible attack point (AP).

The main contributions of the proposed threat model are as follows.

1. The model introduces the template protection techniques module and demonstrates its threats. Even though the module is an integral component in modern biometric systems it was missing in existing threat models. The module is vulnerable to side-channel attack as depicted in Figure 3.1.

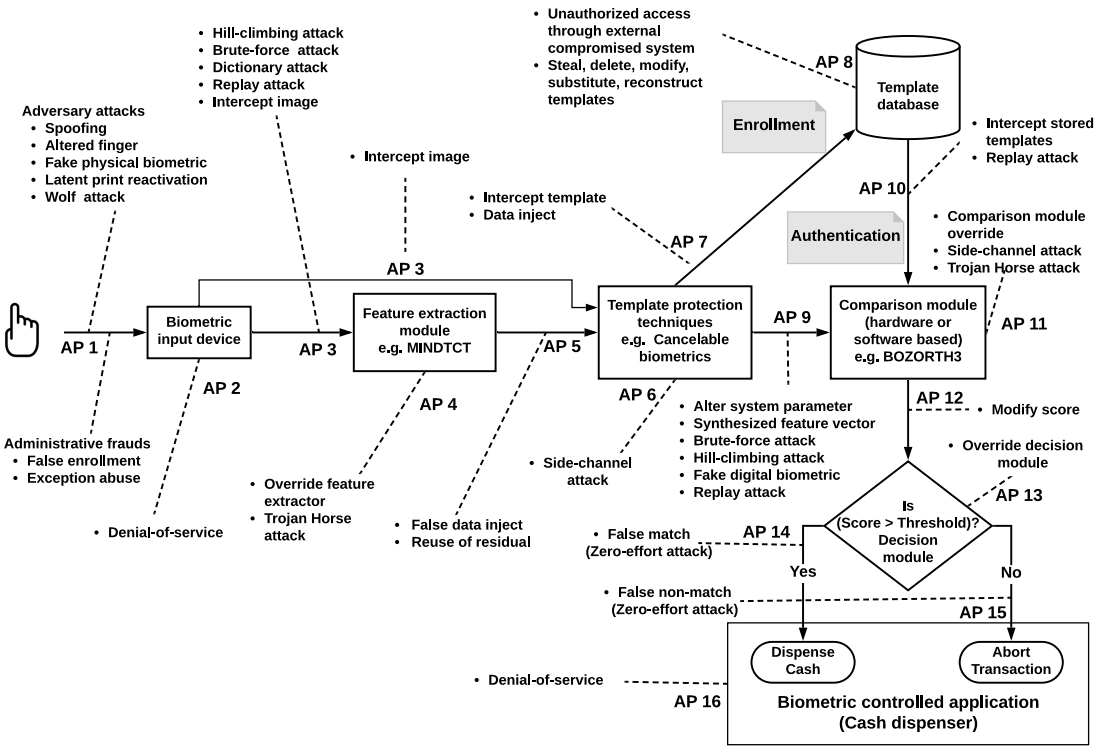


Figure 3.1: The proposed threat model: vulnerabilities and attacks on different components of a match-in-database fingerprint biometric system controlled application, for example, a cash dispenser machine.

2. The model depicts the side-channel attack at the vulnerable components of the biometric system. The existing models did not mention an attempt of a side-channel attack on the system components such as template protection techniques module and comparison module. The attacks are used to extract secret information, such as encryption key, from a vulnerable system component.

3. The model further reports an instance of an adversary intercepting the encrypted template over the communication channel connecting the template protection techniques module and the template database during enrollment. The enrollment stage does not require the matcher module as it creates the template and stores it in the database. However, existing models did not show the link connecting the template protection techniques module and the template database. The proposed model depicted this connection as AP 7 in Figure 3.1.

4. The model shows two possibilities to mount a denial-of-service (DoS) attack, a case that is not present in previous models. The adversary can destroy the sensing instrument or the application controlled by the biometric system to mount a DoS attack. The DoS possibilities were lacking in existing models.

5. The model shows the comparison module as a software or hardware component and pinpoints possible attacks in both the cases. The earlier models considered software-based comparison module and highlighted the threats to which it is vulnerable. However, as hardware-based comparison modules are also employed in biometric application, the proposed model showed the vulnerabilities to software and hardware-based comparison module.

3.1.1 Attack scenarios and countermeasures

The biometric system threats are broadly classified as *direct attacks* and *indirect attacks* [44]. A direct attack involves utilising the publicly available system components to either access the system or damage the system to mount a DoS attack [45]. The *indirect attacks* assume the adversary to have knowledge and expertise in biometric systems [46]. The vulnerable

system components and the insecure communication channels provide multiple opportunities to such an adversary. The following subsections consider each *attack point* (AP) separately to explain the threat model.

3.1.1.1 AP 1 (Attacks through biometric input device)

We categorise an attempt to break the system security as adversary attacks and insider attacks, i.e., administrative frauds. The adversary attacks comprise latent print reactivation, spoofing attacks, fake physical biometric presentation, altered fingerprint submission, and a wolf attack using MasterPrints. The adversary lifts the latent fingerprint left behind by an individual at public places and creates a dummy fingerprint. In a spoofing attack, the attacker manages to access and reverse engineer the stored template to construct an artificial spoofed fingerprint and uses it at the sensing device. Alternatively, the adversary collects and analyses the similarity score for an existing fingerprint and alters the fingerprint until it is accepted by the biometric system. Matsumoto et al. [47] demonstrated that eleven different fingerprint recognition systems accepted artificial fingers. The adversary benefiting an exception abuse and false enrollment are some examples of administrative fraud. The adversary exploits the emergency authentication facility provided to enrolled users to access the system under exception abuse with the help of an existing employee. However, a system administrator may perform a fake enrollment of the adversary into the system and grant essential privileges to perform financial frauds. Figure 3.2 shows various techniques employed by an attacker to spoof a legitimate user's finger. In a cooperative method, an individual assists in constructing a synthetic imprint of his fingerprint. The adversary uses plaster or dental impression material to make a direct mould of a legitimate user [3]. The non-cooperative methods involve creating artificial fingerprints using the impressions left on the surface at public places.

Galbally et al. [48] evaluated fingerprint verification systems for vulnerability against a direct attack using fake fingertips created from minutia templates. The biometric system adopts an anti-spoofing scheme shown in Figure 3.3 to thwart a spoofing attack. The hardware-based anti-spoofing solutions are costly since they require adding a new hardware component to the biometric sensor. Additionally, the newly added hardware component may leak confidential biometric information to an attacker. Hence, hardware-based anti-spoofing countermeasures are seldom preferred. A spoofed and live fingerprint can be distinguished

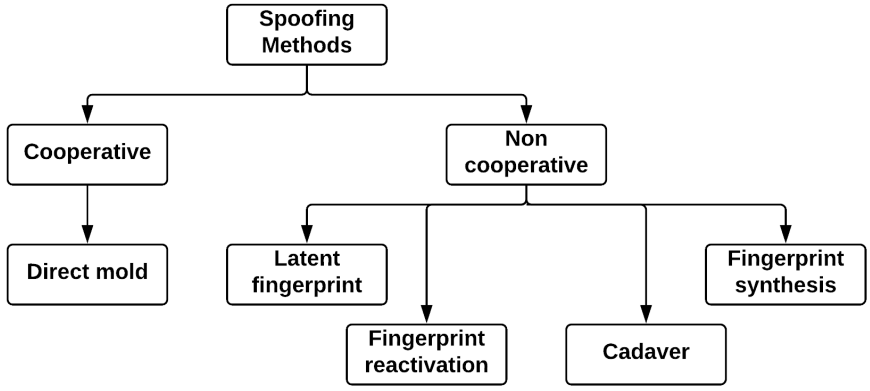


Figure 3.2: Methods employed for creating artificial fingerprints [3]

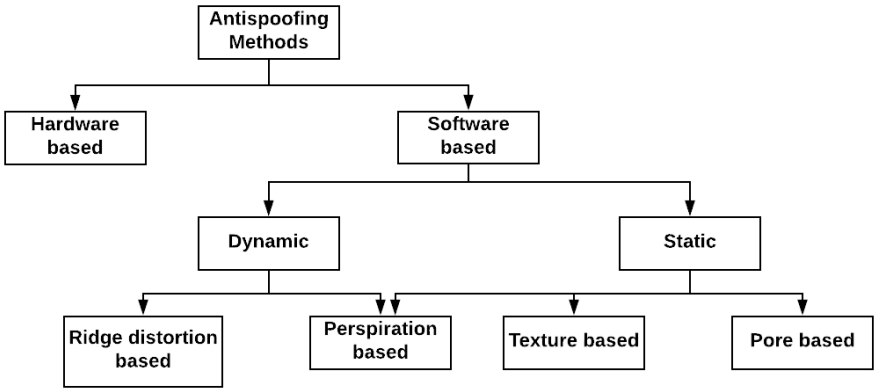


Figure 3.3: Anti-spoofing approaches [3].

using a method to identify and locate pores on it [49]. The method uses a single fingerprint image to detect multiple static features comprising pore spacing. The liveness score was utilised to classify a fingerprint as fake or live. The approaches to detect a live fingerprint utilises local quality features [50], local texture features [51], and gradient-based texture features [52].

3.1.1.2 AP 2 (Attack on biometric input device)

Denial-of-Service (DoS) attack is the most optimal and least complex way to destroy the entire biometric-controlled application. A novice person can damage the biometric sensing device and prohibit enrolled users from accessing the system. Usually, the input device is kept in an open environment as it must be accessible to the users. Hence, security personnel or CCTV camera can be deployed to monitor such attempts. Also, the vendors can employ

rugged devices that can tolerate an attempt of physical damage.

3.1.1.3 AP 3 (Attacks through communication channel connecting input device)

In a hill-climbing attack, the adversary obtains adequate information about the system modules to learn the attack success probability. A dictionary-based guessing attack is a sophisticated brute-force attack. In this case, the attacker chooses only those fingerprints with a high acceptance rate. Ratha et al. [53] provided the correlation between the matches that occurred for various brute-force attempts. In a replay or false data injection attack, an attacker circumvents the scanner, and uses an existing digital image of a legitimate user to get authorised and access the system.

A time-out-lock-out policy allows a specific number of unsuccessful attempts in a short duration and locks the biometric system for a random time span. An encrypted communication channel that disallows traffic from any other source is a primary measure to deal with the hill-climbing attack. Coding techniques, such as Slepian–Wolf coding, superposition coding, and universal channel coding, can be applied to secure the channel and stop reverse-engineering the traffic over the communication medium. Transferring digital data, such as files, over a *secure connection* in a computer network to ensure protection against snooping is termed *Transfer Encryption* [54]. *Secure communication* protocols including SSL (*Secure* Socket Layer), TLS (Transport Layer Security), SSH (*Secure* Socket Shell), HTTPS (Hyper Text *Transfer* Protocol over SSL/TLS) transfer encrypted data over the *communication* channel.

3.1.1.4 AP 4 (Attacks on feature extraction module)

The adversary replaces the feature extraction module output with his chosen feature set. The adversary targets the software or firmware of the biometric system to execute a Trojan horse attack [26]. A Trojan horse is a malicious code that stays hidden or idle once implanted and can be triggered to activate its malicious behaviour at a later time. Once triggered, the Trojan can delete, copy, or modify the data from the target system component. Figure 3.4 shows a scenario of a Trojan horse attack.

Trojan horse attacks can be repelled using a *trusted biometric system* (TBS) comprising various modules that perform mutual authentication, and thus are physically and logically

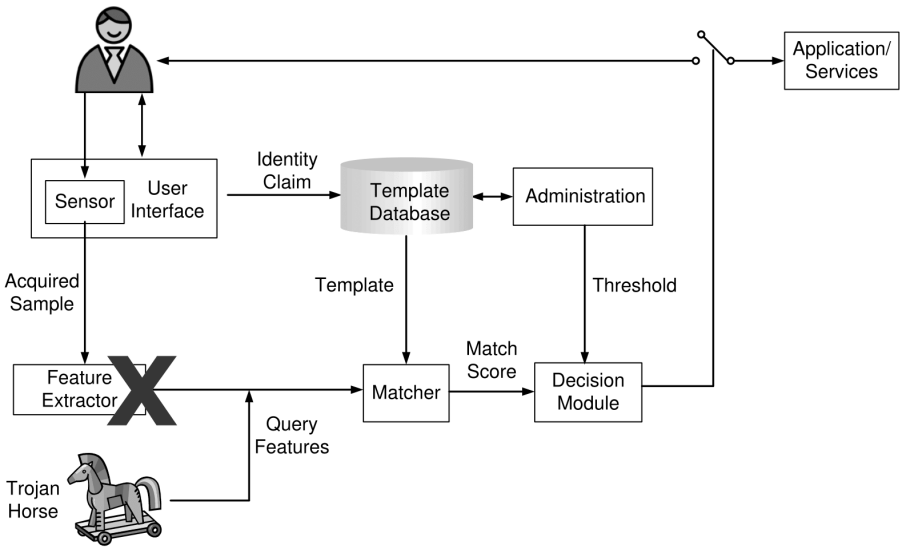


Figure 3.4: The feature extraction module as a target of a Trojan Horse attack [4]

bound together. The manufacturers can use code-signing, light-weight digital signature algorithms, such as, WalnutDSA [55] or Mave to ensure the integrity of the executable code. Specialised tamper-resistant hardware, such as, IBM 4758, iButton, Dallas 5002, etc., prohibits alteration of the module functionalities through enforcing secure software execution.

3.1.1.5 AP 5 (Attacks through outgoing communication channel connecting the feature extraction module)

The adversary uses the communication link leaving the feature extraction module to inject fake or altered data into the system. The feature extraction module holds the features in its primary memory before forwarding it to the template generation module. A technically sound attacker can access the features for a legitimate user and replay it at a later instance to access the system. The stolen features also facilitate mounting a hill-climbing attack to create a synthesised feature vector.

The system implementing a challenge-response technique detects an attack involving a synthesised feature vector. The biometric template retrieved from database usually resides in cache or main memory even after the comparison with input fingerprint template which can be reused by the adversary as residual to get authorised with the identity of an enrolled individual. The system memory must be flushed immediately after the feature extraction and comparison process to prevent reuse of residual. The feature extraction and template

generation module should be blended together, i.e., they should be implemented as a single component, to avoid leaking confidential information through the communication channel between them.

3.1.1.6 AP 6 (Attack on template protection techniques module)

The biometric system generates a secure digital template by applying cryptographic techniques, such as cancelable biometrics and biometric cryptosystem on the feature set retrieved by the feature extraction module [56]. The attacker targets the template protection techniques module to extract information related to the secret key used in the encryption algorithm. Side-channel attack is a class of techniques that target a specific circuit implementation on a technology platform to recover the secret key of the cryptosystem [57]. In these attacks, the adversary exploits the information leakage, through power consumption patterns, execution time variations, or the electromagnetic waves as a side-channel to recover the secret key [58]. Electromagnetic side-channel attacks involve exploiting the correlation between intermediate data dependent on the secret key and variations in electromagnetic emanations [59].

The system designers must address the power analysis attacks while designing a practicable cryptosystem as the information leakage is from the implementation aspect. The masking techniques provides a defensive mechanism against side-channel attacks [56]. The masking scheme generates a random variable r to protect secret information x. Instead of directly working on x, the scheme manipulates the masked data $x' = x \oplus r$ and r independently. Eventually, the scheme thwarts any first-order attack as every intermediate variable possesses a uniform distribution [60]. Integrated Circuits (ICs) with an active shield are intended to avoid side-channel leakages. The cryptographic algorithm implementation wherein all computations consume same power and use an equal number of clock cycles mitigates the side-channel threat. Yang et al. [61] introduced a robust fingerprint matching technique which guards the system against side-channel attacks.

3.1.1.7 AP 7 (Attacks through a communication channel connecting the template protection techniques module and the template database)

The attacker targets the communication channel between the template protection techniques module and the template database for data injection and intercepting the encrypted templates.

The link is explicitly utilised during the enrollment process. Hence, the attacker replays the intercepted template to get authorised as a legitimate system user.

A secure and encrypted communication channel that reveals no confidential information even after reverse engineering ensures that the system remains shielded against any attempt of eavesdropping and replaying. The system must implement a secure communication protocol that supports confidentiality and content-integrity protection.

3.1.1.8 AP 8 (Attacks on template database)

The attacker attempts unauthorised access to the remote template database through an external compromised system to fraudulently steal, delete, modify, substitute, and reconstitute the stored templates. The adversary exploits a bug in the database software, log information on the disk, or information in cache or primary memory to mount a direct or indirect attack on the database. The illegitimately modified templates are utilised in a replay attack on the system. Ross et al. [62] demonstrated the feasibility of an iterative hill-climbing attack to facilitate the fingerprint reconstruction process.

Li et al. [63] proposed a data hiding-based privacy protection scheme that ensures the impracticality of revealing confidential information from an encrypted template. The feasibility of a fingerprint reconstruction can be diminished by creating encrypted minutiae-based templates instead of stored minutiae-based. The countermeasures to alleviate intrusion attacks on the template database includes server hardening, i.e., enhancing the server security against potential vulnerabilities, database access control schemes, signing stored templates, storing encrypted templates, and using Match-on-Card technology.

3.1.1.9 AP 9 (Attacks through communication channel connecting template protection techniques module and comparison module)

The attacker targets the communication medium connecting the template protection techniques module and the comparison module to control the system parameters, such as system threshold, to enforce false acceptance. The adversary uses the channel to inject a synthesised feature vector as inputs to the comparison module. The communication link is also vulnerable to a hill-climbing attack for generating a synthetic template acceptable by the system. In this scenario, the attacker collects the similarity score for a random fingerprint

presented to the biometric system and alters it iteratively until the system accepts it. Usually, the adversary alters the minutiae in the fingerprint. The link is also exposed to fake biometric submission at the comparison module. Here, the adversary targets the biometric systems that perform image level comparison. An image collected by intercepting over the communication channel can be reused at the comparison module to access the system.

An attempt of code tampering and alteration can be detected and prevented by adopting code signing techniques. Oorschot et al. [64] and Chang et al. [65] explored the software code protection schemes against unauthorised access and modifications. The software integrity verification can be performed by applying the self-hashing technique [66]. Martinez-Diaz et al. [67] introduced a score quantization approach to lower the chances of succeeding in a hill-climbing attack. The robust digital defence mechanisms are helpful against an attempt of a fake biometric submission at the comparison module [26].

3.1.1.10 AP 10 (Attacks through communication channel connecting template database and comparison module)

An unguarded communication channel between the template database and comparison module is primarily utilised to intercept the stored templates. The attacker uses the captured templates to mount a replay attack later, either at the same link or some other vulnerable communication channel within the system.

The preventive measures to thwart any attempt to access the communication channel includes implementing a secure, robust, and encrypted communication medium within the system.

3.1.1.11 AP 11 (Attacks on comparison module)

The comparison module in the biometric system can be a hardware or software component susceptible to side-channel attacks and Trojan horse attack. The adversary may also override the comparison module to suppress its built-in functionality. Galbally et al. [58] demonstrated that a higher similarity score on average consumes more time. Berthier et al. [68] analysed the potential side-channel leakages of a hardware fingerprint biometric comparison module. Their observations confirmed a shifting drop in power consumption for diverse minutiae angles.

The system designers should digitally sign a software-based comparison module. Masking and designing ICs with an active shield prevent confidential information leakage through side channels on a hardware-based comparison module [69]. Most of the higher-standard CC EAL cards, including NXP SmartMX family, have mitigation against side-channel attacks [70].

3.1.1.12 AP 12 (Attacks through communication channel connecting comparison module and decision module)

The attacker uses a vulnerable communication channel between the comparison and decision modules to replace the computed similarity score with an injected acceptable score to get authenticated. The system designers should implement a challenge-response strategy as a mutual authentication mechanism to detect a fraudulently injected similarity score into the communication channel. Some of the examples of such challenge response strategy include Challenge-Handshake Authentication Protocol (CHAP) and Open Authentication (OATH) Challenge-Response Algorithm (OCRA).

3.1.1.13 AP 13 (Attack on decision module)

The adversary targets the decision module to override its conclusion. The act initiates rejecting each submitted fingerprint leading to a DoS attack. In another instance, the system accepts the fingerprint of a non-enrolled individual. The mitigation approaches include detecting decision module code tampering through code signing and digital signature.

3.1.1.14 AP 14 (Zero-effort attack (false match))

A zero-effort attack due to the system limitation of a false match illustrates the benefit gained by an adversary, i.e., the adversary puts in zero efforts and gets authenticated due to false match and low system threshold. In general, a system whose threshold is set to a lower value is vulnerable to the attack. The existence of inter-class variations and limited individuality in biometric characteristics lead to a zero-effort attack [2].

He et al. [71] presented a robust fingerprint matching approach that uses multiple parameters to decide the similarity between two templates and alleviates a potential zero-effort

attack. Some other strategies that employ a fusion of local, global, and minutiae-based hybrid features to lower the risk of zero-effort attack includes a hybrid fingerprint matcher [72], an orientation-based minutia descriptor [73], minutiae and texture features-based fingerprint matcher [74], and an orientation-based minutia descriptor-based matcher [75].

3.1.1.15 AP 15 (Zero-effort attack (false non-match))

Another form of a zero-effort attack or Z-attack on the biometric system creates inconvenience to the enrolled users. In this scenario, the system rejects legitimate users due to a false non-match and forces them to resubmit their finger at the sensor repeatedly. In general, a system whose threshold is set to a higher value is vulnerable to such attack.

Poh et al. [76] demonstrated that a symmetric matcher that combines the similarity score and liveness score mitigates a zero-effort attack. The authors presented a classification scheme employing attack-specific characteristics to mitigate zero-effort and nonzero-effort impostor attacks such as spoofing attacks. The realisation of the scheme using the combination of a Gaussian Copula-based Bayesian classifier and linear classifiers outperformed the SVM-based baseline classifier. Zhang et al. [77] introduced the *similarity histogram approach* (SHA), as a global fingerprint alignment method that alleviates the possibility of a false match. Initially, the local similarity matrix is calculated using minutiae features and associated ridge patterns from the fingerprints under comparison. Subsequently, the approach utilises the similarity matrix to construct similarity histograms of transformation parameters. Finally, a statistical method was employed to obtain the optimal transformation parameters.

3.1.1.16 AP 16 (Attacks on biometric controlled application)

The biometric system controlled applications have the input device exposed to the external environment. The attacker damages the sensing device to mount a DoS attack. In another instance, the application server is bombarded with fake authentication requests such that the server becomes inaccessible for valid access requests [32]. The organisation should install robust sensing instruments with tampering alarms to detect a DoS attack by damaging the input device. A secure communication channel prevents fake data from illicitly entering the system.

3.1.2 Discussion

A comparative analysis of the proposed threat model and the existing threat models is shown in Table 3.1. Ratha et al. model [53] is a generic model that lacked details about specific attacks on the system components. Fishbone model [2] depicted the reasons that result in system failure. However, some threats, such as presentation attacks, may not lead to system failure but can allow an imposter to access the application. The model did not highlight such vulnerabilities and threats. Nagar et al. model [2] does not pinpoint side-channel attacks that can disclose secret information through insecure hardware and software components. Bartlow and Cukic framework [29] [26] presented several vulnerable components in the biometric application, but failed to illustrate the methodology for designing security techniques for the biometric system. The proposed model in this thesis addresses the limitations of previous models, presents an exhaustive security analysis depicting sixteen attack points, and provides various countermeasures to mitigate them. Therefore, the proposed model acts as a benchmark while designing a novel match-in-database biometric system.

The proposed model introduces the template protection techniques module, and demonstrates its threats. Even though the module is an integral component in modern authentication systems, it was not mentioned in the existing threat models. The side-channel attacks are used to extract secret information, such as encryption keys, from a vulnerable system component. However, previous models does not mention these threats. The proposed model shows the side-channel attack on the template protection techniques and the comparison module as depicted in Figure 3.1. It reports an instance of an adversary intercepting the encrypted template over the communication channel connecting the template protection techniques module and the template database during enrollment.

The enrollment stage does not require the matcher module as it creates the template and stores it in the database. However, existing models did not show the link connecting the template protection techniques module and the template database. The proposed model shows this connection as AP 7 in Figure 3.1. The model demonstrates two possibilities to mount a denial-of-service (DoS) attack that does not exist in previous models. The adversary can malfunction the sensing instrument, or the application controlled by the biometric system to mount a DoS attack. The DoS possibilities were lacking in existing models. The proposed

Table 3.1: Comparison of the proposed and existing threat models.

Threat model	**Founding criteria**	**Categorisation**	**Pros**	**Cons**
Ratha et al. model [53] (2001)	Access point of information	8 vulnerable points resulting in 8 types of attacks	Provides attacks on major components of biometric system	A general model without details of specific attacks e.g. administrative frauds
Bartlow and Cukic framework [29] [26] (2005)	Wayman's subsystem architecture	3 modules with 20 potential attack points	Specify details about the internal components of a generic biometric system and its vulnerabilities	Does not propose the methodology to design security techniques for biometric system
Fishbone model [2] (2008)	Cause and effect	5 causes leading to biometric system failure	Useful while designing biometric security techniques	Only shows reasons leading to biometric system failure
Nagar et al. model [2] (2008)	Cause and effect similar to Fishbone model	Major vulnerabilities and their 4 underlying causes	Addresses administrative loopholes along with other major attacks	Side-channel attack scenarios are missing
Proposed threat model (2020)	Existing and probable attacks	7 components with 16 probable attack points	Can be treated as a reference model while designing a match-in-database biometric system for any modality	The model can be extended for smart-card-based biometric systems

model showed the comparison module as a software or hardware component and pinpointed possible attacks in both scenarios. The earlier models considered software-based comparison modules and highlighted the threats to them. However, as hardware-based comparison modules are also employed in the biometric application, the proposed model showed the vulnerabilities of software and hardware-based comparison module.

The proposed and existing threat models provided theoretical and practical vulnerabilities associated with biometric systems. The mitigation approaches to secure the system from

various types of threats were also discussed. However, there is a necessity for a common platform wherein a newly built biometric system can be rigorously evaluated and tested against probable threats. El-Abed et al. [78] introduced *Security EvaBio* as an online tool employing a quantitative-based security assessment method to evaluate the security of a biometric system. The authors additionally provided a common vulnerabilities and threats database for biometric systems to assess a newly designed biometric system.

3.1.3 Threat analysis

The proposed threat model in Figure 3.1 depicted probable threats at multiple attack points, such as replay attacks at AP 3, AP 9, and AP 10. It is important to determine the intensity of a probable attack, and design countermeasures to provide maximum resilience against such attacks to the system. Hence, the attack points and potential threats are classified into eight classes to facilitate system designers in building a robust biometric system accepting any modality. The threat analysis of a match-in-database fingerprint biometric system introducing eight classes is shown in Table 3.2. The threats were analysed such that each class specifies similar attacks applicable to different components. The treats are categorised as high, low, and moderate, based on the intensity of damage and ease of execution. The components classified under high-risk factors may be most vulnerable with minimal efforts from the adversary, or a successful attack on such elements may create substantial damage to the system, such as system failure. As the biometric system input device and an application controlled by the biometric system are usually monitored by CCTV or security personnel, mounting a DoS attack through physically damaging these components is less probable, and hence is categorised as a low-risk factor. The proposed threat model depicted in Figure 3.1 shows sixteen attack points, vulnerable to diverse types of attacks. This work results into a unified representative model providing the threats, countermeasures, and existing literature to further explore the system's security aspects. The summary of the proposed threat model including the targeted components at each attack point, potential threats to these components, and suggested countermeasures are presented in Table 3.3. The table shows that the communication channel suffers from maximum attack cases. Moreover, administrative frauds via

AP 1 and DoS through AP 16 are less likely to be mitigated using technological solutions. Implementing feature extraction and template protection technique module into a single unit can reduce the vulnerable components in the system. As the template database is maintained at a different location than the biometric input device, securing the template database and the links connecting to it can alleviate real-world threats to the system. The table provides a complete representation of various threats and countermeasures at each vulnerable attack point. The table forms the foundation in exploring the biometric security domain.

Table 3.2: Threat analysis of a match-in-database fingerprint biometric system. AP stands for *Attack Point*.

Attack category	**Vulnerable components**	**Probable attacks**	**AP**	**Risk factor**
Class I	Biometric system input device, an application controlled by a biometric system	Denial-of-service attack	AP 2, AP 16	Low
Class II	Biometric input sensor	Spoofing attack, altered finger, fake fingerprint submission	AP 1	High
Class III	Communication channel connecting various components within the input device	Data interception, false data injection, replay attack, hill-climbing attack, brute-force attack	AP 3, AP 5, AP 9, AP 12	Moderate
Class IV	Communication channel connecting remote template database	Data interception, false data inject, replay attack	AP 7, AP 10	High

Class V	Feature extraction module, software-based template protection techniques or comparison module, decision module	Override specific module, Trojan horse attack	AP 4, AP 6, AP 11, AP 13	Moderate
Class VI	Hardware-based template protection techniques and comparison module	Side-channel attack, reuse-of-residual	AP 6, AP 11	Moderate
Class VII	Remote template database	Unauthorised access through an external compromised system, steal, delete, modify, substitute, reconstruct templates	AP 8	High
Class VIII	Other than aforementioned	Administrative frauds, zero-effort attacks	AP 1, AP 14, AP 15	High

Table 3.3: Biometric system attack summary with reference to Figure 3.1. AP stands for Attack Point.

AP	**Targeted component**	**Possible attack**	**Countermeasures**	**References**
1	Through the biometric sensor	Adversary attacks and administrative frauds	Liveness detection, anti-spoofing techniques, challenge/ response, separation of privileges	[79], [46], [80], [3], [81], [50]
2	On the biometric input device	Denial-of-service (DoS) attack	Rugged devices	[26], [2]

3	Through communication channel connecting biometric input device	Hill-climbing attack, brute-force attack, dictionary attack, replay attack, fingerprint image intercept	Time-out-lock-out policy, time-stamp, one-time password, digital signature, liveness detection, encryption	[2], [67]
4	Feature extraction module	Trojan horse attack, override feature extraction module	Code signing, a Trusted Biometric System, specialised tamper-resistant hardware	[2]
5	Through outgoing communication channel connecting feature extraction module	False data inject, reuse-of-residuals	Challenge-response based system, disposable Feature Extractors (FE)	[26], [82]
6	Template protection techniques module	Side-channel attacks	Masking, designing ICs with active shield	[83], [69]
7	Communication channel between template protection techniques module & template database	Template intercept, data inject	Time-out-lock-out policy, time-stamp, one-time password, digital signature, liveness detection, encryption	[84]

8	Template database	Unauthorised access through external compromised system, steal, delete, modify, substitute template	Server hardening, database access controls, digitally signing templates, template encryption, storing template on card	[2], [5], [85]
9	Communication channel between template protection module & comparison module	Alter system parameters, synthesised feature vector, hill-climbing attack, brute-force attack, fake digital biometric, replay attack	Time-out-lock-out policy, time-stamp, one-time password, digital signature, liveness detection, encryption	[86], [87]
10	Communication channel between template database and comparison module	Replay attack, intercept stored template	Utilise time-stamps or Time-to-live (TTL) tag	[26]
11	Comparison module	Override comparison module, side-channel attack, Trojan horse	Masking, designing ICs with active shield, code signing	[83], [61]
12	Communication channel between comparison module & decision module	Modify score	Mutual authentication between comparison & decision module	[2]

13	Decision module	Override decision	Code signing	[26]
14	Communication channel between decision module & biometric application	Zero-effort attack (false match)	Design robust comparison module	[87], [88]
15	Communication channel between decision module & biometric application	Zero-effort attack (false non-match)	Design robust comparison module	[88]
16	On biometric controlled application (e.g. cash dispenser)	DoS attack	CCTV monitoring, deploy security guards	[2], [26]

3.2 Latent MasterPrint: a case study

The finger imprints fortuitously left behind at a crime location are referred to as *latent fingerprints*. Forensics agencies exercise latent fingerprints to investigate a probable suspect. Latent fingerprints are usually partial, and lack clarity in ridge patterns [89]. Hence, an *automated fingerprint identification system* (AFIS) is employed to acquire, process and analyse these blurry and distorted fingerprints. An overview of different steps involved in a latent fingerprint identification system is shown in Figure 3.5. Forensic investigators visit the crime scene to locate latent fingerprints on various surfaces. Sweat formed on the finger, and oil secreted from the pores result in a latent fingerprint. The experts utilise different powders according to the surface material to enhance such prints [90]. Due to distorted and low feature clarity, they are often preprocessed to improve the quality. Yoon et al. [91] suggested a metric *latent fingerprint image quality* (LFIQ) to distinguish the quality of latent fingerprints. The parameters related to ridges and minutiae are analysed to determine the LFIQ.

The latent fingerprints lifted from a crime scene are segmented as foreground and background for identification of ridge patterns. The segmentation process includes normaliza-

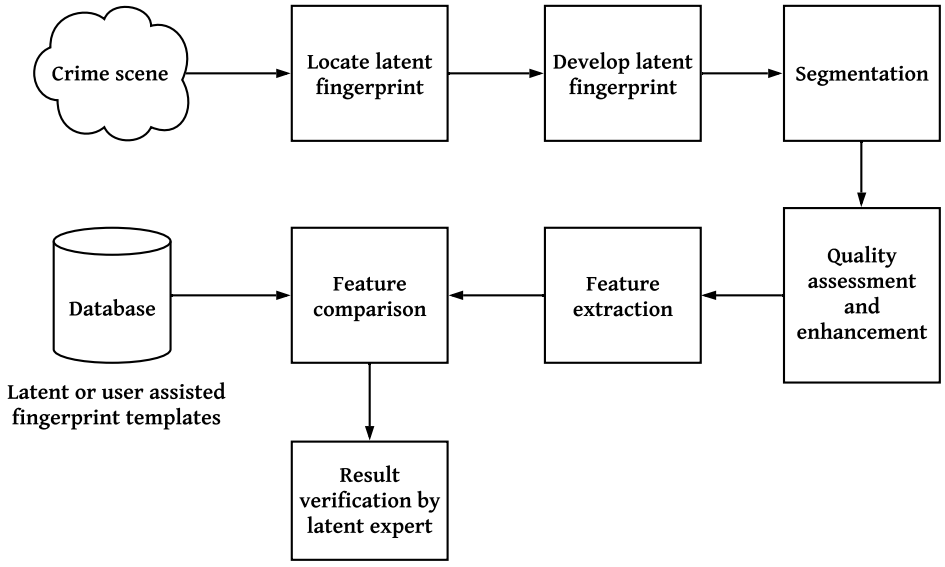


Figure 3.5: Overview of various steps involved in a latent fingerprint identification system.

tion, orientation and ridge frequency estimation to rebuild discontinuous ridges and curve patterns [92]. Moreover, methods such as Canny Edge Detection, Laplacian, and Gaussian Low Pass filters are used to enhance the latent fingerprint quality. Fingerprint ridges are generally categorised using 17 characteristic features [90]. Among them, *level* 2 features include the minutiae. Latent fingerprints usually show up with a significantly low count of minutiae compared to rolled or plain fingerprints due to their poor quality and small area. The average number of minutiae observed in NIST Special Database 27 (NIST SD27) [93] images is 21 and 106 for latent fingerprints, and corresponding rolled prints, respectively [94]. Due to low minutiae count observed in latent fingerprints, most techniques for latent fingerprint identification use correlation-based or non-minutiae feature-based matching [95]. However, Krish et al. [96] introduced an approach based on extended minutiae types. The results showed improved accuracy when integrated with the existing minutiae-based methods.

As the judiciary accepts forensic reports on fingerprint matching as supportive evidence, the AFIS must produce accurate results to ensure that the system conclusion does not erroneously strengthen fabricated evidence. However, as latent fingerprints are partial and possess unclear ridge patterns, it is essential to investigate the possibility of a latent fingerprint acting like a MasterPrint. Latent impressions lifted from a crime scene are usually compared against a fingerprint database comprising paper-based inked fingerprints or fin-

gerprints captured conventionally using AFIS from criminals and suspects. The term *Latent MasterPrint* is referred to a latent fingerprint that identifies at least 4% distinct subjects from such a database storing fingerprints of past offenders. The investigation on the prospects of a Latent MasterPrint is presented in the following subsections.

3.2.1 Experimental setup

The experiments were conducted under a closed-set identification setup assuming that every input has a matching fingerprint in the database. In the *identification test*, each input latent fingerprint is compared with every other stored template from the dataset, and the similarity score for each comparison is computed [97]. A *correct detect and identify* (CDI) represents that the highest score corresponds to a sample from the actual subject. However, a *false alarm* (FA) shows that the highest score belongs to the sample from some other subject. The rejected latent fingerprints showed a similarity score less than the system threshold. Let $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{R}$ denote the CDI count, FA count, and the rejected latent fingerprint count. Assuming that the system is enrolled with $\mathcal{K}$ latent fingerprints, such that $\mathcal{K} = \mathcal{C}+\mathcal{F}+\mathcal{R}$, the *detect and identification rate* (DIR), *false alarm rate* (FAR), and *rejection rate* (RR) were computed as: $DIR = \frac{\mathcal{C}}{\mathcal{K}} \times 100$, $FAR = \frac{\mathcal{F}}{\mathcal{K}} \times 100$, and $RR = \frac{\mathcal{R}}{\mathcal{K}} \times 100$.

In addition to DIR, FAR, and RR, the percentage of Latent MasterPrint generated and the maximum number of subjects identified by a Latent MasterPrint were also recorded. The identification tests were conducted for threshold value 0, 5, 15, 25, 35 and 45. A *cumulative matching characteristic* (CMC) curve shows the DIR for each threshold at different ranks [98]. The identification test results were utilised to plot the CMC curves. Ideally, if n subjects were enrolled with a system, the rank-n identification rate should be 100%. Hence, the best approach reaches 100% DIR at the earliest, i.e., at lower ranks. The experiments were conducted using the *Multisensor Optical, and Latent Fingerprint* (MOLF) DB4 dataset comprising 4400 latent fingerprints cropped from simultaneous prints from 100 subjects [9]. The volunteers deposited their simultaneous latent fingerprints on a ceramic tile during the dataset preparation. Fingerprints were then directly captured using the camera apparatus. The camera setup consisted of a USB-programmable UEye camera with a capture size of

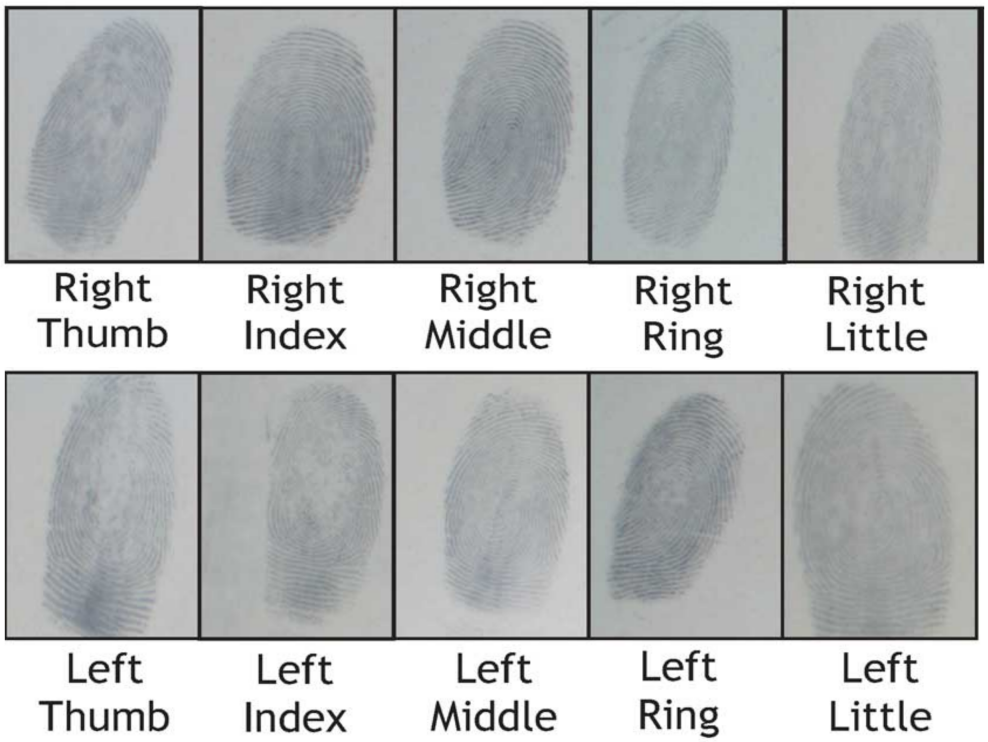


Figure 3.6: Sample latent fingerprint images from MOLF DB4 dataset.

3840×2748 pixels. It has a 0.5-inch CMOS sensor and captures at a maximum rate of three frames per second. The latent fingerprint images from the MOLF DB4 dataset were employed during the experiment without applying any pre-processing, i.e., enhancement techniques. Figure 3.6 illustrates sample latent fingerprint images from the MOLF DB4 dataset. It was important to employ a combination of minutiae detection and minutiae comparing algorithms that are tuned and blended to deliver high matching accuracy. The NIST Biometric Image Software (NBIS) is a standardised software that includes the MINDTCT algorithm to detect minutiae and the BOZORTH3 algorithm to perform minutiae matching [99]. Hence, these algorithms were utilised during the experiments.

3.2.2 Result analysis

The results for the identification test at various thresholds on the MOLF DB4 dataset are shown in Table 3.4. The results show that $64\% - 84\%$ Latent MasterPrints were generated up to a threshold of 15. A Latent MasterPrint at each threshold up to 15 identified every subject from the dataset. The DIR and FAR under these scenarios were observed between

Table 3.4: Results on MOLF DB4 dataset for various thresholds. DIR = Detect and Identification Rate, FAR = False Alarm Rate, RR = Rejection Rate, MPs = percentage of Latent MasterPrints generated, LMPs = Maximum no. of subjects identified by a Latent MasterPrint at a given threshold.

Threshold	DIR	FAR	RR	MPs	LMPs
0	38.2%	48.94%	12.86%	87.14%	100
5	38%	49.02%	12.98%	85.89%	100
15	36.4%	41.96%	21.64%	64.23%	100
25	31.4%	27.3%	41.3%	7.86%	91
35	26.8%	5.09%	68.11%	0.16%	47
45	21.4%	0.6%	78%	0.04%	13

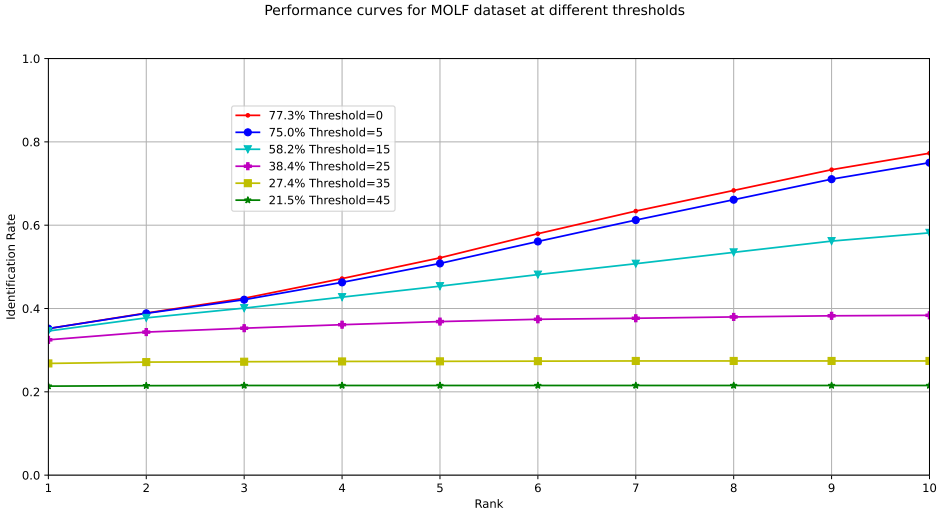


Figure 3.7: CMC plot: Rank-10 performance on MOLF DB4 dataset at various thresholds

$36 - 38\%$ and $40 - 50\%$, respectively. The DIR decreased moderately as the threshold was raised. However, the percentage of Latent MasterPrints generated and the maximum number of subjects identified by a Latent MasterPrint reduced significantly while increasing the threshold. At a threshold of 15, no Latent MasterPrint has been observed, and the DIR reduced to a mere 18.1% while rejecting the remaining latent fingerprints. The dataset possesses more than 96% latent fingerprints having an NBIS Fingerprint Image Quality (NFIQ) score of 5, i.e., the worst quality images [100]. Due to the poor quality of a large portion of the latent fingerprints in the dataset, NBIS showed high error rates in terms of FAR and RR.

The CMC curves at various thresholds for the MOLF DB4 dataset is shown in Figure 3.7. The rank-10 performance at the lower thresholds was above 75%, wherein about 13% of latent fingerprints were rejected by the NBIS. With an increasing threshold, the DIR was

reduced below 22%. Moreover, there was no significant improvement in the DIR at higher ranks above the threshold of 25. The results imply that at higher thresholds, the rejection rate crosses 75% as we achieve zero Latent MasterPrint generation. The investigation thus demonstrated the possibility of Latent MasterPrint. It also confirmed that future research on Latent fingerprint identification must include the latent MasterPrint as a metric to ascertain the robustness of a novel latent fingerprint identification approach. This work thus proved the significance of designing resilient methods to mitigate Latent MasterPrint vulnerability.

3.3 Summary of the chapter

This chapter introduced the proposed threat model depicting sixteen attack points as potential threats to a match-in database fingerprint biometric system. The model was comprehensively examined by pinpointing various threats and respective countermeasures. The chapter included a comparative analysis of the proposed and existing threat models and the organisation of system threats into eight classes. Further, as latent fingerprints are partial and possess distorted ridge structure, an investigation was carried out to explore the prospect of Latent MasterPrint. The results using the MOLF dataset and NBIS generating 87.14% Latent MasterPrints without any threshold and 0.04% Latent MasterPrints at a threshold of 45 confirmed that Latent MasterPrints exist. This work motivates to carry out future research in developing a novel and accurate latent fingerprint identification technique that mitigates the Latent MasterPrint.

Chapter 4

MasterPrint Mitigation Using Minutiae-based Coordinate System

Roy et al. [7] deduced that enrolling the same finger multiple times and accepting an input partial fingerprint that shows a similarity score above the threshold for any of the enrolled templates facilitates the MasterPrint generation. Due to environmental factors and user ignorance, multiple samples of the same finger often vary. Hence, image preprocessing using binarization and thinning are employed to produce nearly similar fingerprints. However, precisely similar minutiae positions over these fingerprints can not be guaranteed. Therefore, minutiae-based fingerprint biometric systems usually accept approximately comparable features within a tolerance range, δ, between the two fingerprints. In addition, these systems eventually result in a high probability of MasterPrints generation. Hence, the authors suggested a better partial fingerprint matcher to mitigate the MasterPrint vulnerability. In this regard, this chapter addresses the MasterPrint vulnerability by proposing a novel partial fingerprint identification method that extracts minutiae-oriented local features from binarized and thinned partial fingerprint images over eight axes emerging from a reference minutia. The chapter introduces a metric to compute the similarity score between the input and enrolled templates. The features are compared without considering any tolerance to reduce the chances of MasterPrint.

The main highlights of the work in this chapter are as follows,

1. Presenting a comparative analysis of existing partial fingerprint recognition systems.

2. Building a novel feature vector that represents a minutia uniquely.

3. Introducing a similarity score computation metric for minutiae-based fingerprint biometric system.

4. Experimentations on five benchmark fingerprint datasets, namely, the standard FVC2002 DB1_A [101], FVC2002 DB2_A [101], CrossMatch Sample DB dataset from NeuroTechnology [102], NIST Special Databases 302 (sd302b and sd302d) [103], to show the robustness of the approach towards reducing MasterPrint generation.

The contents of this chapter are organised as follows. Initially, Section 4.1 provides a glimpse of various approaches employed in partial fingerprint recognition. Section 4.2 is dedicated to the description of various steps involved in the proposed partial fingerprint identification method. The experimental setup and experimental results are presented in Section 4.3 and Section 4.4, respectively. The discussion and the significance of the experimental results are shown in Section 4.5. Finally, Section 4.6 summarises the chapter.

4.1 Partial fingerprint recognition: a glimpse

A partial fingerprint recognition system performs matching in three ways. It can be compared with another partial fingerprint template, a full fingerprint template, or a template created by employing image fusion techniques over several partial fingerprints [104]. Usually, small sensors from smartphones capture about $3 - 5$ minutiae on average [105]. Hence, smartphones such as Samsung Galaxy versions and iPhones employ texture-based approaches for partial fingerprint matching to ensure instantaneous and highly accurate results. Kumar et al. [106] proposed a level-zero feature-based method for contact-less fingerprint identification, which has potential application in partial fingerprint identification on smartphones. The fusion of several images involves an alignment of influential information from diverse sources using mathematical modelling practices for generating a single complete image [107]. The existing approaches performing partial-to-partial template comparison includes support vector machine-based method [108], minutiae-triplet-based approach [109], finger-

print alignment-based method [110], etc. Partial-to-full fingerprint template comparison was performed in the convex hull-based method [111], checkerboard sampling method [112], and minutiae-oriented localised secondary features-based approach [113], etc.

Jea et al. [111] employed minutiae and its secondary features to create convex-hull for local feature extraction. Zeng et al. [114] proposed *residual network* (ResNet) to extract features from a partial fingerprint. Mil'shtein et al. [115] employed *line scan algorithm* (LSA) over 2D fingerprint images for partial fingerprint matching. Zhang et al. [116] proposed a partial fingerprint recognition method utilising low-level minutia and high-level global features. Bae et al. [117] used minutiae and image segmentation during feature extraction process. The feature comparison was performed using feature-weighted block scoring, segment-based rotation matching, and segmented matching. Aravindan et al. [118] utilised the *scale-invariant feature transform* (SIFT) algorithm and wavelet transformed fingerprint image to recognise a partial fingerprint. The experiments comprised partial-to-partial and partial-to-full fingerprint matching. Ramírez et al. [119] introduced the computational geometry-based fingerprint verification method. The approach uses global features from the Delaunay Triangulation formed using minutiae over the entire fingerprint. A comparative analysis of various partial fingerprint recognition approaches from the literature is presented in Table 4.1.

Table 4.1: Partial fingerprint recognition : summary of existing approaches.

References	**Base for fingerprint recognition**	**Approach/ Method employed**	**Comparison type**	**Dataset used**
Jea et al. [113] (2004)	Localised secondary features are generated from minutiae	Brute-force, and secondary feature matching	Partial-to-full fingerprint	FVC2002 DB1 database

Jea et al. [111] (2005)	Minutiae and secondary features generated from minutiae	Convex hull-based method, MCF technique	Partial-to-full fingerprint	FVC 2002 DB1 and DB2 databases
Chen et al. [120] (2007)	Level 3 features namely, dots and incipients	A commercial minutiae-based matcher Neurotechnology VeriFinger 4.2	Partial to full fingerprint	NIST SD30
Gang Fang et al. [121] (2007)	Minutiae and representative ridge points	BOZORTH and the compound minutiae (or k-minutiae) matcher	Partial-to-partial and partial-to-full fingerprint	FVC2002 DB1 and DB3 datasets
Mil'shtein et al. [115] (2008)	Converts a 3D object like a finger into a 2D image with minimal distortion	Spaced frequency transformation algorithm (SFTA) and line scan algorithm (LSA)	Partial-to-partial and partial-to-full	Database of 150 partial fingerprints
Vijayaprasad et al. [108] (2010)	Minutiae	Global minutiae matching and *support vector machine* (SVM)	Partial-to-partial fingerprint	FVC 2002 database

Arada et al. [112] (2012)	Checkerboard sampling method	Artificial neural network (ANN)	Partial-to-full fingerprint	FVC2004 DB1_A
Chang et al. [122] (2012)	Minutiae	Triangular matching scheme	Partial-to-partial fingerprint	FVC database
Gao [123] (2012)	Constructing subtemplates from minutiae-sparse multiple partial fingerprints	Subtemplate matching schemes	Subtemplates-to-supertemplates matching	CASIA DBv5
Chen et al. [124] (2013)	Fusion of the minutiae matching score and the fingerprint area	Modified SVM	Fusion of partial-to-partial fingerprint	FVC2002 DB1 and DB2, and the THU database
Zanganeh et al. [125] (2014)	Comparing pixel-wise correlation coefficient	Fisher transform	Correlation of regions created by dividing full-fingerprint images	FVC2002 DB1
Zanganeh et al. [110] (2015)	Partial fingerprint alignment	Fisher transformation technique	Partial-to-partial fingerprint	FVC 2002 DB1 dataset
Zhou et al. [126] (2016)	Local minutiae triplet features	Fuzzy-based fusion scheme	Partial-to-partial fingerprint	FVC 2000 DB2a, FVC 2002 DB1a, and NIST SD 14

Ramírez Flores et al. [119] (2016)	Minutiae-based connected graph composed of triangles	Delaunay triangulation	Full image matching	FVC2002 databases
Lee et al. [104] (2017)	Minutiae and ridge shape features	Minutiae matching and ridge-feature-matching	Partial-to-partial fingerprint	FVC2002, FVC2004 and BERC (self-constructed) databases
Aravindan et al. [118] (2017)	Wavelet transformed fingerprint image	SIFT algorithm	Partial-to-full and partial-to-partial fingerprint	MESCEF, FVC2002 and CASIA database
Qin et al. [127] (2017)	Partial fingerprint alignment	Phase-only correlation and deep *convolutional neural network* (CNN)	Partial-to-partial fingerprint	FVC2000, FVC2002, FVC2004, FingerPass DB7, in-house partial fingerprint database
Bae et al. [117] (2018)	Minutiae, image segmentation	Segment-based rotation matching, segmented matching, feature-weighted block scoring	Full image matching, Segmented area matching	Database acquired by the capacitive fingerprint sensor for the latest mobile device

Boujnah et al. [128] (2018)	Redefined characteristics of minutiae	Euclidean distance between feature vectors from different partial fingerprints	Partial-to-partial fingerprint	POLYU HRF database and FVC 2004 database
Khodadoust et al. [109] (2018)	Minutiae triplets, and orientation field (OF)	Indexing algorithms	Partial-to-partial fingerprint	FVC 2000, 2002, 2004 and NIST SD4, SD14
Zeng et al. [114] (2019)	Residual network (ResNet)	Deep learning	Partial-to-partial fingerprint	NIST-DB4 and self-built (NCUT-FR) dataset
Zhang et al. [116] (2019)	High-level global feature and low-level minutia feature	Combined global matching and minutiae-based matching	Partial-to-partial fingerprint	FVC 2000, 2002, 2004, 2006, and in-house dataset
Deshpande et al. [129] (2020)	Local minutiae arrangements (MA)	Delaunay triangulation	CNNAI, a convolutional neural network (CNN) matching model	FVC2004 and NIST SD27 fingerprint datasets

4.2 Proposed approach

The existing approaches for partial fingerprint recognition discussed in the previous section were mostly experimented under verification scenario. However, MasterPrint vulnerability is associated with identification systems and none of these methods was evaluated for MasterPrint generation. The flowchart for partial fingerprint-based MasterPrint identification using the proposed method is presented in Figure 4.1. Each partial fingerprint passes through seven stages during the enrollment stage. Initially, the region of interest is detected from the input partial fingerprint. The gray-scale images for the same finger on multiple instances produced by the sensing devices often vary. Hence, an image binarization process is applied to render several fingerprints of an individual appear similar. An adaptive thresholding approach carries out the binarization process. Minutiae-based feature extraction is more accurate in a thinned image. Therefore, a binarized image is thinned to extract local features near a minutia. Zhang-Suen thinning algorithm [130] is applied for the thinning operation. A sample partial gray-scale fingerprint image and its binarized and thinned version is shown in Figure 4.2.

The conventional metric, *Crossing Number* (CN) [131], detects minutiae in the thinned image. As full fingerprints are cropped to create the partial dataset for the experiments, the pixels near the image boundary emerge as *false* minutiae, i.e. as a ridge ending. Moreover, incorrect binarization and thinning operations introduce additional *false* minutiae. Hence, it is essential to eliminate such spurious minutiae for an improved identification rate. Kim et al. [132] algorithm is used to detect and remove false minutiae. The algorithm employs ridge flow, connectivity, ridge orientation, and distance between minutiae as the underlying parameters for post-processing the detected minutiae. It detects and removes five types of false minutiae, namely, short ridge, hole, triangle, broken ridge, and bridge [132].

The orientation for the true minutiae detected in the recent step is found using the input partial fingerprint image. The feature extraction process involves creating an eight-axes coordinate system in the binarized and thinned image around each true minutia. The ridge-to-valley and valley-to-ridge transitions are denoted as $r \rightleftharpoons v$. The count of $r \rightleftharpoons v$ transitions over an axis in a binarized image is termed as *intensity-shift*, IS. The valley-to-valley tran-

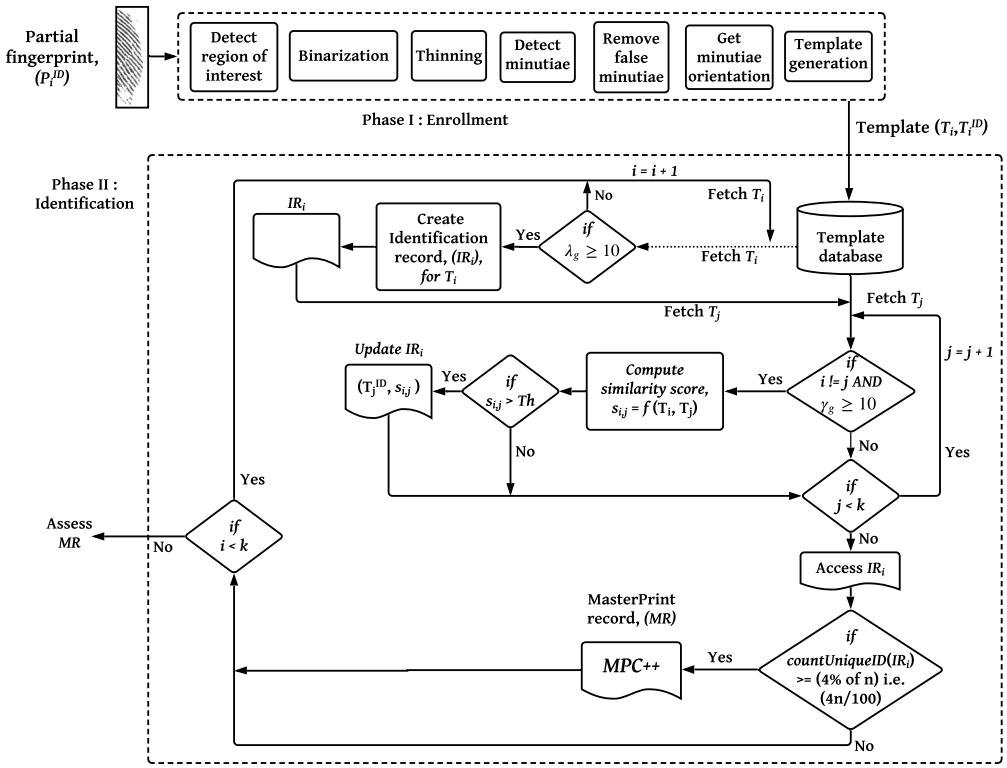


Figure 4.1: Flowchart for partial fingerprint-based MasterPrint identification. The enrollment phase takes a partial fingerprint, P_i^{ID} as input, and enrols the template T_i and corresponding user identifier, T_i^{ID}. The database contains k templates for n users such that $i \in \{1, 2, \ldots, k\}$, $ID \in \{1, 2, \ldots, n\}$, and $k > n$. The dotted arrow originating from the template database denotes the start of the identification phase. The system threshold is set at Th, and the MasterPrint count (MPC) is initialised to zero. Initially, T_i and T_j corresponds to the first template in the database, λ_g and γ_g represents the number of good quality minutiae in T_i and T_j, respectively, and $s_{i,j}$ represents the similarity score between T_i and T_j.

(a) Gray-scale image (b) Binarized image (c) Thinned image

Figure 4.2: Sample gray-scale partial fingerprint, its binarized and thinned version. The white portion indicates the sensor surface left untouched by the finger.

sitions are termed as $v \rightarrow v$. The count of $v \rightarrow v$ transitions over an axis in a thinned image is termed as *intensity-factor*, IF. A successful feature extraction labels a given minutia as *good quality minutiae*. A feature template, T, comprises the feature vectors correspond-

ing to the true minutiae in a partial fingerprint. If the i^{th} partial fingerprint to be enrolled with the system belongs to a user with the identifier as ID, it is represented as P_i^{ID}, and its feature template is denoted as T_i. The flowchart in Figure 4.1 assumes that k partial fingerprints belonging to n subjects are enrolled into the template database. The database contains multiple feature templates generated from different samples of each subject. Hence, in the experimental setup k is greater than n, i.e., $k > n$.

An all-versus-all comparison of the enrolled feature templates is performed during the identification phase. The feature templates possessing less than ten good quality minutiae are not considered for comparison. A template, T_i, undergoes comparison with every other template, T_j, to compute a similarity score, $s_{i,j}$ between them. An identification record, IR_i, corresponding to each T_i is maintained to store the user identifier, T_j^{ID}, for T_j and the score $s_{i,j}$. If $s_{i,j}$ is greater than the system threshold, Th, such entry is recorded in IR_i. Furthermore, IR_i is assessed to verify if T_i identified at least 4% distinct subjects. If so, the MasterPrint count, MPC, in the *MasterPrint record*, MR, is incremented to mark it as a MasterPrint. The subsequent subsections explain the feature extraction, template generation processes, and similarity score computation method.

4.2.1 Feature extraction from binarized fingerprint

The feature template comprises local features extracted in eight directions around a reference minutia to ensure a distinctive feature vector for every minutia. Hence, a local region is created using an eight-axes coordinate system around a reference minutia to facilitate local feature extraction. The proposed eight-axes coordinate system for a true minutia, m, coloured red at the centre in a binarized image is shown in Figure 4.3. The figure showed a red coloured reference minutia, m, at the center and the proposed intensity matrix over each axis surrounding the reference pixel, C, and its adjacent pixels, $C_{n_i} \in \{0, 1\}, i \in \{1, 2, \ldots, 8\}$, along axis, ax_i, $i \in \{1, 2, \ldots, 8\}$, respectively. Each axis around a reference minutia in a binarized image is traversed to measure the intensity-shift, IS_i, $i \in \{1, 2, \ldots, 8\}$, over axis, ax_i, $i \in \{1, 2, \ldots, 8\}$, respectively. The intensity-shifts are initialized as, $IS_i = 0$, $i \in \{1, 2, \ldots, 8\}$.

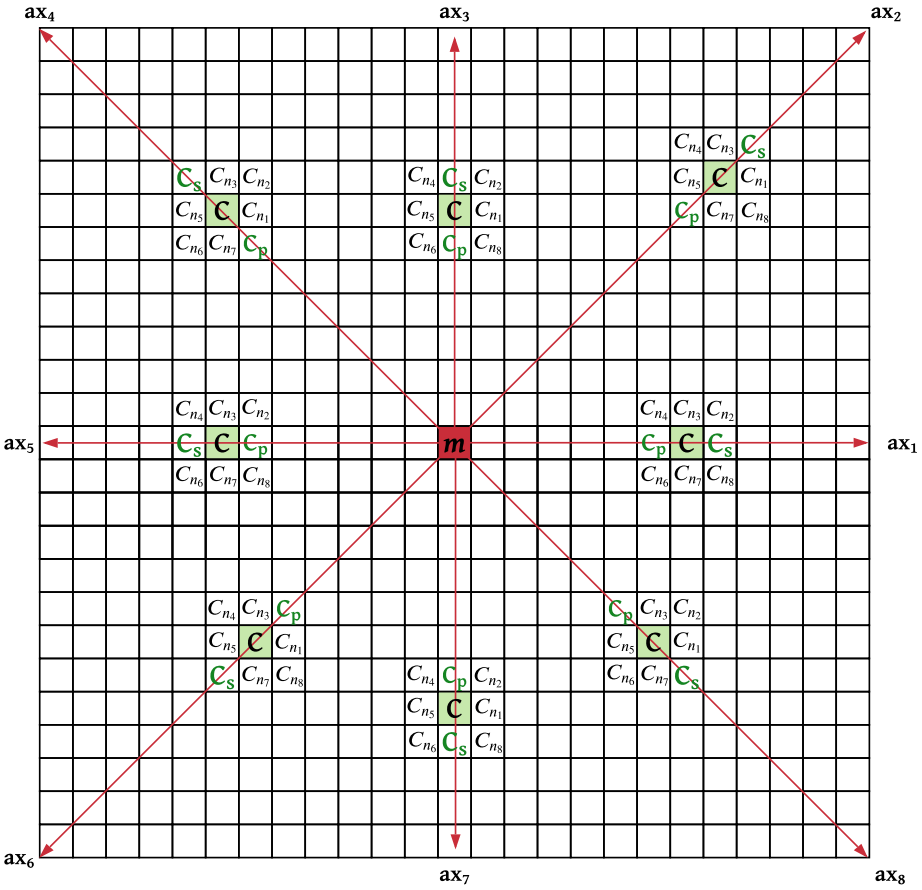
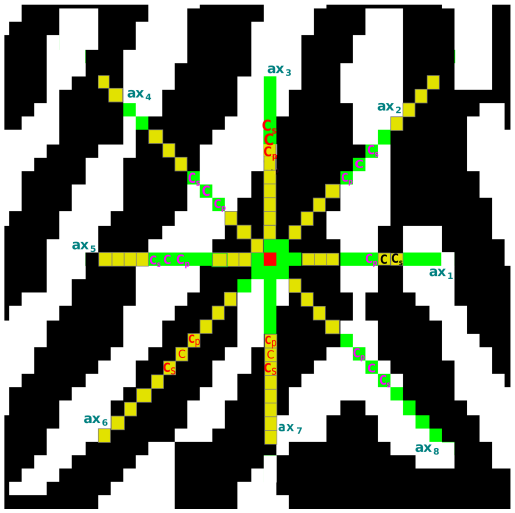


Figure 4.3: The eight-axes coordinate system corresponding to a reference minutia, m, marked in red at the center. Each square represents a pixel in the binarized fingerprint image. The intensity matrix over each axis surrounding the reference pixel, C, employed during feature extraction depicts $C_{n_i}, i \in \{1, 2, \ldots, 8\}$ as the current neighbor along axis $ax_i, i \in \{1, 2, \ldots, 8\}$, respectively. Further, C_p and C_s shows the *preceding* and *succeeding* pixels while traversing along a given axis starting from m.

While traversing away from the reference minutia along any axis and reaching a reference pixel, C, the recently observed pixel position is marked as C_p. In contrast, the pixel at the following location in the same path is marked as C_s. The IS value for a reference minutia on each axis is incremented for every occurrence of $r \rightleftharpoons v$ transition. It was observed that such instances occur whenever the RHS of equation (4.1) evaluates to $\sigma = 0$ and $\omega = 1$.

$$(\sigma, \omega) = (C_p \oplus C) + (C_p \oplus C_s) \tag{4.1}$$



Axis	C_p	C	C_s	(σ, ω)
ax_1	1	0	0	(0,1)
ax_2	1	1	1	(0,0)
ax_3	0	1	1	(0,1)
ax_4	1	1	1	(0,0)
ax_5	1	1	1	(0,0)
ax_6	0	0	0	(0,0)
ax_7	0	0	0	(0,0)
ax_8	1	1	1	(0,0)

Figure 4.4: Sample binarized fingerprint image to illustrate the feature extraction. The minutia is red, the eight-axes coordinate system around the minutia are green and yellow. The yellow pixels show the axis passing over a ridge whereas the green pixels show the axis passing over a valley. The black squares are the binarized ridges. The table on the right side show the computation according to equation 4.1 for a reference pixel, C, on the respective axis. As only the green coloured values for σ and ω in the table satisfy the condition demonstrating $r \rightleftharpoons v$ transition, the value of IS_1 and IS_3 will be incremented.

where, C, C_p, $C_s \in \{0, 1\}$, $\oplus$ indicates a XOR operation, and $+$ denotes 1-bit binary addition returning sum, σ, and carry, ω. A sample binarized image illustrating the instances of incrementing the IS is shown in Figure 4.4. The red coloured minutia is centered at the green and yellow coloured eight-axes coordinate system. The part of axis crossing a binarized ridge, i.e., black region, is coloured yellow. The part of axis crossing a valley region is coloured green. The yellow, black, and red pixels have an intensity bit value of 0. The green and white pixels possess an intensity bit value of 1. The table on the right part of Figure 4.4 illustrates the computation using equation 4.1 for the reference pixel, C, over the respective axis. The values for σ and ω coloured green satisfies the condition demonstrating $r \rightleftharpoons v$ transition. However, the red coloured entries from the table does not satisfy the condition indicating no possibility of $r \rightleftharpoons v$ transition. Hence, the values of IS for these axes will not be updated. Hence, the value for IS_1 and IS_3 will be incremented.

4.2.2 Feature extraction from thinned fingerprint

The thinned fingerprint image contains single pixel ridges, as shown in Figure 4.5. Each axis around a reference minutia in the thinned image is traversed away from the minutia to measure the intensity-factor, IF_i, $i \in \{1, 2, \ldots, 8\}$, along axis, ax_i, $i \in \{1, 2, \ldots, 8\}$, respectively. The intensity-factors are initialised as, $IF_i = 0$, $i \in \{1, 2, \ldots, 8\}$. The value of IF over each axis in a thinned image is incremented for every instance of $v \rightarrow v$ transition. The observations showed that these transitions occur whenever the RHS of equation (4.2), (4.3), (4.4), (4.5) and (4.6) evaluates to $\sigma = 0$ and $\omega = 1$.

$$(\sigma, \omega) = (C_p \oplus C) + (C \oplus C_s) \tag{4.2}$$

$$(\sigma, \omega) = (C_p \oplus C_{n_7}) + (C_{n_5} \oplus C) \tag{4.3}$$

$$(\sigma, \omega) = (C_p \oplus C_{n_1}) + (C_{n_7} \oplus C) \tag{4.4}$$

$$(\sigma, \omega) = (C_p \oplus C_{n_1}) + (C_{n_3} \oplus C) \tag{4.5}$$

$$(\sigma, \omega) = (C_p \oplus C_{n_5}) + (C_{n_3} \oplus C) \tag{4.6}$$

where, C, C_p, C_s, C_{n_1}, C_{n_3}, C_{n_5}, $C_{n_7} \in \{0, 1\}$, $\oplus$ denotes a XOR operation and $+$ indicates 1-bit binary addition returning sum, σ, and carry, ω. The equation 4.2 holds for all eight-axes. Further, the intensity-factors, IF_2, IF_4, IF_6, and IF_8 over the diagonal axes, ax_2, ax_4, ax_6, and ax_8, are also incremented using equation (4.3), (4.4), (4.5), and (4.6), respectively.

A sample intensity matrix over a single-pixel ridge in a thinned image illustrating the feature extraction, IF, is shown in Figure 4.5. A square denotes a single pixel. The minutia and the eight-axes coordinate system around the minutia are coloured red. A thinned ridge having an intensity bit value of 0 is denoted with green colour. The valley region having an intensity bit value of 1 is shown as white squares. The blue circled numbers indicate the equations whose condition on σ and ω evaluates to true, illustrating $v \rightarrow v$ transition. Hence, the value of IF over the respective axis will be incremented.

In a certain instance, we may reach the image boundary before finishing the l^{th} iteration

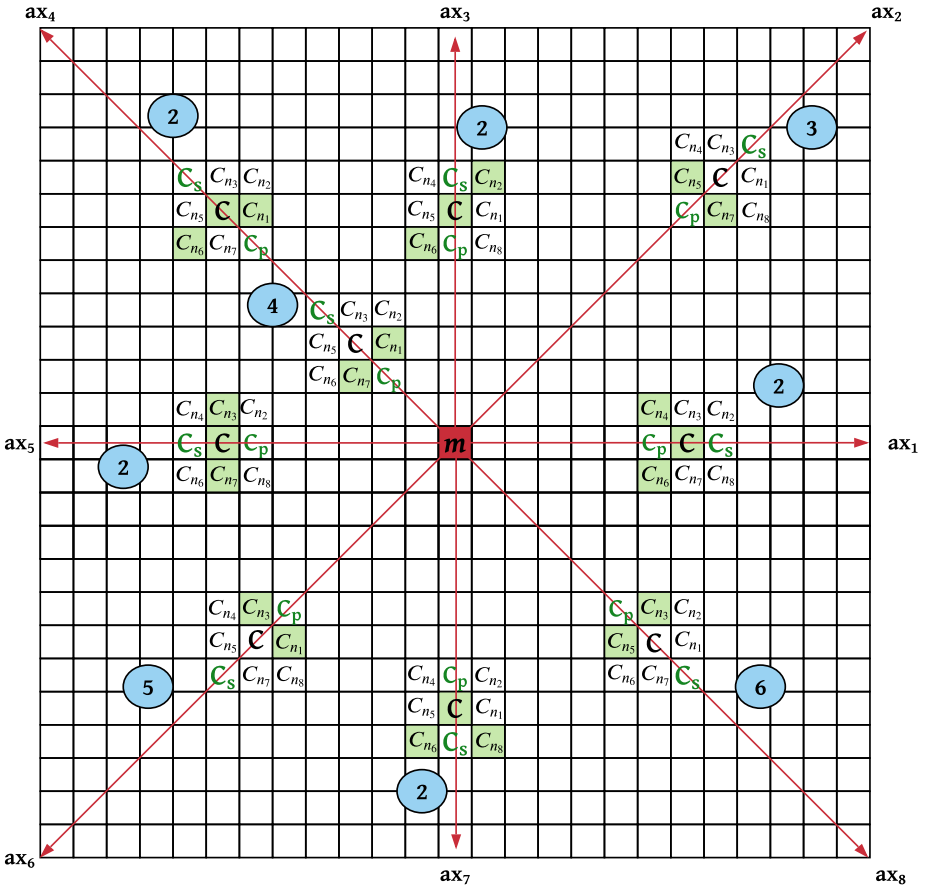


Figure 4.5: Sample intensity matrix for a thinned, i.e., possessing single-pixel ridge, fingerprint image to illustrate the feature extraction, IF. The minutia is red, the eight-axes coordinate system around the minutia are coloured red. The green squares are the thinned ridges. The algorithm will increment the IF on each axis as at least one equation from $4.2 - 4.6$ satisfies the condition on σ and ω. The numbers in blue circles show the equations for which the condition on σ and ω becomes true, demonstrating $v \rightarrow v$ transition.

while traversing a given axis in the binarized and thinned image. The IS and IF value over such axis is marked as -1, indicating incomplete feature extraction. The value of l, which produced feature vectors with a minimum value of -1 for IS and IF in a feature template, was decided empirically. The observations from Peralta et al. [133] were used to set the value of $l = 13$ indicating the limit for traversing an axis to measure $IS_i, i \in \{1, 2, \ldots, 8\}$ and $IF_i, i \in \{1, 2, \ldots, 8\}$.

4.2.3 Template generation

A feature vector representing the IS and IF is created for each true minutia. The feature template comprises feature vectors for every true minutia in a partial fingerprint. A true minutia should be positioned at least 13 pixels within the partial fingerprint image to extract non-negative values for IS and IF. A true minutia near the image boundary does not facilitate complete feature extraction in eight directions. Hence, such minutiae are termed as *bad quality*. A minutia quality, φ, is computed using equation (4.7).

$$\varphi = \begin{cases} 1\,, & \text{if } \left(\sum_{i=1}^{8} |\, IS_i| - \sum_{i=1}^{8} IS_i \right) = 0 \\ 0\,, & \text{otherwise} \end{cases} \tag{4.7}$$

The value $\varphi = 1$ indicates a *good quality* minutia, i.e., having positive values over each axis in binarized and thinned image. A partial fingerprint template possessing less than ten good quality minutiae is discarded from comparison in the identification phase. The feature vector, $\mathcal{V}$, for a true minutia in the partial fingerprint is represented as,

$$\mathcal{V} = (\varphi, b_1, b_2, \ldots, b_8, t_1, t_2, \ldots, t_8)$$

where, $\varphi \in \{0, 1\}$ indicates the minutia quality. The elements $b_i, i \in \{1, 2, \ldots, 8\}$ and $t_i, i \in \{1, 2, \ldots, 8\}$ corresponds to the IS and IF along the eight-axes in the *binarized* and *thinned* image traversed in anti-clockwise direction, respectively. The feature template, T, is denoted as, $T = (\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_{\lambda_t})'$, where λ_t is the number of true minutiae in the partial fingerprint and $'$ indicates the transpose. The dimension of feature template, T, is $\lambda_t \times 17$.

The minutia orientation, θ, is used to decide the axis in a binarized and thinned image whose IS and IF measures occupy b_1 and t_1 in the feature vector, respectively. The subsequent axes in anti-clockwise direction are followed to replace b_2, b_3, ..., b_8 and t_2, t_3, ... ,t_8 with their respective IS and IF measures. Table 4.2 shows the range of minutia orientation,

Table 4.2: Relation between minutia orientation (θ) and the axis selected for b_1 and t_1.

Range for θ	**Axis for b_1 and t_1**
>=338° and <=22°	ax_1
[23° - 67°]	ax_2
[68° - 112°]	ax_3
[113° - 157°]	ax_4
[158° - 202°]	ax_5
[203° - 247°]	ax_6
[248° - 292°]	ax_7
[293° - 337°]	ax_8

θ, to determine the first axis whose IS and IF values will replace b_1 and t_1 during feature vector formation, respectively. Suppose the orientation of a reference minutia m is $200°$. The initial member in the feature vector representing b_1 and t_1 will be the IS and IF over axis ax_5, i.e., IS_5 and IF_5, respectively. The following members in the feature vector will be over the axis ax_6, ax_7, ax_8, ax_1, ax_2, ax_3, and ax_4. Suppose, in another attempt the finger orientation varies by $60°$ resulting a new orientation of $260°$. Now, the first member in the feature vector, i.e., b_1 and t_1, will be the IS and IF over axis ax_7, i.e., IS_7 and IF_7, respectively. And, the subsequent members in the feature vector will be over the axis ax_8, ax_1, ax_2, ax_3, ax_4, ax_5, and ax_6. This strategy ensured similar feature vectors in both scenarios. Hence, the proposed approach produces a rotation-invariant feature vector that is adaptable to the changes in the orientation.

Suppose $\mathcal{V}_p$ is a feature vector for p^{th} true minutia within an input template T_i and $\mathcal{V}_q$ is a feature vector for q^{th} true minutia within the database template T_j. The feature vectors $\mathcal{V}_p$ and $\mathcal{V}_q$ are comparable provided they both possess good quality. The condition to decide the comparability of $\mathcal{V}_p$ and $\mathcal{V}_q$ is given as, $\mathcal{V}_p \shortparallel \mathcal{V}_q$, *if* $\varphi_p == 1 \ and \ \varphi_q == 1$, where the notation $\shortparallel$ reads *is comparable*.

4.2.4 Similarity score computation

Suppose m_i and m_j denotes the minutiae count in the input and database templates under comparison, respectively. Let, the count of minutiae matched between them is denoted as,

m_m, i.e., possessing minutiae-based features within tolerance limits. The two most widely used similarity score between these templates are given by equation (4.8) [111] and (4.9) [134].

$$s_1 = \frac{m_m^2}{m_i \times m_j} \tag{4.8}$$

$$s_2 = \frac{2 \times m_m}{m_i + m_j} \tag{4.9}$$

The conventional way to compute a similarity score between two templates allows tolerance, e.g. $|\theta - \theta^{'}| \leq \alpha$, where α is the tolerance accepted while comparing the minutiae orientation. However, our proposed approach addresses the MasterPrint vulnerability by implementing zero-tolerance during comparison, i.e., requiring exact feature vectors. The analysis showed that, on average, most partial fingerprints possess $10 - 15$ minutiae. Consequently, imposing an exact match will result in a single-digit minutiae count matching among these templates. Hence, a metric was required to produce a high score for a low matched minutiae count during the comparison to avoid a substantial percentage of the dataset being rejected even at a low threshold value. The existing similarity score computing methods from equation (4.8) and (4.9) generated a low similarity score even when sufficient number of minutiae have matched. Hence, a new metric was employed to compute the similarity score that is necessary and sufficient for producing comparatively high similarity score when small count of minutiae matches between input and database template.

Suppose an input template, T_i, contains λ_t true minutiae feature vectors among which λ_g feature vectors contain good quality minutiae, i.e., $\lambda_g \leq \lambda_t$. Similarly, a database template, T_j, contains γ_t true minutiae feature vectors among which γ_g feature vectors contain good quality minutiae, i.e., $\gamma_g \leq \gamma_t$. The input template is represented as $\mathrm{T_i} = (\mathcal{V}_{i_1}, \mathcal{V}_{i_2}, \ldots, \mathcal{V}_{i_{\lambda_t}})'$ and the database template is represented as $\mathrm{T_j} = (\mathcal{V}_{j_1}, \mathcal{V}_{j_2}, \ldots, \mathcal{V}_{j_{\gamma_t}})'$. Let $\mathcal{V}_p$, $p \in \{i_1, i_2, \ldots, i_{\lambda_g}\}$, and $\mathcal{V}_q$, $q \in \{j_1, j_2, \ldots, j_{\gamma_g}\}$, be feature vectors representing good quality minutia from the input and database templates under comparison, respectively,

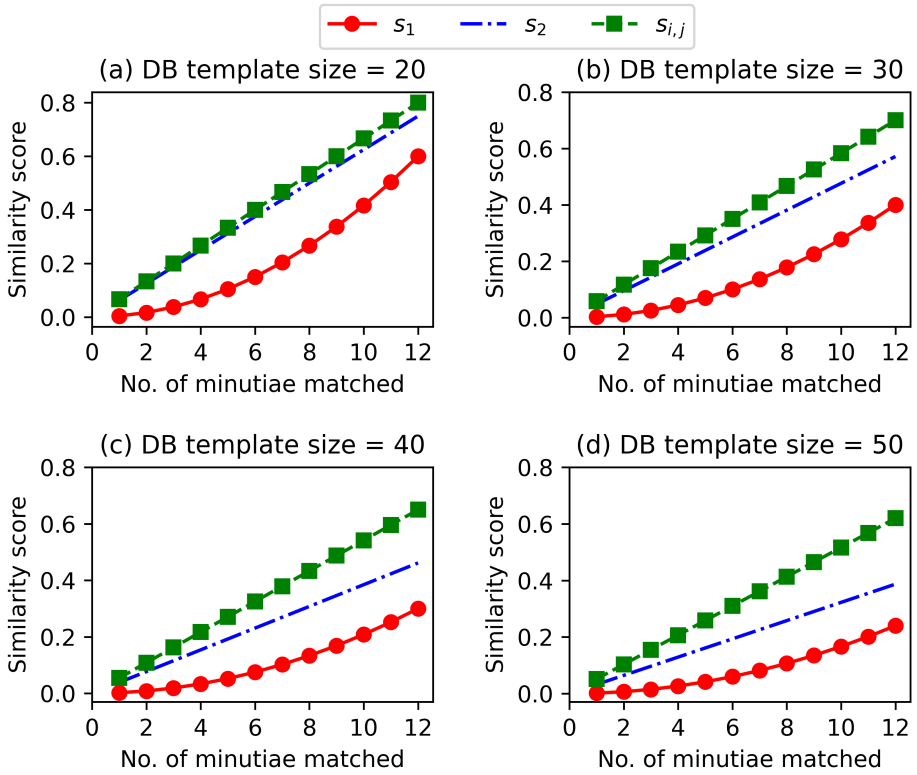


Figure 4.6: Comparison of various similarity score computing metrics. The input template possesses 12 good quality minutiae, and there are 20, 30, 40, and 50 good quality minutiae in database templates for (a), (b), (c), and (d) plots, respectively. If all 12 feature vectors from the input template are equivalent with feature vectors from the database template, the score is approximately close to 1. However, only the similarity score metric $s_{i,j}$ generates a reasonably acceptable score compared to s_1 and s_2.

such that $\mathcal{V}_p \in \mathrm{T_i}$ and $\mathcal{V}_q \in \mathrm{T_j}$. The feature vectors are considered *equivalent*, if $\mathcal{V}_p - \mathcal{V}_q = \mathcal{O}$, where $\mathcal{O}$ is a zero vector. The number of *equivalent* vectors is denoted by $\mathbb{C}$ as,

$$\mathbb{C} = \sum_{r=1}^{min\left(\lambda_g, \gamma_g\right)} \mathrm{X}_r$$

where, $\mathrm{X}_r \in \{0, 1\}, r \in \{1, 2, \ldots, min(\lambda_g, \gamma_g)\}$. The value $\mathrm{X}_r = 1$ signifies the existence of a pair of equivalent vectors between T_i and T_j. The similarity score, $s_{i,j}$, between T_i and T_j is then computed as,

$$s_{i,j} = \frac{1}{2}\left(\frac{\mathbb{C}}{\lambda_g} + \frac{\mathbb{C}}{\gamma_g}\right) \tag{4.10}$$

The similarity scores computed by employing s_1, s_2, and $s_{i,j}$, are shown in red, blue, and green color in Figure 4.6, respectively. In each plot an input template contains 12 good quality minutiae. However, the database template for (a), (b), (c), and (d) plots comprises 20, 30, 40, and 50 good quality minutiae, respectively. The metric, $s_{i,j}$ generated a high score compared to existing schemes. For instance, if all 12 feature vectors from the input template are *equivalent* with the feature vectors from the database template, the expected score should be near to 1. However, only $s_{i,j}$ generated reasonably acceptable higher score values as compared to score values of s_1 and s_2. The remaining plots also depict similar observations.

4.2.5 Algorithm for partial fingerprint-based MasterPrint identification

The steps discussed above for the partial fingerprint-based MasterPrint identification are summarised in Algorithm 1. The algorithm accepts partial fingerprints from a dataset as an input and returns the MasterPrint record (MR) containing MasterPrint Count (MPC) as an output. The algorithm works in two phases, namely, *enrollment* and *identification*. The enrollment of every partial fingerprint is carried out by the first $foreach$ statement at step number 1. However, the second $foreach$ statement at step number 14 compares each enrolled template with every other stored template during the identification process. Initially, preprocessing and true minutiae detection is done till step number 8. The feature extraction is performed for every true minutia in a partial fingerprint under consideration using the $foreach$ statement at step number 9. The function $features(t_m, thin, bin, \theta)$ on step number 10 utilises binarized and thinned partial fingerprint image and returns a feature vector corresponding to each true minutia, t_m, within the image. The feature vector comprises the minutia quality followed by the intensity shift, IS, and intensity factor, IF, values. Before

Algorithm 1: Algorithm for MasterPrint identification

Input: Partial fingerprint, P_i^{ID}, $i \in \{1, 2, \ldots, k\}$, and threshold, Th, $ID \in \{1, 2, \ldots, n\}$

Output: MasterPrint record, *MR*, containing MasterPrint count *MPC*

1 **foreach** $i \in \{1, 2, \ldots, k\}$ **do**
2 $i_m \leftarrow read(P_i^{ID})$ // read an image
3 $roi \leftarrow regionOfInterest(i_m)$ // get region of interest
4 $bin \leftarrow binarization(i_m)$ // perform binarization
5 $thin \leftarrow thinning(bin)$ // apply thinning on binarized image
6 $minu \leftarrow detectMinutiae(thin)$ // minutiae detection
7 $true_{minu} \leftarrow removeMinu(minu, i_m, roi)$ // remove false minutiae
8 $\theta \leftarrow getOrientation(true_{minu})$ // get minutiae orientation
9 **foreach** t_m *in* $true_{minu}$ **do**
10 $[\varphi, IS, IF] \leftarrow features(t_m, thin, bin, \theta)$ // extract features
11 $[T_i, T_i^{ID}] = generateTemplate(\varphi, IS, IF)$ // template generation
12 $db_T \leftarrow [T_i, T_i^{ID}]$ // store in database (db_T)
13 $MPC \leftarrow 0$ // initialise MasterPrint count (MPC) to zero
14 **foreach** $i \in \{1, 2, \ldots, k\}$ **do**
15 $\lambda_g \leftarrow countGoodMinutia(T_i)$ // count good quality minutiae
16 **if** $\lambda_g \geq 10$ **then**
17 $IR_i \leftarrow \emptyset$ // create null *identification record* (IR_i) for T_i
18 **foreach** $j \in \{1, 2, \ldots, k\}$ **do**
19 $\gamma_g \leftarrow countGoodMinutia(T_j)$ // count good quality minutiae
20 **if** $\gamma_g \geq 10$ **and** $i \neq j$ **then**
21 $s_{i,j} \leftarrow computeScore(T_i, T_j)$ // compute score using eq.4.10
22 **if** $s_{i,j} \geq Th$ **then**
23 $IR_i \leftarrow (T_j^{ID}, s_{i,j})$ // make entry in IR_i
24 **if** $countUniqueID(IR_i) \geq (4 \times n/100)$ **then**
25 $MPC \leftarrow MPC + 1$ // increment MPC in MR
26 return MR

starting the identification phase, the MasterPrint count, MPC, is set to zero at step number 13. The $foreach$ statement at step number 18 compares the feature vector corresponding to good quality minutiae from input and database templates. The if statement at step number 24 confirms if the input template under comparison identified 4% distinct subjects from the database and increments the MasterPrint record, MR, if the condition evaluates to $True$. Finally, the MR is assessed to compute the percentage of MasterPrint generated.

Table 4.3: Summary of the cropped, 150×150 px, partial fingerprint datasets used in the experiments.

Dataset	FVC2002 DB1_A	FVC2002 DB2_A	Cross-Match	NIST sd302b	NIST sd302d
Sensor technology	Optical sensor	Optical sensor	Optical sensor	Touch-free	Touch-free
Full fingerprints	800	800	408	920	800
Total subjects	100	100	51	92	100
Samples per subjects	8	8	8	10	8
Image resolution	500 dpi	569 dpi	500 ppi	1000 dpi	500 dpi
Dataset size	2133	1323	1996	1602	1304

4.3 Experimental setup

Five full fingerprint datasets, namely, the standard FVC2002 DB1_A [101], FVC2002 DB2_A [101], CrossMatch Sample DB dataset from NeuroTechnology [102], NIST Special Databases 302 (sd302b and sd302d) [103] were cropped to create partial datasets of 150×150 pixels having 50% overlap between adjacent partial fingerprints. Partial fingerprints cropped from the boundary of the original fingerprint may not contain ridges or contain a tiny ridge pattern with less than ten minutiae. Such partial fingerprints were discarded as they did not fulfil the minimum criteria to participate in the experiment. Hence, the final dataset size employed in the investigations vary. The fingerprint dataset details used during the experiment are provided in Table 4.3.

The average memory space required for storing the partial fingerprint dataset templates was between $2.5 - 3.5$ KB. A desktop system with 64-bit Ubuntu $20.04.3$ LTS (Focal Fossa) operating system having 64 GB internal memory (RAM) and Intel® Xeon(R) CPU $E5-1620$ v3 @ 3.50 GHz $\times 8$ processor was used to perform the experiments employing the MATLAB R2020a computing software.

4.3.1 Existing approaches under comparison

Research on fingerprint identification is relatively less explored compared to verification [133]. Moreover, user identification through partial fingerprint by comparison against thousands of stored templates is rarely reported in the literature. Five diverse existing methods were selected to assess the performance of the proposed approach. The subsequent paragraphs justify employing a particular approach in the experiments.

The NIST Biometric Image Software (NBIS) [99] is the standard software distribution for fingerprint biometric recognition. It has demonstrated high accuracy and is extensively used to evaluate novel approaches, mostly under verification scenarios. However, it was seldom explored for partial fingerprint identification. Hence, NBIS was exercised for identifying partial fingerprints and investigating its effectiveness in addressing the MasterPrint vulnerability.

The *baseline minutiae matching* (BMM) approach is widely accepted as a classical method of fingerprint minutiae matching [118]. The approach stores a minutia as its location coordinates, (x, y), and orientation, θ. A pair of minutiae, m_i and m_j, are reported as matched, if their spatial distance, $s \leq r_0$, and the direction difference, $d \leq h_0$, where r_0 and h_0 denote the tolerance values for spatial distance, s, and direction difference, d. The BMM approach was employed to analyse the impact of tolerance on MasterPrint generation.

As the proposed approach uses local features, it was essential to compare its results against an existing method solely based on local features. The *Speeded-Up Robust Features* (SURF) scheme acts as a local feature detector and descriptor, and is favourably used in computer vision applications [135]. Recently, Kuban et al. [136] amended the original scheme to extend it for fingerprint recognition. The modified SURF method was employed in the investigation to evaluate its accuracy in identifying an individual from a partial fingerprint.

Lee et al. [104] introduced minutiae and *ridge shape features* (RSF) based partial fingerprint recognition method. The method demonstrated higher matching accuracy compared to earlier methods, such as *Minutia Cylinder-Code* (MCC) [19], *representative ridge point* (RRP) [137], Histogram of Oriented Gradients *(HoG)* based matcher [138], and Accelerated-KAZE *(A-KAZE)* based matcher [139], etc. However, the RSF approach was not explored in analysis with MasterPrint. As the RSF approach outperformed the existing four approaches

used in the original paper, it was used in the experiments as the best among those compared.

Roy et al. [7] employed commercial VeriFinger SDK by Neurotechnology while investigating the MasterPrint vulnerability. The original work considered only two fingerprint datasets. The VeriFinger SDK software was experimented with five benchmark datasets, and its performance was evaluated on various parameters, such as DIR, FAR, and the percentage of MasterPrint generated.

4.3.2 Experimental protocols and tests

During the investigation, the MasterPrint vulnerability is considered as an identification problem and a closed-set and open-set identification set-up is followed. In the case of a closed-set scenario, we assume that a person's template already exists in the database. However, there is no such assurance in the latter case wherein the system determines the identity of an unknown user by comparing an input fingerprint with every enrolled template. A watch-list task or application is an example of an open-set identification. A biometric system at a railway station compares the fingerprint of every passenger against a database of criminals or missing individuals. Identification systems often function as an open-set identification task. However, researchers use closed-set identification to evaluate the results of a method as it forms a sound criterion to verify general strengths & weaknesses [97].

In the *identification test*, each input fingerprint is compared with every other stored template from the dataset, and the similarity score for each comparison is computed [97]. A *correct detect and identify* (CDI) indicates that the highest score corresponds to a sample from the actual subject. However, a *false alarm* (FA) shows that the highest score belongs to the sample from some other subject. The rejected fingerprints showed a similarity score less than the system threshold. Let $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{R}$ denote the CDI count, FA count, and the rejected fingerprint count, respectively. Assuming that the system is enrolled with $\mathcal{K}$ fingerprints, such that $\mathcal{K} = \mathcal{C} + \mathcal{F} + \mathcal{R}$, the detect and identification rate (DIR), false alarm rate (FAR), and rejection rate (RR) were computed as $DIR = \frac{\mathcal{C}}{\mathcal{K}} \times 100$, $FAR = \frac{\mathcal{F}}{\mathcal{K}} \times 100$, and $RR = \frac{\mathcal{R}}{\mathcal{K}} \times 100$.

In addition to DIR, FAR, and RR, the percentage of MasterPrint generated and the min-

imum and the maximum number of subjects identified by a MasterPrint were also recorded. An ideal approach suitable for practical use should produce a negligible MasterPrint at a higher identification rate and low FAR.

A *cumulative matching characteristic* (CMC) curve depicts the rank-k DIR for an identification system [98]. The CMC curves used the data from the identification test results for each dataset. If the system is enrolled with k subjects, the rank-k identification rate should be 100%. The partial dataset is cropped, and at least ten good-quality minutiae are required to participate in the experiments. Hence, every partial fingerprint may not have its matching partial fingerprint from the same subject in the database. Thus, the rank-k DIR may not achieve 100% DIR, as expected.

If an approach generates MasterPrints, another experiment, *zero MasterPrint detection test* was conducted by setting a threshold, Th_z, where no MasterPrint were observed, and the DIR and FAR were computed. Further, Th_z was split into ten intermediate, equidistant thresholds to conduct more experiments, and DIR and FAR at each threshold was computed. The statistics from the zero MasterPrint detection test were employed to plot the watchlist *receiver operating characteristics* (ROC) or an identification ROC. The ideal and practical identification system shows minor variations in the DIR and FAR for the given scenario.

4.4 Experimental results

The result analysis is presented in four subsections to report the best and worst-performing methods under various criteria on each dataset. The following subsections analyse the results and plots for the tests conducted in the experiments.

4.4.1 Identification test performance

The performance of various approaches on four evaluation criteria during the identification test conducted on each partial dataset is shown in Table 4.4. The results for the best-performing method on each parameter are marked in boldface. The proposed method outperformed other approaches on every parameter except the FAR. It showed above 90% DIR

Table 4.4: Results for the identification test on various partial datasets. The bold-faced values correspond to the best performing approach on a given dataset for each criteria.

Evaluation Criteria	Approach	DB1_A	DB2_A	Cross Match	sd302b	sd302d
Detect & Identification Rate (DIR) (%)	BMM	37.22	38.47	29.81	3.37	18.71
	VeriFinger	83.97	81.18	37.53	0.75	32.9
	NBIS	87.25	72.49	33.87	16.73	22.62
	SURF	0.09	1.97	3.51	1.87	2.22
	RSF	8.53	8.92	4.91	3.87	2.68
	Proposed	**91.8**	**84.96**	**86.07**	**88.76**	**92.33**
False Alarm Rate (FAR) (%)	BMM	62.45	61.53	64.28	96.5	80.6
	VeriFinger	**0.19**	**0.23**	**0.35**	**0.44**	**0.92**
	NBIS	10.74	11.56	4.71	71.72	8.36
	SURF	99.91	98.03	92.59	98.13	97.78
	RSF	91.14	91.08	30.47	92.2	29.14
	Proposed	8.2	15.04	13.83	9.86	7.67
MasterPrints generated (percentage) (%)	BMM	99.67	100	94.09	99.88	99.31
	VeriFinger	84.15	81.78	37.68	1.06	33.74
	NBIS	97.98	84.05	33.58	88.45	30.98
	SURF	100	100	96.09	100	100
	RSF	99.67	100	35.47	96.07	31.83
	Proposed	**2.25**	**1.74**	**1.45**	**0.37**	**0.54**
No. of subjects identified by a MasterPrint (min, max)	BMM	(98, 98)	(87, 88)	(46, 47)	(83, 86)	(84, 87)
	VeriFinger	(85, 97)	(78, 84)	(39, 42)	(15, 16)	(63, 70)
	NBIS	(43, 98)	(25, 84)	(18, 45)	(42, 80)	(8, 67)
	SURF	(17, 98)	(60, 88)	(38, 46)	(39, 86)	(75, 87)
	RSF	(95, 98)	(82, 88)	(38, 44)	(77, 86)	(51, 70)
	Proposed	**(5,9)**	**(5,8)**	**(5,8)**	**(5,9)**	**(5,6)**

on two datasets, namely DB1_A and sd302d, and $85 - 90\%$ DIR on the remaining datasets. VeriFinger requires higher count of minutiae from partial fingerprints for enrollment. As partial fingerprints missed larger potion of full fingerprints, VeriFinger could not enroll a large portion from the three datasets, namely CrossMatch, sd302b, and sd302d, and offered the lowest FAR on each dataset. The SURF-based scheme produced the worst DIR and FAR while creating a MasterPrint from almost every partial fingerprint. The BMM method identified maximum enrolled subjects with a MasterPrint. The RSF showed poor performance on almost every parameter. NBIS generated more than 66% MasterPrint, on average. The average DIR delivered by the proposed method was 88.8%. The proposed method falsely identified 10.9% partial fingerprints, on an average. However, the major difference the pro-

Table 4.5: Results for zero MasterPrint detection test on various partial datasets. The results for the best-performing method on each parameter are marked in boldface.

Evaluation Criteria	Approach	DB1_A	DB2_A	Cross Match	sd302b	sd302d
Detect & Identification Rate (DIR) (%)	BMM	0.19	12.55	0	0	0
	VeriFinger	**83.83**	**80.88**	36.87	0.62	32.44
	NBIS	53.12	50.04	26.85	3.81	17.64
	SURF	0	0	0.2	0.37	0.31
	RSF	0.8	0.91	1.8	0.62	0.92
	Proposed	81.43	65	**52.2**	**44.26**	**67.48**
False Alarm Rate (FAR) (%)	BMM	0.89	5.52	**0**	0.31	**0**
	VeriFinger	**0**	**0.08**	**0**	**0.06**	0.38
	NBIS	**0**	0.38	0.15	1.87	**0**
	SURF	4.36	23.58	14.53	22.91	20.09
	RSF	4.08	6.27	5.91	13.86	6.67
	Proposed	2.06	4.46	1.85	0.81	3.3

posed method had shown was on the percentage of MasterPrint generated. It generated the lowest, i.e., 1.27% MasterPrint, on average.

4.4.2 Zero MasterPrint detection test performance

The DIR and FAR of various approaches in the zero MasterPrint detection test is shown in Table 4.5. The results corresponding to the best-performing method are in boldface. The proposed method demonstrated the highest DIR on CrossMatch, sd302b, and sd302d dataset. The commercial VeriFinger SDK produced the lowest FAR on each dataset and high DIR on the DB1_A and DB2_A dataset. NBIS delivered marginal FAR but failed to identify even 55% partial fingerprints from any dataset. The SURF and RSF-based methods accurately identified less than 1% and 2% partial fingerprints from each dataset, respectively. The BMM method rejected more than 98% partial fingerprints from most datasets.

4.4.3 CMC and Watchlist ROC curve performance

The CMC plots depicting the DIR at various ranks by each method in the experimentation is illustrated on the left-side portion in Figure 4.7, 4.8, 4.9, 4.10, and 4.11. The best approach

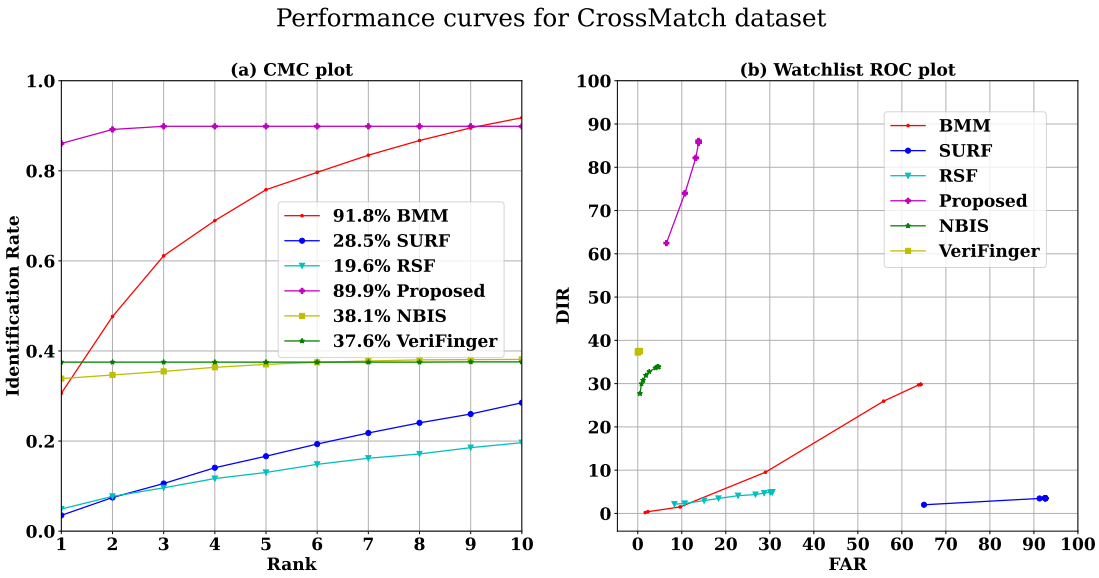


Figure 4.7: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of various approaches on CrossMatch Sample DB partial dataset.

is expected to achieve 100% identification rate at the earliest, i.e., at lower ranks. Hence, the CMC plots showed rank-10 performance of each method under consideration. The proposed method delivered the highest average rank-10 DIR, i.e., 90.74%, followed by 76.36% for the BMM method. However, the RSF and SURF-based approach showed the lowest average rank-10 DIR of 25.28% and 17.04%, respectively. Moreover, NBIS and VeriFinger achieved above 83% and 81% accuracy on the FVC datasets, respectively. The Watchlist ROC plots depicting variation in DIR versus FAR on each dataset by every method during the experimentation are illustrated on the right-side portion in Figure 4.7, 4.8, 4.9, 4.10, and 4.11. The observations showed that in most plots, the DIR for the BMM, SURF, and the RSF method declines drastically and reaches below 10%. However, the DIR varies insignificantly in the case of the remaining three approaches.

4.5 Discussion

This section provides a discussion over the performance of each approach on various parameters of the experiments. The results analysis from the previous section showed that the proposed method produced an average rank-10 DIR of 90.74%. Hence, it can be deduced

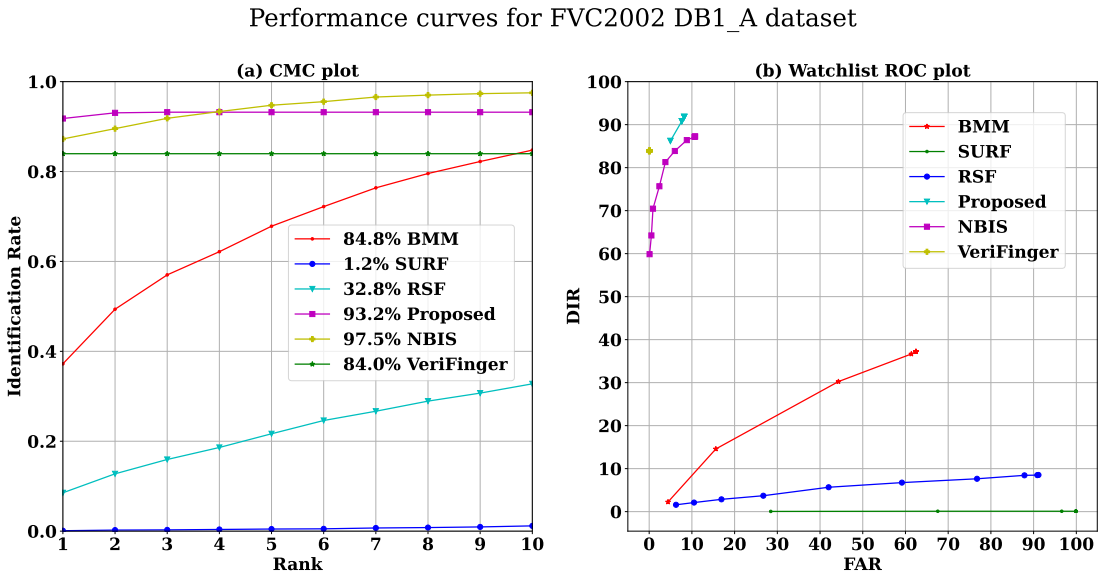


Figure 4.8: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of various approaches on FVC2002 DB1_A partial dataset.

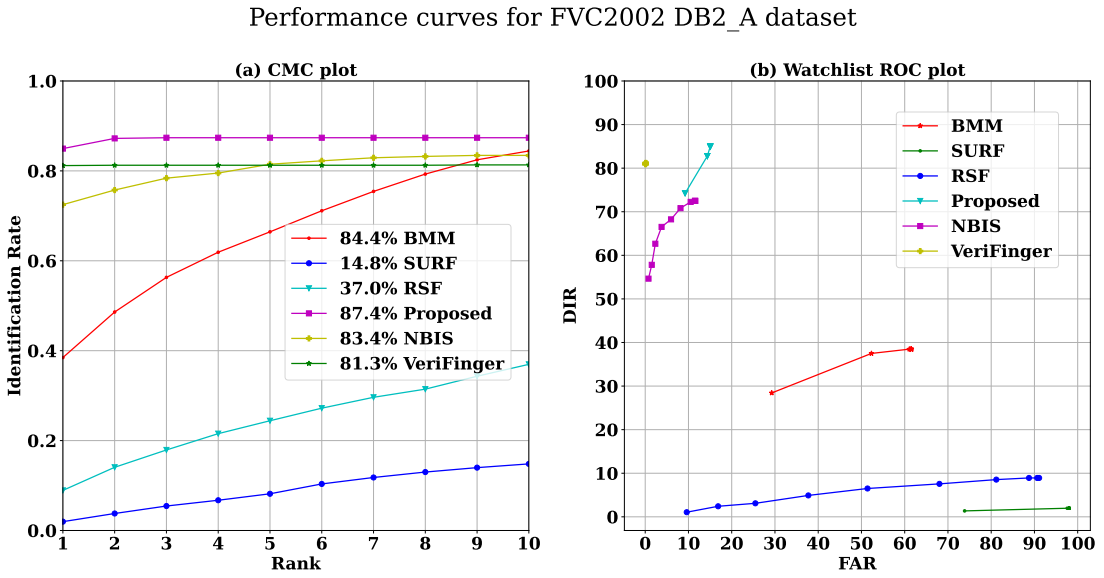


Figure 4.9: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of various approaches on FVC2002 DB2_A partial dataset.

that the proposed method identifies an individual with a partial fingerprint more accurately than other methods. Moreover, it generated 1.27% of MasterPrints on average, which is a significantly smaller percentage. The results correspond to experiments while identifying the lowest average number of distinct users, i.e., 8, from each dataset. Therefore, the proposed method emerges appropriate for IoT-based home appliances, smart devices, etc.

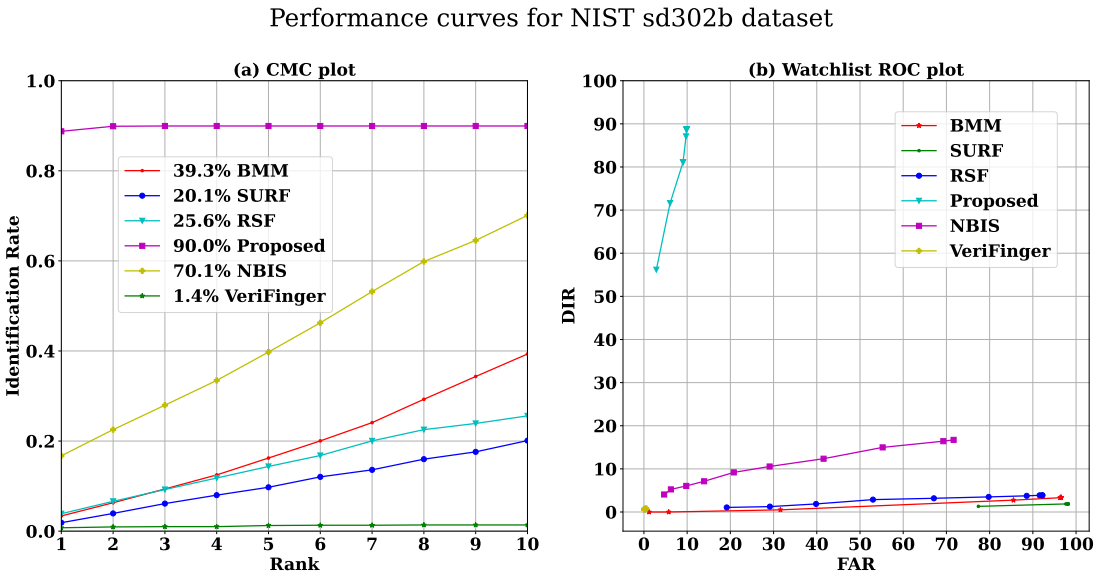


Figure 4.10: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of various approaches on NIST sd302b partial datasets.

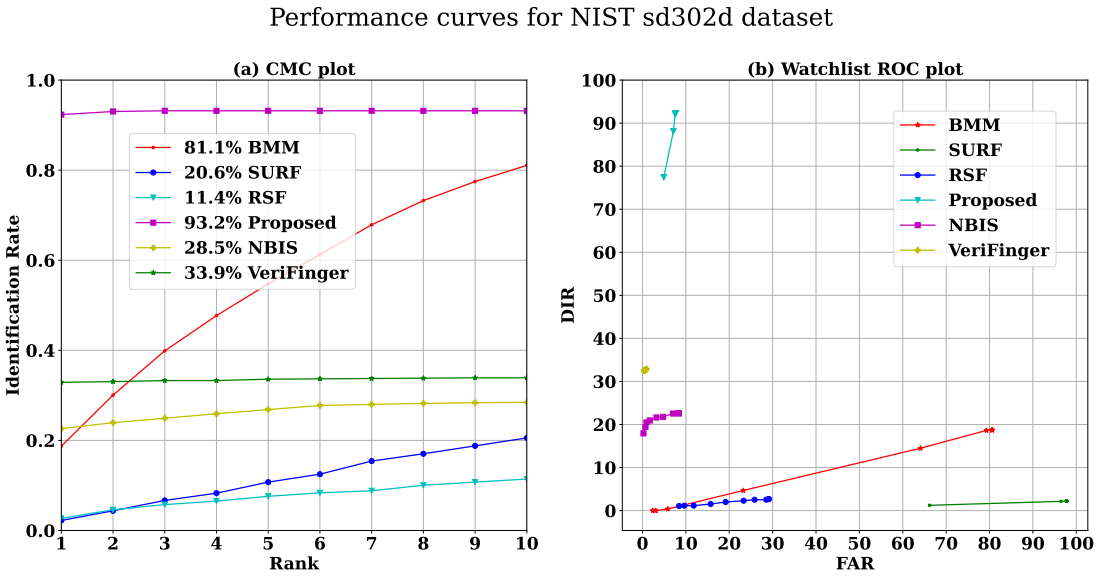


Figure 4.11: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of various approaches on NIST sd302d partial dataset.

Lee et al. [104] termed the concave and convex ridge patterns as fixed-size ridge shape features, *RSFs*. However, due to the small-sized partial fingerprints employed in our experiments, these RSFs were missing in a large portion of each dataset. Therefore, the RSF approach delivered higher RR on every dataset. The approach extracted RSFs over gray-scale partial fingerprints and employed a weighting factor, λ, $(0 \leq \lambda \leq 1)$ during the match-

ing stage. These two steps triggered generation of large MasterPrints for the method. The proposed approach extracted local features within a small region around a minutia, applied preprocessing steps, and enforced strict matching to overcome the limitations of the RSF method.

NBIS is solely minutiae-based and is considered a robust approach for full fingerprint verification. The accuracy of NBIS depends on the number of minutiae detected within a fingerprint. Due to fewer minutiae observed in partial fingerprints, its RR was impacted significantly. However, its marginal FAR demonstrated that it is highly accurate in fingerprint verification. The proposed method extracts an integer-valued feature vector within a small region around a reference minutia. Hence, it rejected a small percentage of each dataset compared to NBIS.

VeriFinger SDK requires a large portion of a fingerprint to extract minutiae-based features. Its accuracy reduces if a small number of minutiae are detected in a fingerprint. The SDK rejected a substantial percentage of partial fingerprints during the enrollment phase, raising an exception as "*Extraction failed: BadObject*". Consequently, VeriFinger SDK demonstrated a high RR and low DIR on most datasets. However, the marginal FAR during the identification test and zeroed FAR in the zero MasterPrint detection test reflected its robustness as a successful commercial product in the market.

The preprocessing stage in the proposed approach comprising the binarization and thinning operation ensured that the local features seldom deviate between different samples of a finger. As the feature vectors for a finger rarely vary, the method achieved an improved identification rate. However, the modified SURF-based method failed to accurately extract the essential local feature descriptor at interest points from a partial dataset during the experiments. These local features were similar to other partial fingerprints due to non-uniqueness. Consequently, the DIR for the modified SURF-based method reduced immensely and eventually produced a large number of MasterPrints.

The proposed method imposed strict feature vector matching to ensure high identification accuracy and low chances of MasterPrint generation. The results demonstrated that the BMM approach rejected a small portion of the fingerprints from each dataset. However, the method detected similar features between samples from different subjects in the dataset. Moreover, the BMM method accepts a tolerance while feature matching. As a result, it gen-

erates similarity scores for slightly similar fingerprints from distinct subjects. The approach employs minutiae location and orientation as its features, which are more likely to match other subjects' partial fingerprints, if tolerance is accepted. Hence, it is evident that the primary reason behind the method's poor DIR, high FAR, and huge MasterPrint generation lies in allowing tolerance during the matching stage.

4.6 Summary of the chapter

This chapter presented a novel minutiae-based MasterPrint mitigation method and introduced a similarity score computation metric. The approach extracted local features around a reference minutia in eight directions to create a unique feature vector and performed a zero-tolerance feature vector comparison. The method was experimented on partial fingerprint datasets cropped from five benchmark fingerprint datasets, and its results were compared with five existing methods, namely, NIST Biometric Image Software (NBIS), baseline minutiae matching (BMM) approach, modified Speeded-Up Robust Features (SURF) scheme, Ridge Shape Features (RSF) based method, and VeriFinger SDK. The proposed method demonstrated a high identification rate and generated insignificant MasterPrints on each dataset. The proposed method has achieved an average accuracy of 88.78% and generated on an average 1.27% MasterPrints. Thus, it is evident that MasterPrint vulnerability can be reduced with a local feature-based partial fingerprint identification scheme with strict feature matching.

Chapter 5

MasterPrint Mitigation Employing Minutiae Geometry

The investigation in the previous chapter ascertained that a minutiae-based local-feature extraction method delivers high identification accuracy and generates an insignificant count of MasterPrints from cropped partial fingerprint datasets. However, the feature extraction was carried out within a 27×27 pixel-wide square around a reference minutia. In another attempt to further improve the identification rate, and reduce the percentage of MasterPrint generated in the previous method, this chapter employs minutiae geometry to build a unique feature vector corresponding to every minutia in a partial fingerprint. The approach requires exact feature vectors during the comparison. The experiments used partial fingerprint datasets cropped from five benchmark fingerprint datasets, namely, FVC2002 DB1_A, FVC2002 DB2_A, CrossMatch Sample DB dataset from NeuroTechnology, NIST Special Databases 302 (sd302b and sd302d), and the results of the proposed method were compared against six existing methods, namely, NIST Biometric Image Software (NBIS), baseline minutiae matching (BMM) approach, modified Speeded-Up Robust Features (SURF) scheme, Ridge Shape Features (RSF) based method, Combination of Nearest Neighbor Arrangement Indexing (CNNAI)-based method, and VeriFinger SDK.

The main highlights of the work in this chapter are as follows,

1. Constructing a novel feature vector for the distinctive representation of a minutia,

2. Studying the partial fingerprint identification to investigate MasterPrint generation.

The contents of this chapter are organised as follows. Initially, Section 5.1 describes various steps involved in the proposed partial fingerprint identification method. The experimental setup, including the fingerprint dataset details and other existing methods employed in the experiments, is discussed in Section 5.2. Further, the experimental results on various parameters and plots are presented in Section 5.3. The discussion and the significance of the experimental results are provided in Section 5.4. Finally, Section 5.5 summarises the chapter.

5.1 Proposed approach

The proposed method for MasterPrint identification from partial fingerprint dataset is shown in Figure 5.1. The flowchart is divided into enrollment and identification phases. A partial fingerprint undergoes binarization, thinning, minutiae detection, and template generation stages in the enrollment phase. The template database stores the encrypted partial fingerprint templates. The i^{th} partial fingerprint corresponding to a user having identifier as ID is denoted as P_i^{ID}. T_i denotes the feature template for P_i^{ID}. The system is enrolled with k partial fingerprints for n distinct subjects. The template database stores multiple templates associated with each subject. Hence, in the experimentation k is greater than n, i.e., $k > n$.

In the identification phase, a partial fingerprint template, T_i, is compared with every other template, T_j, from the database, i.e., the system performs comparison in all-versus-all manner. For every comparison between T_i and T_j, a similarity score, $s_{i,j}$, is computed. The system maintains an identification record, IR_i, for every T_i. An entry in IR_i comprises T_i, T_j, and $s_{i,j}$, where $s_{i,j}$ must be greater than the system threshold, Th. Once the comparison between T_i with every other T_j in the database is finished, its IR_i is assessed to determine if T_i identified at least 4% distinct subjects enrolled with the system. If so, the system labels it as a MasterPrint by incrementing the MasterPrint count, MPC, in the MasterPrint record, MR. Finally, the MR is assessed to get the percentage of MasterPrints generated during the experiment. The details about the feature extraction, feature vector formation, and the similarity score computation are provided in the following sections.

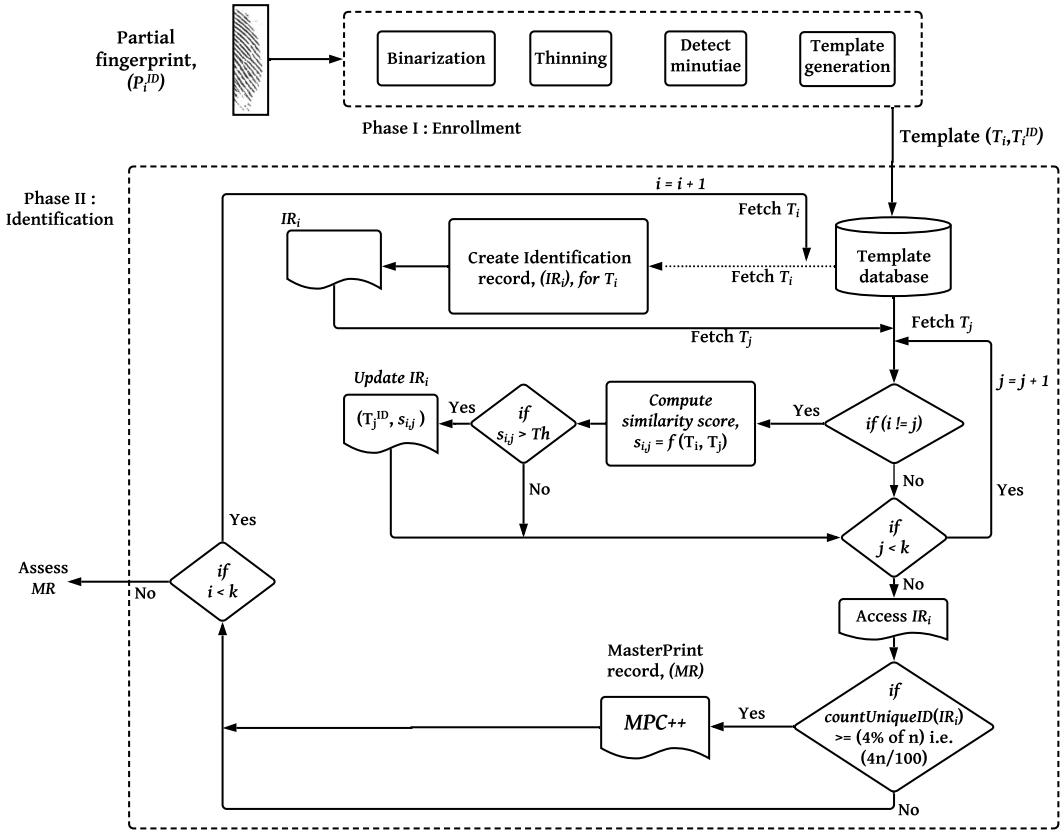


Figure 5.1: Flowchart for partial fingerprint-based MasterPrint identification. The enrollment phase takes a partial fingerprint, P_i^{ID} as input and enrols the template T_i and corresponding user identifier, T_i^{ID}. The database contains k templates for n users such that $i \in \{1, 2, \ldots, k\}$, $ID \in \{1, 2, \ldots, n\}$, and $k > n$. The dotted arrow originating from the template database denotes the start of the identification phase. The system threshold is set at Th, and the MasterPrint count (MPC) is initialised to zero. Initially, T_i and T_j corresponds to the first template in the database, and $s_{i,j}$ represents the similarity score between T_i and T_j.

5.1.1 Preprocessing and minutiae detection

Usually, the gray-scale fingerprints obtained at different instances for the same finger vary due to moisture or dryness at the fingertip, finger orientation, and variation in pressure applied while touching the sensor. Hence, a binarization process is usually applied to the gray-scale fingerprints to render them appear more similar to each other. The proposed approach performed the binarization of gray-scale partial fingerprints using an adaptive thresholding method [140]. The performance of the fingerprint matching algorithm highly depends on accurate minutiae detection. As thinned images possess a single-pixel ridge, they are preferred

for minutiae detection than binarized images. Therefore, the proposed method performs a thinning operation on the binarized partial fingerprints using Zhang-Suen thinning algorithm [130]. The minutiae detection process on the thinned images is done by computing the Crossing Number (CN) [131].

5.1.2 Feature extraction

Each minutia is stored as its (x, y) coordinates within the partial fingerprint. The Euclidean distance between a reference minutia, $m(x, y)$, and its n adjacent minutiae, $m_1(x_1, y_1)$, $m_2(x_2, y_2)$, ..., $m_n(x_n, y_n)$, is maintained and utilised during the feature extraction process. The set of tuples, M, represents the x and y coordinates for a reference minutia and its adjacent minutiae as

$$M = \{(x, y), (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{5.1}$$

The first minutiae-based geometric construct is a closed curve, Δ, comprising the minimal elements from M, such that no other member of M is left outside the boundary of Δ. An instance of Δ for $n = 3$ is shown in Figure 5.2. It is evident that the minimum and the maximum number of vertices for a given Δ can be 3 and $(n + 1)$, respectively. Hence, the cardinality of Δ, $\eta(\Delta)$, ranges from $3 \leq \eta(\Delta) \leq (n + 1)$. The closed curve, Δ, is revisited to verify the presence of the reference minutia coordinates as its vertex. The binary-valued metric, ϱ, indicates whether the reference minutia belongs to the closed curve Δ.

$$\varrho = \begin{cases} 1, & \text{if } (x, y) \in \Delta \\ 0, & \text{otherwise} \end{cases} \tag{5.2}$$

A tuple, T, maintains the Euclidean distances between a reference minutia and its every adjacent minutiae. A minutiae pair corresponding to the smallest Euclidean distance in T is chosen as the first edge of another minutiae-based construct, ∇, comprising non-intersecting

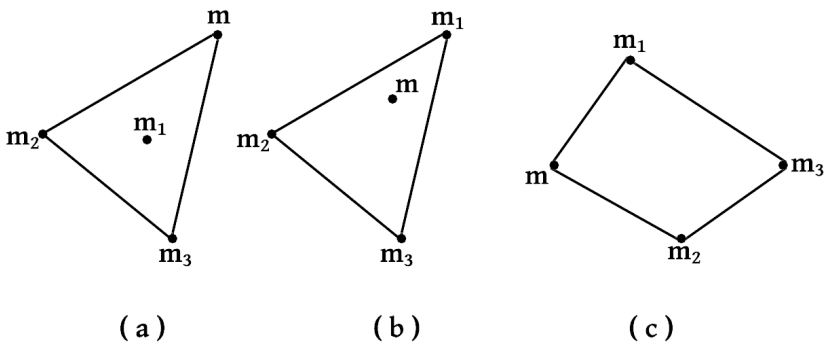


Figure 5.2: Possible closed curves formed with $n = 3$ enclosing all the members of M.

edges. The subsequent edges are added to ∇ in ascending Euclidean distances while ensuring that a new edge does not intersect with the existing edges in ∇. If the $n + 1$ minutiae points are non-collinear, ∇ will possess several triangular shapes. The possible constructions of ∇ formed using $n = 3$ are shown in Figure 5.3. The count of all possible triangles formed in a ∇ construct is denoted as τ. The ∇ construct is revisited to find the number of triangles with the reference minutia, $m(x, y)$, as its vertex. This number is represented as $\mathfrak{f} \in \{0, 1, 2, \ldots, \tau\}$.

The third minutiae-based geometric construct is a non-self-intersecting poly shape or polygon, $\diamondsuit$. The construct $\diamondsuit$ possesses $n + 1$ vertices, i.e., it has every member from M as a vertex. A possible $\diamondsuit$ construct with $n = 3$ is depicted in Figure 5.4. The construct shows four vertices, namely, $v_1(x_1, y_1)$, $v_2(x_2, y_2)$, $v_3(x_3, y_3)$, and $v_4(x_4, y_4)$. The Euclidean distance between two vertices v_i and v_j is denoted as, len_{ij}, such that, $i, j \in \{1, 2, 3, 4\}$. The area, λ, and perimeter, μ, of the $\diamondsuit$ are computed using equation (5.3) and (5.4), respectively and stored as integer values.

$$\lambda = \frac{1}{2}\left(\sum_{i=1}^{4}(x_i \times y_{i+1} - x_{i+1} \times y_i)\right), \text{where } x_5 = x_1 \, and \, y_5 = y_1 \tag{5.3}$$

$$\mu = len_{12} + len_{23} + len_{34} + len_{41} \tag{5.4}$$

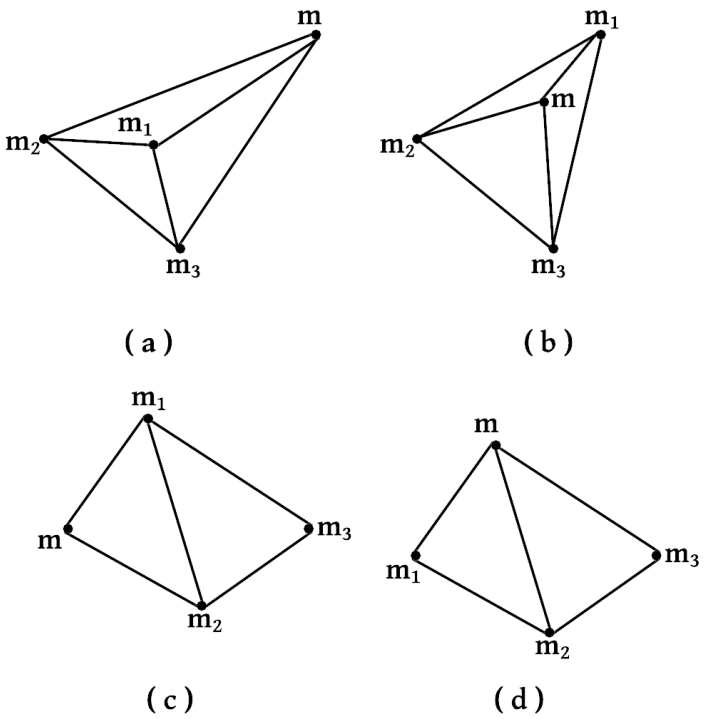


Figure 5.3: Possible non-intersecting constructs with $n = 3$.

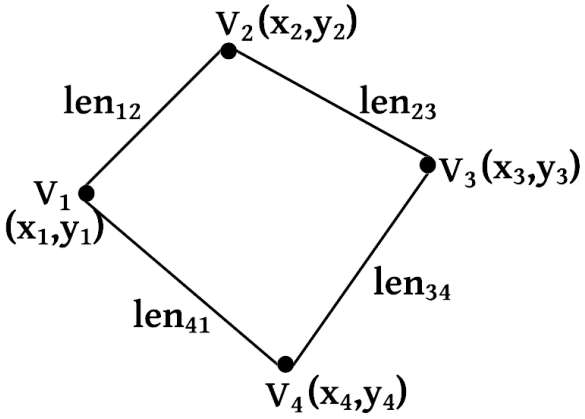


Figure 5.4: A sample non-self-intersecting poly shape or polygon, for $n = 3$, connecting every member of M.

The feature vector, υ, corresponding to a reference minutia in the partial fingerprint is represented as,

$$\upsilon = (\eta(\Delta), \varrho, \tau, \mathfrak{f}, \lambda, \mu) \tag{5.5}$$

If a reference minutia, m, and its n adjacent minutiae are collinear, then the feature vector for m comprises every member of υ as -1. The feature vectors for such minutiae are discarded during similarity score computation.

5.1.3 Similarity score computation

The comparison between a feature vector υ_p from the input template, T_i, and a feature vector, υ_q, from the database template, T_j, is performed as per equation (5.6), where ω is a binary variable.

$$\omega = \begin{cases} 1\,, & \text{if } \upsilon_p - \upsilon_q == \mathcal{O} \\ 0\,, & \text{otherwise} \end{cases} \tag{5.6}$$

where, $\mathcal{O}$ represents a zero vector. Suppose T_i possesses $\mathcal{X}$ minutiae, and T_j contains $\mathcal{Y}$ minutiae such that $\mathcal{X} \geq 10$, $\mathcal{Y} \geq 10$, and $\alpha = min(\mathcal{X}, \mathcal{Y})$. Then the total count of minutiae matched, M_m, is calculated using equation (5.7).

$$M_m = \sum_{i=1}^{\alpha} \omega_i \tag{5.7}$$

Finally, the similarity score between T_i and T_j is calculated using equation (5.8) [111].

$$s_{i,j} = \frac{{M_m}^2}{\mathcal{X} \times \mathcal{Y}} \tag{5.8}$$

A partial fingerprint must possess at least ten minutiae to participate in the experiments.

The proposed method was experimented with different values of adjacent minutiae, $n \in \{3, 4, 5, 6\}$. The observations from each experiment demonstrated that $n = 4$ delivered the best results. Hence, the results used in this chapter correspond to $n = 4$.

5.2 Experimental setup

Five full fingerprint datasets, namely, the standard FVC2002 DB1_A [101], FVC2002 DB2_A [101], CrossMatch Sample DB dataset from NeuroTechnology [102], NIST Special Databases 302 (sd302b and sd302d) [103] were cropped to create partial fingerprint datasets of 150×150 pixels having 50% overlap between adjacent partial fingerprints. Partial fingerprints cropped from the boundary of the original fingerprint may lack ridges or possess a tiny ridge pattern with less than ten minutiae. Such partial fingerprints were discarded as they did not fulfil the minimum criteria to participate in the experiment. Hence, the final dataset size employed in the investigations varies. The fingerprint dataset details used during the experiment are provided in Table 5.1. A desktop system with 64-bit Ubuntu $20.04.3$ LTS (Focal Fossa) operating system having 64 GB internal memory (RAM) and Intel® Xeon(R) CPU $E5-1620$ v3 @ 3.50 GHz $\times 8$ processor was used to perform the experiments employing the MATLAB R2020a computing software.

Table 5.1: Summary of the cropped, 150×150 px, partial fingerprint datasets used in the experiments.

Dataset	FVC2002 DB1_A	FVC2002 DB2_A	Cross-Match	NIST sd302b	NIST sd302d
Sensor technology	Optical sensor	Optical sensor	Optical sensor	Touch-free	Touch-free
Full fingerprints	800	800	408	920	1600
Total subjects	100	100	51	92	200
Samples per subjects	8	8	8	10	8
Image resolution	500 dpi	569 dpi	500 ppi	1000 dpi	500 dpi
Cropped dataset size	3549	2767	2134	2098	2960

5.2.1 Existing approaches under comparison

Research on fingerprint identification is relatively less explored compared to verification [133]. Moreover, user identification through partial fingerprint by comparison against thousands of stored templates is rarely reported in the literature. Six diverse existing methods were selected to assess the performance of the proposed approach. The subsequent paragraphs justify employing a particular approach in the experiments.

The NIST Biometric Image Software (NBIS) [99] is the standard software distribution for fingerprint biometric recognition. It has demonstrated high accuracy and is extensively used to evaluate novel approaches, mostly under verification scenarios. However, it was seldom explored for partial fingerprint identification. Hence, NBIS was exercised to investigate identifying partial fingerprints and addressing MasterPrint vulnerability.

The baseline minutiae matching (BMM) approach is widely accepted as a classical means of fingerprint minutiae matching [118]. The approach stores a minutia as its location coordinates, (x, y), and orientation, θ. A pair of minutiae, m_i and m_j, are reported as matched, if their spatial distance, $s \leq r_0$, and the direction difference, $d \leq h_0$, where r_0 and h_0 denote the tolerance values of spatial distance, s, and direction difference, d. BMM approach was employed to analyse the impact of tolerance on MasterPrint generation.

As the proposed approach uses local features, it was essential to compare its results against an existing method solely based on local features. The Speeded-Up Robust Features (SURF) scheme acts as a local feature detector and descriptor and is highly utilised in computer vision applications [135]. Recently, Kuban et al. [136] amended the original scheme to extend it for fingerprint recognition. The modified SURF method was employed in the investigation to evaluate its accuracy in identifying an individual from a partial fingerprint.

Lee et al. [104] introduced minutiae and ridge shape features (RSF) based partial fingerprint recognition method. The method demonstrated higher matching accuracy compared to earlier methods, such as Minutia Cylinder-Code (MCC) [19], representative ridge point (RRP) [137], HoG-based matcher (HoG) [138], and A-KAZE-based matcher (A-KAZE) [139], etc. However, the RSF approach was not explored in addressing MasterPrint vulnerability. As the RSF approach outperformed the existing four approaches used in the original paper, it was considered as the best among those compared in the original experiments.

Roy et al. [7] employed commercial VeriFinger SDK by Neurotechnology while investigating the MasterPrint vulnerability. The original work considered only two fingerprint datasets. The VeriFinger SDK software was experimented with five benchmark datasets, namely, the standard FVC2002 DB1_A, FVC2002 DB2_A, CrossMatch Sample DB dataset from NeuroTechnology, NIST Special Databases 302 (sd302b and sd302d), and its performance was evaluated on essential parameters, such as DIR, FAR, percentage of MasterPrints generated, and number of subjects identified by a MasterPrint.

Deshpande et al. [129] developed *Combination of Nearest Neighbor Arrangement Indexing* (CNNAI) as a local minutia-based convolutional neural network (CNN) matching model. The model uses features from n nearest neighbour minutiae to generate feature vectors that are rotation-scale invariant. CNNAI performs minutiae detection using a deep neural-network-based minutiae extractor MINU-EXTRACTNET [141]. The model demonstrated more than 80% rank-1 identification accuracy on FVC2004 and NIST SD27 latent fingerprint datasets. CNNAI was exercised as a deep learning-based approach to compare the results of the proposed method.

5.2.2 Experimental protocols and test

The investigation considered the MasterPrint vulnerability as an identification problem, and followed a closed-set and open-set identification set-up. In the case of a closed-set scenario, we assume that a person's template already exists in the database. However, there is no such assurance in the latter case wherein the system determines the identity of an unknown user by comparing an input fingerprint with every enrolled template. A watch-list task or application is an example of an open-set identification. A biometric system at a railway station compares the fingerprint of every passenger against a database of criminals or missing individuals. Identification systems often function as an open-set identification task. However, researchers experiment under closed-set identification to evaluate and report the results of a method as it forms a sound criterion to verify general strengths and weaknesses [97].

In the *identification test*, each input fingerprint is compared with every other stored template from the dataset, and the similarity score for each comparison is computed [97]. A

correct detect and identify (CDI) represents that the highest score corresponds to a sample from the actual subject. However, a *false alarm* (FA) shows that the highest score belongs to the sample from some other subject. The rejected fingerprints showed a similarity score less than the system threshold. Let $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{R}$ denote the CDI count, FA count, and the rejected fingerprint count. Assuming that the system is enrolled with $\mathcal{K}$ fingerprints, such that $\mathcal{K} = \mathcal{C} + \mathcal{F} + \mathcal{R}$, the detect and identification rate (DIR), false alarm rate (FAR), and rejection rate (RR) were computed as, $DIR = \frac{\mathcal{C}}{\mathcal{K}} \times 100$, $FAR = \frac{\mathcal{F}}{\mathcal{K}} \times 100$, and $RR = \frac{\mathcal{R}}{\mathcal{K}} \times 100$. In addition to DIR, FAR, and RR, the percentage of MasterPrint generated and the minimum and the maximum number of subjects identified by a MasterPrint were also recorded. An ideal approach suitable for practical use should produce a negligible MasterPrint at a higher identification rate and low FAR.

A *cumulative matching characteristic* (CMC) curve depicts the rank-k DIR for an identification system [98]. The CMC curves were plotted based on the data from the identification test results for each dataset. If the system is enrolled with k subjects, the rank-k identification rate should be 100%. If an approach generates MasterPrints, another experiment, *zero MasterPrint detection test*, was conducted by setting a threshold, Th_z, where no MasterPrint were observed, and the DIR and FAR were computed. Furthermore, Th_z was split into ten intermediate, equidistant thresholds to conduct more experiments, and DIR and FAR at each threshold were computed. The statistics from the zero MasterPrint detection test were employed to plot the watchlist *receiver operating characteristic* (ROC) or an identification ROC. The ideal and practical identification system shows minor variations in the DIR and FAR for the given scenario.

5.3 Experimental results

The result analysis is presented in four subsections to report the best and worst-performing methods under various criteria on each dataset. The following subsections analyse the results and plots for the tests conducted in the experiments.

5.3.1 Identification test performance

The identification test performance of various approaches on each partial dataset is presented in Table 5.2. Each method is evaluated on four criteria, namely, DIR, FAR, percentage of MasterPrints generated, and minimum and maximum number of subjects identifies by a MasterPrint, for each dataset. The boldfaced values from the table demonstrate the best results for a given parameter in a dataset. An approach is expected to deliver at least 90% DIR and generate marginal MasterPrint on the datasets, which identify fewer distinct subjects enrolled with the system. Any method providing the desired results can be used for practical purposes to identify a user through a partial fingerprint.

The proposed method showed better results on every parameter, except the FAR. The DIR for the proposed method on each dataset was above 90%. It generated less than 1% MasterPrint for three datasets and $1\% - 2\%$ MasterPrints for the other two datasets. A MasterPrint from the proposed method could identify the lowest number of distinct subjects on each dataset. However, a MasterPrint from the BMM approach on the FVC2002 DB1_A dataset could identify a minimum of 95 subjects, and the number rises to 100, i.e., identifying every enrolled user.

The best FAR was demonstrated by the VeriFinger and CNNAI, i.e., up to 0.5% on each dataset. The SURF-based scheme demonstrated low DIR, average DIR of 1.69%, and high FAR, average FAR of 98.16%. Moreover, it reported every partial fingerprint from most datasets as a MasterPrint. A MasterPrint from the SURF, BMM, and RSF methods identified the maximum distinct subjects from the dataset. The average highest percentage of MasterPrint were produced by the SURF-based method, i.e., 99.09%, followed by 96.52% from the BMM method. The average lowest percentage of MasterPrint were generated using the proposed method, i.e., 0.882%.

5.3.2 Zero MasterPrint detection test performance

The results obtained during the zero MasterPrint detection test on each partial dataset are presented in Table 5.3. Each approach is evaluated on three criteria, DIR, FAR, and RR, on

Table 5.2: Results for the identification test on various partial datasets. The boldfaced values correspond to the best performing approach on a given dataset under each criteria.

Evaluation Criteria	Approach	FVC 2002 DB1_A	FVC 2002 DB2_A	CrossMatch Sample DB	NIST sd302b	NIST sd302d
Detect and Identification Rate (DIR) (%)	BMM	16.17	14.13	9.33	7.05	9.36
	VeriFinger	50.46	38.81	35.1	0.57	14.49
	SURF	1.3	0.94	3.8	1.72	0.71
	RSF	2.38	2.42	6.89	1.53	1.72
	NBIS	30.57	22.05	33.74	9.49	41.96
	CNNAI	13.09	9.68	6.8	3.27	17.96
	Proposed	**93.43**	**91.69**	**94.24**	**97.04**	**92.64**
False Alarm Rate (FAR) (%)	BMM	78.67	83.66	84.21	90.23	89.8
	VeriFinger	**0.06**	**0.07**	0.19	0.38	0.41
	SURF	97.97	99.06	96.2	98.28	99.29
	RSF	37.39	27.39	31.02	43.28	34.59
	NBIS	9.07	7.77	4.26	35.18	30.88
	CNNAI	0.09	0.08	**0.03**	**0.2**	**0.07**
	Proposed	6.57	8.31	5.76	2.96	7.36
Percentage of MasterPrints generated (%)	BMM	94.84	97.8	93.53	97.28	99.16
	VeriFinger	50.52	38.89	37.57	0.95	14.9
	SURF	99.27	100	96.2	100	100
	RSF	39.62	29.82	37.91	44.8	36.32
	NBIS	39.64	29.82	38	44.6	72.84
	CNNAI	13.18	9.76	6.83	3.47	18.03
	Proposed	**1.04**	**0.4**	**0.84**	**0.1**	**2.03**
Number of subjects identified by a MasterPrint (min, max)	BMM	(95, 100)	(94, 100)	(13, 49)	(69, 79)	(170, 172)
	VeriFinger	(73, 96)	(78, 84)	(39, 42)	(15, 16)	(63, 70)
	SURF	(51, 100)	(42, 100)	(40, 49)	(62, 79)	(153, 172)
	RSF	(98, 100)	(87, 95)	(42, 48)	(39, 72)	(120, 150)
	NBIS	(35, 100)	(26, 95)	(15, 48)	(30, 72)	(30, 147)
	CNNAI	(5, 9)	(6, 8)	(5, 11)	(6, 13)	(12, 20)
	Proposed	**(5, 6)**	**(5, 6)**	**(5, 9)**	**(5, 5)**	**(10, 11)**

every dataset. The best results for each parameter on a given dataset are marked in boldface. A method demonstrating marginal deviation in the DIR and FAR compared with its identification test results is considered the robust identification approach. However, a scheme delivering high RR during the zero MasterPrint detection test is regarded as practically inappropriate for identifying a user from thousands of database templates.

The proposed approach delivered the highest average DIR during the test, i.e., 75.15%.

The lowest FAR on maximum datasets was from the VeriFinger and NBIS schemes. The SURF and RSF-based methods demonstrated less than 1% DIR on average. However, the BMM method couldn't identify a user through the partial fingerprint from any dataset during the test. The results showed that four methods, namely, the SURF-based, BMM, RSF-based, and CNNAI, produced low DIR than FAR and above 90% RR on average. The CNNAI showed, on average, 1.2% less DIR than the identification test performance. However, no partial fingerprint falsely identified a subject by the CNNAI during the test.

Table 5.3: Results for zero MasterPrint detection test on various partial datasets. The results for the best performing approach on each criteria are marked in boldface.

Evaluation Criteria	Approach	FVC 2002 DB1_A	FVC 2002 DB2_A	CrossMatch Sample DB	NIST sd302b	NIST sd302d
Detect and Identification Rate (DIR) (%)	BMM	–	–	–	–	–
	VeriFinger	50.38	38.67	34.49	0.48	14.29
	SURF	0.17	0.07	–	0.1	0.1
	RSF	0.51	0.18	1.78	0.1	0.37
	NBIS	15.38	14.53	26.9	3.05	25.71
	CNNAI	10.76	8.6	6.2	2.97	16.33
	Proposed	**76.5**	**64.87**	**85.29**	**83.03**	**71.05**
False Alarm Rate (FAR) (%)	BMM	–	–	–	–	–
	VeriFinger	–	**0.04**	–	**0.05**	**0.17**
	SURF	9.83	5.89	3.98	5.48	17.33
	RSF	4.93	2.17	3.09	6.29	2.7
	NBIS	**0.08**	0.14	**0.28**	0.38	5.95
	CNNAI	–	–	–	–	–
	Proposed	4.54	4.23	5.25	2.76	3.45
Rejection Rate (RR) (%)	BMM	100	100	100	100	100
	VeriFinger	49.62	61.29	65.51	99.48	85.54
	SURF	90	94.04	96.02	94.42	82.57
	RSF	94.56	97.65	95.13	93.61	96.93
	NBIS	84.53	85.33	72.82	96.57	68.34
	CNNAI	89.24	91.4	93.8	97.03	83.67
	Proposed	**18.96**	**30.9**	**9.47**	**14.2**	**25.51**

5.3.3 CMC and Watchlist ROC curve performance

The CMC plots depicting the DIR at various ranks by each method in the experimentation is illustrated on the left-side portion in Figure 5.5, 5.6, 5.7, 5.8, and 5.9. The best approach is expected to achieve 100% identification rate at the earliest, i.e., at lower ranks. Hence, the CMC plots showed rank-10 performance of each method under consideration. The proposed method reached 100% DIR at rank-3 on each partial dataset. The lowest average rank-10 DIR, i.e., less than 20%, was delivered by the RSF, CNNAI, and SURF-based methods on most datasets. However, VeriFinger and NBIS achieved an average rank-10 identification rate of 28.08% and 40.88%, respectively.

The Watchlist ROC plots depicting variation in DIR versus FAR on each dataset by every method during the experimentation are illustrated on the right-side portion in Figure 5.5, 5.6, 5.7, 5.8, and 5.9. A method appropriate for practical implementation shows the curve lying at the top-left position in the plots [97]. It is apparent from the plots that the curve for the proposed method occupied the top-left region in every dataset. The plots illustrated that the DIR for every method except the NBIS and the proposed decreased significantly and reached less than 10% on most datasets. However, a minor deviation on both the parameters, 0.324% average change in DIR and 0.2% average change in FAR, was shown by the VeriFinger.

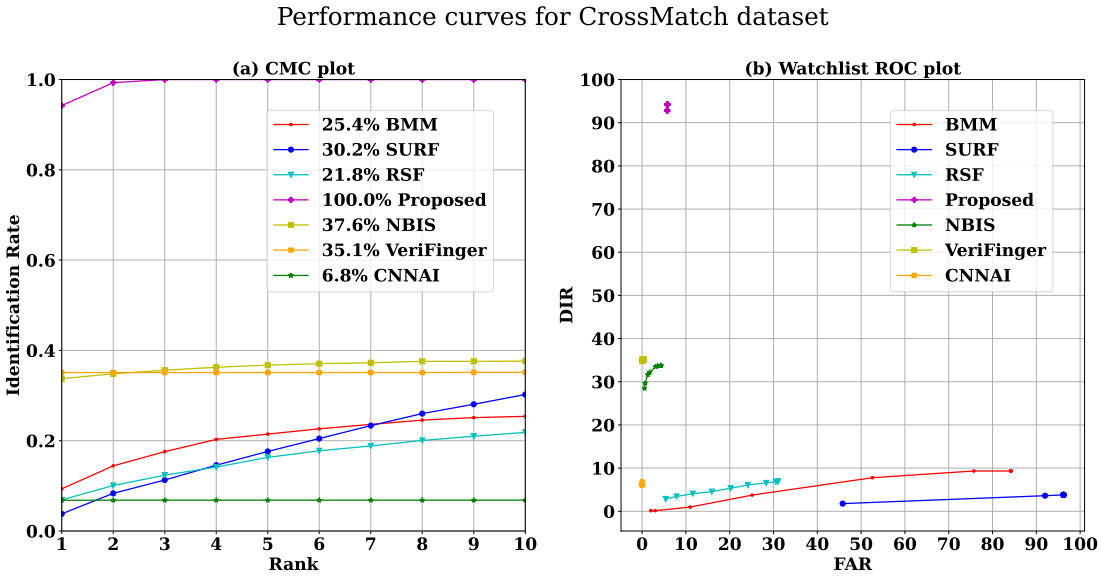


Figure 5.5: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of CrossMatch Sample DB partial dataset.

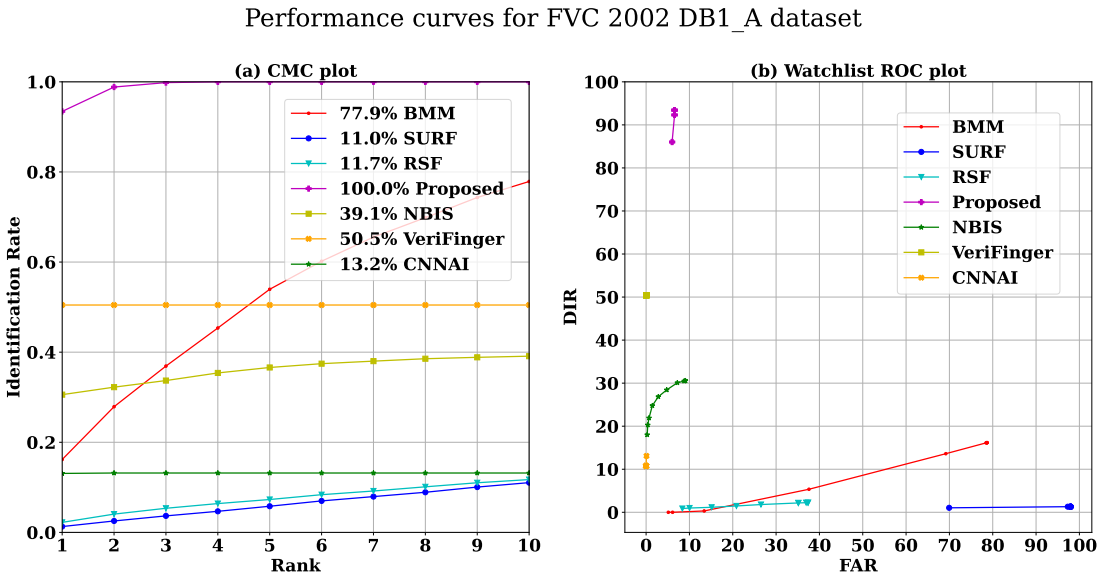


Figure 5.6: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of FVC2002 DB1_A partial dataset.

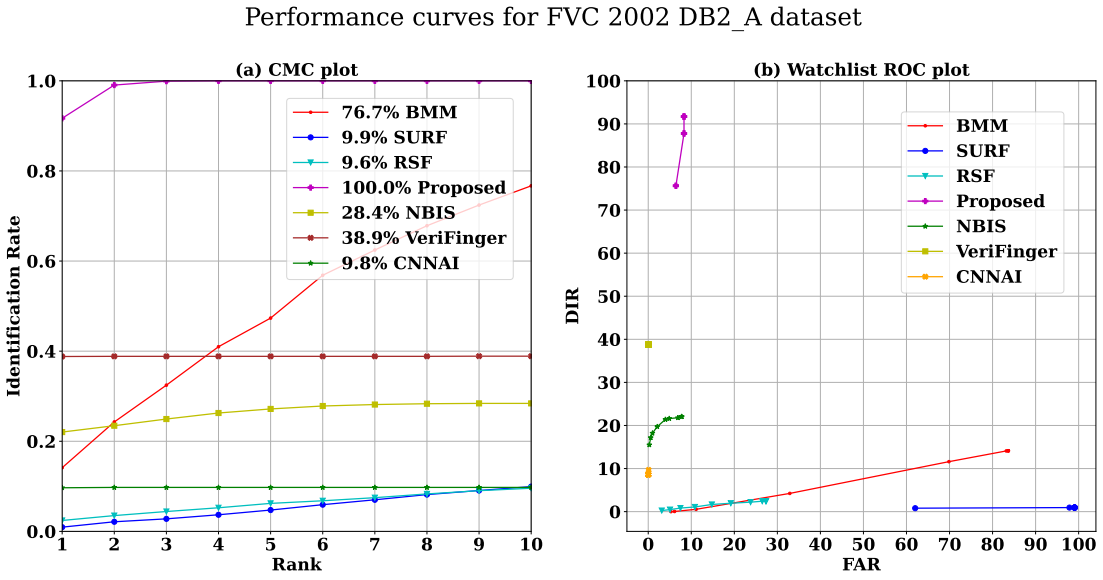


Figure 5.7: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of FVC2002 DB2_A partial dataset.

5.4 Discussion

This section provides a discussion over the performance of each approach on various parameters of the experiments, DIR, FAR, percentage of MasterPrints generated, and number of subjects identified by a MasterPrint. The results analysis from the previous section showed

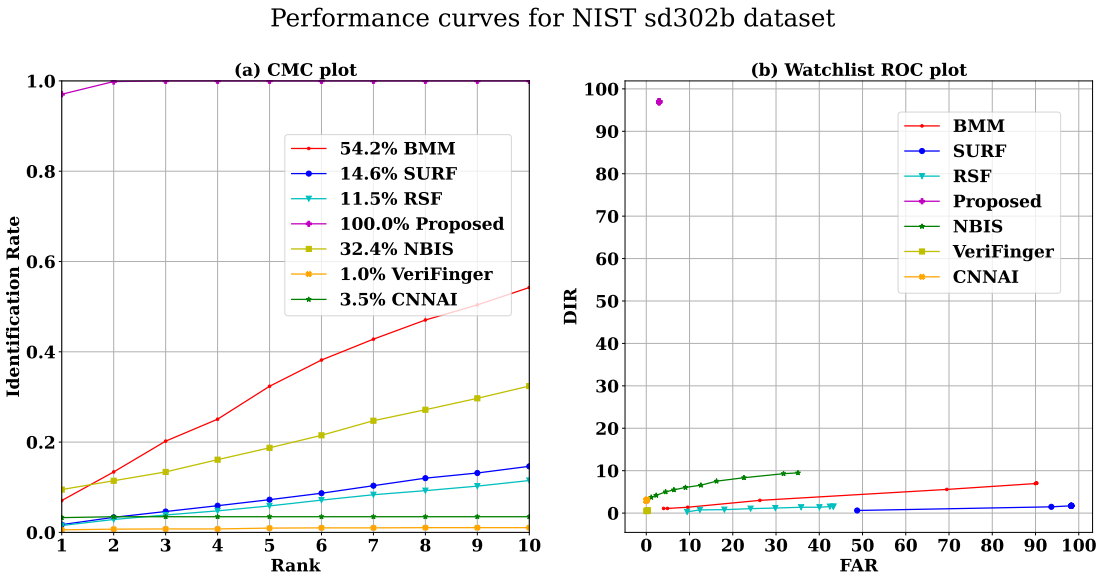


Figure 5.8: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of NIST sd302b partial dataset.

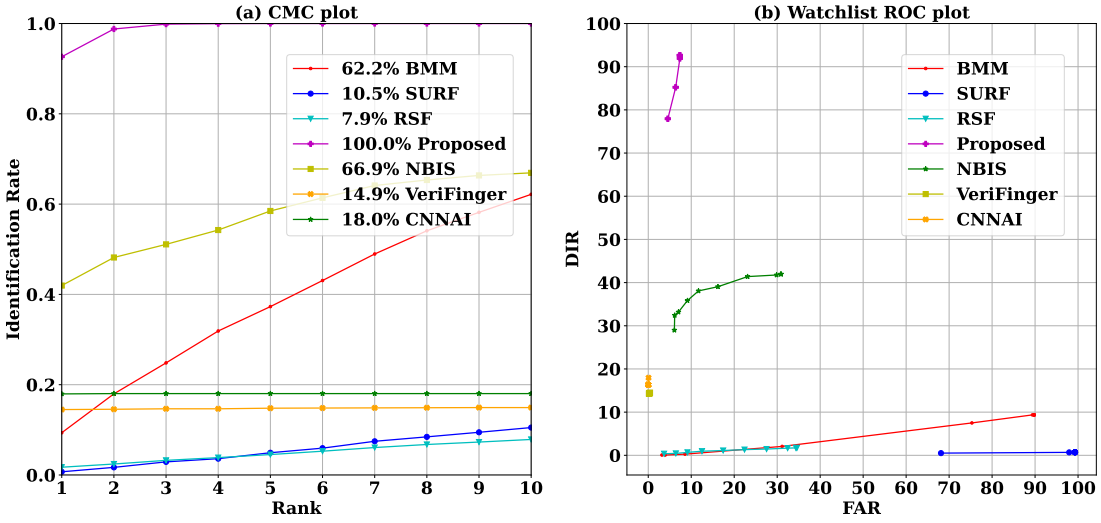


Figure 5.9: CMC curve for rank-10 identification and Watchlist ROC for the DIR and FAR performance of NIST sd302d partial dataset.

that the proposed method produced an average rank-10 DIR of 100%. Hence, it can be be deduced that the proposed method identifies an individual with a partial fingerprint more accurately than other methods. Moreover, it generated a significantly fewer percentage, 0.1% on average, of MasterPrints while identifying the lowest average number of distinct users, i.e., 5, from each dataset. Therefore, the proposed method emerges appropriate for IoT-based

home appliances, smart devices, etc. Subsequently, we justify the high DIR performance of the proposed method through a comparative assessment of the other approaches.

Lee et al. [104] termed the concave and convex ridge patterns as fixed-size *ridge shape features*, *RSFs*. However, due to the small-sized partial fingerprints employed in the experiments, these RSFs were missing in a large portion of each dataset. Therefore, the RSF approach delivered higher RR on every dataset. The approach extracted RSFs over gray-scale partial fingerprints and employed a weighting factor, λ, $(0 \leq \lambda \leq 1)$ during the matching stage. These two steps triggered the generation of large MasterPrints for the method. The proposed approach extracted local features within a small region around a minutia, employed preprocessing steps, and enforced strict matching to overcome the limitations of the RSF method.

NBIS is solely minutiae-based and is considered a robust approach for full fingerprint verification. Its accuracy depends on the number of minutiae detected within a fingerprint. Due to fewer minutiae observed in partial fingerprints, its RR was impacted significantly. However, its marginal FAR demonstrated that it is highly accurate in fingerprint verification. The proposed method extracts an integer-valued feature vector within a small region around a reference minutia which benefited in achieving high identification accuracy compared to NBIS.

VeriFinger SDK requires a large portion of a fingerprint to extract minutiae-based features. Its accuracy reduces if a small number of minutiae are detected in a fingerprint. The SDK rejected a substantial percentage of partial fingerprints during the enrollment phase, raising an exception as "*Extraction failed: BadObject*". Consequently, VeriFinger SDK demonstrated a high RR and low DIR on most datasets. However, the marginal FAR during the identification test and zeroed FAR in the zero MasterPrint detection test reflected its robustness as a successful commercial product in the market.

The preprocessing stage in the proposed approach comprising the binarization and thinning operation ensured that the local features seldom deviate between different samples of a finger. As the feature vectors for a finger rarely vary, the method achieved an improved identification rate. However, the modified SURF-based method failed to accurately extract the essential local feature descriptor at interest points from a partial dataset during the experiments. These local features were similar to other partial fingerprints due to non-uniqueness.

Consequently, the DIR for the modified SURF-based method reduced immensely and eventually produced a large number of MasterPrints.

The proposed method imposed strict feature vector matching to ensure high identification accuracy and lower the chances of MasterPrint generation. The results demonstrated that the BMM approach rejected a small portion of the fingerprints from each dataset. However, the method detected similar features, namely, spatial distance and direction difference, between samples from different subjects in the dataset. Moreover, the BMM method accepts a tolerance while feature matching. As a result, it generates similarity scores for slightly similar fingerprints from distinct subjects. The approach employs minutiae location and orientation as its features, which, if tolerance is accepted, are more likely to match other subjects' partial fingerprints. Hence, it is evident that the primary reason behind the method's poor DIR, high FAR, and huge MasterPrint generation lies in allowing tolerance during the matching stage.

Five full fingerprint datasets were cropped to create the partial datasets employed in the experiments, namely, the standard FVC2002 DB1_A [101], FVC2002 DB2_A [101], CrossMatch Sample DB dataset from NeuroTechnology [102], NIST Special Databases 302 (sd302b and sd302d) [103]. A partial fingerprint participating in the tests must possess a minimum of ten minutiae. The remaining partial fingerprints were discarded, leading to the asymmetrical distribution of samples per subject in each dataset. Consequently, the dataset partitioning for the train-test split was uneven, resulting in a higher rejection rate. Moreover, CNN has a drawback in that it is not invariant to rotation and scale [142]. Hence, due to changes in fingerprint rotation in the partial fingerprint datasets, the CNNAI delivered low DIR.

For training the CNNAI, around 70% of partial fingerprints were employed, and the remaining partial fingerprints were tested for matching. A partial fingerprint possesses low minutiae count. However, MINU-EXTRACTNET couldn't detect all of them [129]. The CNNAI model uses the seven nearest minutiae to construct triangular structures designated as minutiae arrangements (MA). Hence, inaccurate minutiae detection impacted the MA-based feature extraction in the CNNAI model, and led to low DIR on each partial fingerprint dataset.

5.5 Summary of the chapter

The work in this chapter introduced a minutiae geometry-based partial fingerprint identification method to alleviate the MasterPrint generation. The experiments comprised partial fingerprint datasets cropped from five full fingerprint benchmark datasets, and the results were compared with six existing approaches. The result analysis showed that the proposed approach achieved up to 97% identification accuracy while generating about 0.1% MasterPrints. Thus, it can be observed that the local minutiae feature-based partial fingerprint identification method enforcing strict feature matching can mitigate MasterPrint vulnerability while delivering high identification accuracy.

Chapter 6

Investigating the Impact of Preprocessing Approaches on the Performance of Partial Fingerprint Identification Systems

The fingerprint sensors often capture a distorted fingerprint image due to variations in pressure exerted by the user, sweat on the fingertip, moisture in the surroundings, etc. Hence, the quality of the sensed fingerprint is enhanced by employing thresholding and thinning methods to facilitate minutiae extraction. It is not always satisfactory to enhance an image and achieve better results as sometimes undesirable data may be introduced or critical information may be lost by using these methods. Moreover, an improper binding of these methods can adversely introduce false minutiae leading to false matches and a low identification rate. Furthermore, the system may be susceptible to MasterPrint generation. In this regard, this chapter is devoted to investigating the impact of four well-known thresholding and thinning approaches exercised in Joshi et al. [143] method, discussed in Chapter 5. This work employs the partial fingerprint datasets used in the original work to study the impact on identification accuracy and MasterPrint generation.

The contents of this chapter are organised as follows. Initially, Section 6.1 discusses the various thresholding approaches employed in the investigation. Section 6.2 presents

various thinning methods used in the experiments. The experimental setup are presented in Section 6.3. Further, the evaluation metrics and result analysis is given under Section 6.4. The discussion and the significance of the experimental results are provided in Section 6.5. Finally, Section 6.6 summarises the chapter.

6.1 Thresholding approaches

Image preprocessing techniques improve an image's quality thereby preserving its original contents [144]. However, it is not always useful to enhance an image as, more often, crucial information may get lost due to such techniques. There is an infinitesimally small probability of acquiring exactly similar fingerprints by the biometric system each time a user touches the biometric sensor. The weather conditions such as, moisture or heat, sweating around the fingertip, the orientation of the finger, and pressure exerted on the sensor surface are some of the reasons leading to dissimilar fingerprint samples of the same finger acquired at different instances. Therefore, thresholding or binarization approaches are generally employed to produce approximately similar ridge patterns from several samples of the same finger. These techniques are broadly categorised as global, local, and hybrid methods [145]. The fingerprint biometric researchers and vendors use these techniques for research activities and consumer products. However, every method may not be beneficial for a biometric application and the strategies adopted may adversely affect the system accuracy by generating false minutiae. Consequently, it results in an incorrect and low identification rates and high MasterPrint generation. Therefore, it is imperative to investigate the impact of various combinations of preprocessing techniques and examine the system performance.

Shaikh et al. [146] evaluated six binarization methods for performance bench-marking of various global and local binarization methods towards fingerprint-based biometric recognition. Their work forms the basis for selecting the thresholding approaches employed in our experimentation. The thresholding algorithms used during the experiments include iterative optimal thresholding [144], Otsu's global image thresholding [147], Niblack local thresholding [148], and Bernsen local image thresholding [149]. The following subsection briefly discusses these methods.

6.1.1 Iterative optimal thresholding

The iterative optimal thresholding approach models the image pixels as a histogram generating normal distributions for the area of interest, i.e., the ridge portion and the background region, also known as the valley portion [150]. The approach considers the minimum probability lying between the maxima of two distributions as the initial threshold, T_i. Further, the method iteratively updates T_i to minimise the segmentation error [144]. The algorithm considers the value of T_i for which the segmentation error cannot be further minimised as the optimal threshold, T_o.

6.1.2 Otsu's global image thresholding

The approach returns an intensity as a threshold to divide the image pixels as background and foreground. The algorithm iteratively tries to maximise the inter-class intensity variance or minimise the intra-class intensity variance [147]. It computes histogram, $H(i)$, with L bins and probability, $p(i)$, for each intensity, i, within the image. Let t be the threshold under consideration. The probability of a pixel to be a background pixel, W_b, and foreground pixel, W_f, is computed as below,

$$W_b(t) = \sum_{i=0}^{t-1} p(i) \tag{6.1}$$

$$W_f(t) = \sum_{i=t}^{L-1} p(i) \tag{6.2}$$

The approach then calculates intra-class variance σ_w as

$$\sigma_w^2(t) = W_b(t) \times \sigma_b(t) + W_f(t) \times \sigma_f(t) \tag{6.3}$$

where, $\sigma_b(t)$ and $\sigma_f(t)$ are the background and foreground gray level variances, respectively. It returns the threshold corresponding to $min(\sigma_w^2(t))$ as the desired threshold.

6.1.3 Niblack local thresholding

Niblack algorithm is a local thresholding approach [148]. It uses a fixed-sized rectangular window, w, surrounding a reference pixel, p, and slides the window over the entire image, I. The window size is application dependent and default value is 15. The approach computes the local mean, μ_w, and standard deviation, σ_w, for the window region [151]. The following equation decides the local threshold, T_w, for the given window, w.

$$T_w = \mu_w + (-0.2) \times \sigma_w \tag{6.4}$$

Khurshid et al. [152] showed that the approach generates binarization noise in the non-desired gray region.

6.1.4 Bernsen's local image thresholding

Bernsen's approach is another local thresholding method [149]. For a given image, I, the approach initialises local contrast, l, and neighbourhood window size, w, for instance, $l = 15$ and $w = 3$. The algorithm then assigns the lowest and highest gray levels within the window size $w \times w$ as I_{min} and I_{max}, respectively. The local threshold, Th_l, and the contrast measure, C_m, are computed using the following equations [149],

$$Th_l = \frac{I_{max} + I_{min}}{2} \tag{6.5}$$

$$C_m = I_{max} - I_{min} \tag{6.6}$$

If $C_m > l$, i.e., a non-uniform gray scale image, then the neighbourhood belongs to the same class, background or foreground. Otherwise, a global thresholding approach decides the local threshold [153].

6.2 Thinning approaches

A thinning algorithm produces a single-pixel skeletal structure that highlights prominent features from the original image. In general, a binarized image is employed for the thinning process to ensure connectivity among various regions within the image. It helps in determining the topological and metric-based properties to count, measure, and classify relevant features. However, local noises in the image easily affect the resultant skeleton [154]. Thinning algorithms are mainly used for object representation, detection, manipulation, comparison, tracking, recognition, and compression.

Minutiae are the most widely exercised and accepted features utilised in fingerprint biometric systems [155]. In general, minutiae points are stored as their (x, y) coordinates, orientation angle, and type, i.e., ridge or bifurcation. Minutiae-based fingerprint biometric systems employ minutiae correlations within an image during their comparison. Therefore, locating minutiae most accurately within two samples of the same finger is highly desirable. A robust thinning approach accepting a correctly binarized fingerprint image can improve system performance in such circumstances. However, a given thinning approach may also adversely reduce the recognition accuracy if it detects substantial false minutiae or misses a large number of true minutiae.

Nazarkevych et al. [156] evaluated the effectiveness of image thinning methods in biometric security systems. The authors analysed Zhang-Suen [157] and Hilditch thinning algorithm [158], among others. Saha et al. [159] presented a comprehensive review of existing thinning methods and their applications. The authors discussed thinning approaches applicable to fingerprint analysis. The work by Nazarkevych et al. [156] and Saha et al. [159] encouraged us to experiment with Hilditch thinning algorithm [158] and Stentiford thinning method [160]. The designers of Saeed et al. [161] algorithm and its modified version by Tabedzki et al. [162] have claimed these approaches as a universal algorithm for image thin-

ning. Hence, these two methods were used to verify the claim for their robustness in partial fingerprint identification. The following subsections briefly explain the thinning methods used in the investigation.

6.2.1 KMM thinning algorithm

The KMM (Khalid, Mariusz, Marek) approach accepts a binarized image wherein binary 1 represents the dark region to be thinned. Next, it converts the 1's adjacent to the boundary 0's and in the open elbow bends into 2 and 3, respectively [163]. The method considers non-zero positions in the image and figures out the locations, x, having 2, 3, or 4 sticking neighbours. It changes all such x to 4. A predefined table, *Deletion Array*, provides the sum for x that is the probable target for removal. It iteratively eliminates such x, assuring that the connectivity is intact. Finally, the approach excludes unnecessary 2's and 3's until it produces a single-pixel width thinned image [164].

6.2.2 K3M thinning algorithm

The K3M (Khalid, Marek, Mariusz, Marcin) algorithm is a modified version of KMM [161]. The algorithm iterates over seven phases until it generates a thinned image. These phases can be summarised as below,

1. Mark boundry pixels, b.
2. Remove b's with 3 adjacent neighbours.
3. Remove b's having 3 or 4 adjacent neighbours.
4. Remove b's with 3, 4, or 5 adjacent neighbours.
5. Remove b's with 3, 4, 5, or 6 adjacent neighbours.
6. Remove b's with 3, 4, 5, 6, or 7 adjacent neighbours.
7. Unmark remaining boundary pixels.

If the current iteration of these seven phases modifies the image, the image undergoes another iteration over the image comprising the above seven steps. Otherwise, the algorithm stops resulting in a thinned image [162].

6.2.3 Hilditch thinning algorithm

The Hilditch thinning algorithm has two variants; one uses a 3×3 window while the other uses 4×4 window [158]. We have employed the version using the 3×3 window and will discuss the same with reference to the neighbourhood pixel nomenclature shown in Figure 6.1. The algorithm iteratively decides if a pixel P_1 should be removed based on the following five conditions [165],

1. Eliminate P_1, if it is a part of the skeleton.

2. Preserve P_1, if it lies on the border of a skeleton.

3. Preserve P_1, if it is an isolated pixel.

4. If P_1 is a connecting pixel, preserve it.

5. Remove P_1, if it has at least one neighbour.

The algorithm considers all the above conditions to decide if P_1 should be preserved or eliminated. It finally stops when the recent iteration encounters no pixels for removal.

P_9	P_2	P_3
P_8	P_1	P_4
P_7	P_6	P_5

Figure 6.1: Neighbourhood pixel nomenclature in Hilditch approach.

6.2.4 Stentiford thinning algorithm

The templates used in the Stentiford method to decide the pixels for removal is shown in Figure 6.2. It considers only three locations in a reference pixel neighbourhood, marked in circles. The steps followed in the algorithm are as below [160],

1. Traverse an image left-to-right downwards and locate pixels possessing T_1 pattern,
2. If the central pixel at such location is not an endpoint, i.e., last pixel, and have *connectivity value* [166] as 1, mark it for elimination,
3. Repeat step 1 and 2 for each pixel in the image,
4. Repeat steps 1, 2 and 3 for pattern T_2 while traversing upwards left-to-right, for T_3 traversing right-to-left upwards, and for T_4 traversing downwards right-to-left,
5. Remove the marked pixels,
6. If any pixel was removed in step 4 of the recent iteration, repeat steps 1 to 5 else stop.

6.3 Experimental setup

Joshi et al. [143] proposed a minutiae geometry-based partial fingerprint identification approach targeted towards MasterPrint mitigation. The method experimented on partial fingerprint datasets cropped from five benchmark full fingerprint datasets delivered up to 97%

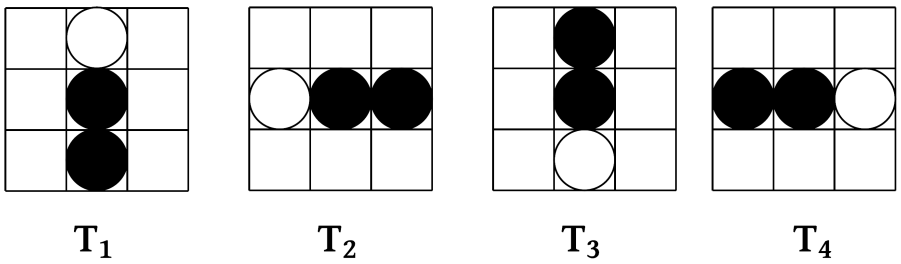


Figure 6.2: The templates for deciding pixels for removal in Stentiford approach. It considers only three locations, marked in circle, for thinning operation.

accuracy and generated 0.1% MasterPrints. The authors employed adaptive thresholding approach [140] for binarization and Zhang–Suen thinning algorithm [157] to thin the binarized partial fingerprints. Minutiae detection was carried out by employing the metric, *Crossing Number* [131]. However, inappropriately binarized or thinned fingerprint may generate false minutiae affecting the system performance. Kim et al. [167] algorithm detects and removes such false minutiae in our experimentation. It performs post-processing on the detected minutiae using various parameters, such as, ridge flow, ridge orientation, connectivity, and distance between minutiae. The algorithm detects and eliminates five different types of false minutiae, namely broken ridge, bridge, short ridge, hole, and triangle [167].

The investigation in this chapter aims to study the impact of various combinations of thresholding and thinning methods in Joshi et al. [143] method towards identification accuracy and percentage of MasterPrint generated. The original papers on these thresholding and thinning methods have shown satisfactory results. However, the impact of their cross combinations has not yet been reported. This work does not attempt to comment on a particular method or ascertain that a specific pair is preferable. Instead, this work evaluates the robustness of the Joshi et al. [143] method under diverse preprocessing conditions. The experiments were carried out using Joshi et al. [143] method on the partial datasets used in the original work. However, instead of adaptive thresholding approach [140] for binarization and Zhang–Suen thinning algorithm [157] for thinning, 16 combinations of the selected thresholding and thinning methods were applied. So, there were 80 individual experiments as a part of this investigation.

The experiments were conducted on a desktop system with 64-bit Ubuntu $20.04.2$ LTS (Focal Fossa) operating system having 64 GB internal memory (RAM) and Intel® Xeon(R) CPU E5-1620 v3 @ 3.50 GHz $\times 8$ processor. The terminology used for various combinations of thresholding and thinning approaches on different datasets is shown in Table 6.1. We have followed *D_T_S* format for each combination, where *D* refers to the dataset, *T* specifies a thresholding approach, and *S* denotes the thinning approach. The values 1, 2, 3, 4, and 5 for fingerprint dataset indicates CrossMatch Sample DB dataset, FVC 2002 DB1_A dataset, FVC 2002 DB2_A dataset, NIST sd302b dataset, and NIST sd302d dataset, respectively. The values 1, 2, 3, and 4 for thresholding approach denotes iterative optimal thresholding, Otsu's method, Niblack local thresholding, and Bernsen's local image thresholding, respec-

tively. Further, the values 1, 2, 3, and 4 for thinning approach represents KMM thinning algorithm, K3M thinning algorithm, Hilditch thinning algorithm, and Stentiford thinning algorithm, respectively. For example, entry 1_1_1 refers to the combination of iterative optimal thresholding and KMM thinning algorithm experimented on CrossMatch Sample DB dataset. Figure 6.3 shows a sample image from CrossMatch Sample DB dataset and its thinned version from each combination of thresholding and thinning approach as mentioned in Table 6.1. The thinned image for the same fingerprint generated from different combinations appears significantly diverse. This variation may produce considerable differences in the identification performance of Joshi et al. [143] method.

Table 6.1: Nomenclature for various combinations of thresholding and thinning approaches on different datasets. SK_1 - KMM thinning algorithm, SK_2 - K3M thinning approach, SK_3 - Hilditch thinning algorithm, SK_4 - Stentiford thinning algorithm.

Thresholding approach	Thinning approach	CrossMatch Sample DB	FVC 2002 DB1_A	FVC 2002 DB2_A	NIST sd302b	NIST sd302d
Iterative optimal thresholding	SK_1	1_1_1	2_1_1	3_1_1	4_1_1	5_1_1
	SK_2	1_1_2	2_1_2	3_1_2	4_1_2	5_1_2
	SK_3	1_1_3	2_1_3	3_1_3	4_1_3	5_1_3
	SK_4	1_1_4	2_1_4	3_1_4	4_1_4	5_1_4
Otsu's method	SK_1	1_2_1	2_2_1	3_2_1	4_2_1	5_2_1
	SK_2	1_2_2	2_2_2	3_2_2	4_2_2	5_2_2
	SK_3	1_2_3	2_2_3	3_2_3	4_2_3	5_2_3
	SK_4	1_2_4	2_2_4	3_2_4	4_2_4	5_2_4
Niblack local thresholding	SK_1	1_3_1	2_3_1	3_3_1	4_3_1	5_3_1
	SK_2	1_3_2	2_3_2	3_3_2	4_3_2	5_3_2
	SK_3	1_3_3	2_3_3	3_3_3	4_3_3	5_3_3
	SK_4	1_3_4	2_3_4	3_3_4	4_3_4	5_3_4
Bernsen's local image thresholding	SK_1	1_4_1	2_4_1	3_4_1	4_4_1	5_4_1
	SK_2	1_4_2	2_4_2	3_4_2	4_4_2	5_4_2
	SK_3	1_4_3	2_4_3	3_4_3	4_4_3	5_4_3
	SK_4	1_4_4	2_4_4	3_4_4	4_4_4	5_4_4

(a) Sample image

(a1) 1_1_1

(a2) 1_1_2

(a3) 1_1_3

(a4) 1_1_4

(a5) 1_2_1

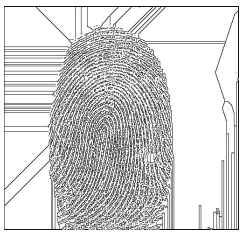
(a6) 1_2_2

(a7) 1_2_3

(a8) 1_2_4

(a9) 1_3_1

(a10) 1_3_2



(a11) 1_3_3

(a12) 1_3_4

(a13) 1_4_1

(a14) 1_4_2

(a15) 1_4_3

(a16) 1_4_4

Figure 6.3: A sample fingerprint image from the CrossMatch Sample DB dataset and corresponding thinned images generated from various combinations of thresholding and thinning approaches.

6.4 Evaluation metrics and result analysis

The MasterPrint vulnerability is a threat to an identification system. Hence, the investigation in this work followed both closed-set and open-set identification set-up. The investigation involved two tests: an identification test [168] and a zero MasterPrint detection test. During the identification test, each template was compared with every other template from the dataset, and the similarity score for each comparison was computed. If the highest score corresponds to the actual subject sample, it was quoted as a correct detect and identify (CDI). In a false alarm (FA) scenario, the highest score belongs to some other subjects' templates. The system may reject a partial fingerprint due to no similarity with any stored templates. Suppose the system is enrolled with $\mathcal{P}$ partial fingerprints, and $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{R}$ denotes the count of CDI, FA, and rejected partial fingerprints, respectively. The detect and identification rate (DIR), δ, false alarm rate (FAR), $\digamma$, and rejection rate (RR), Υ, are computed as : $\delta = \frac{\mathcal{C}}{\mathcal{P}} \times 100$, $\digamma = \frac{\mathcal{F}}{\mathcal{P}} \times 100$, and $\Upsilon = \frac{\mathcal{R}}{\mathcal{P}} \times 100$.

An identification system producing lowest MasterPrints at higher DIR and lower FAR would become ideal for practical use in the fingerprint biometric system. The identification test results on each dataset involve computing the DIR, FAR, RR, and the percentage of MasterPrints generated without setting a predefined threshold. A Cumulative Matching Characteristic (CMC) curve shows the rank-k performance of an identification system, depicting the identification of the correct subject at different ranks [169]. The results from the identification test make up the data for the CMC curve. Suppose we have k subjects enrolled with a system. Ideally, the rank-k identification rate should be 100%. The best approach is expected to reach 100% performance at the earliest. Hence, the CMC plots presented here showed the DIR performance till rank-10.

For each combination that produced MasterPrints in the identification test, a zero MasterPrint detection test was conducted. In this test, the system threshold was raised gradually until no MasterPrints are observed. Suppose, τ is the threshold at which no MasterPrints were observed. The DIR, δ_0, at τ is calculated using the formula for δ. A good approach should show marginal variation between δ and δ_0. Subsequently, δ_0 is divided into three intermediate thresholds to compute DIR and FAR at each of these thresholds. The DIR and

FAR at δ, δ_0, and the three intermediate thresholds provides the data to plot the Watchlist Receiver Operating Characteristic (ROC) curve for each dataset. The curve occupying top-left region in the Watchlist ROC plot is considered robust as it shows slight variation in DIR and significant reduction in FAR as the system threshold is increased to accept highly similar partial fingerprints.

6.4.1 Identification and zero MasterPrint detection test results

The partial fingerprint identification method with different combinations during the preprocessing stage was evaluated for DIR (δ), FAR (F), DIR while generating zero MasterPrints (δ_0), RR (Υ), and the percentage of MasterPrints generated. The results of each combination of thresholding and thinning approach for the identification and zero MasterPrint detection test on CrossMatch Sample DB dataset is presented in Table 6.2. The highest DIR observed was 92.65% by 1_1_3 while producing nearly 10% MasterPrints. The combination 1_2_3 generated more than 21% MasterPrints. The results of each combination of thresholding and thinning approach for the identification and zero MasterPrint detection test on FVC2002 DB1_A dataset is presented in Table 6.3. Only two combinations, namely, 2_4_2 and 2_4_3 could achieve more than 90% DIR. However, 2_4_2 generated the maximum percentage of MasterPrints during the experimentation. The results of each combination of thresholding and thinning approach for the identification and zero MasterPrint detection test on FVC2002 DB2_A dataset is presented in Table 6.4. The combination 3_4_1 delivered more than 93% DIR, but produced above 16% MasterPrints. The lowest percentage of MasterPrints during the experiments was around 8% by the combination 3_1_4. The results of each combination of thresholding and thinning approach for the identification and zero MasterPrint detection test on NIST sd302b dataset is presented in Table 6.5. Here, three combinations, namely, 4_1_1, 4_2_1, and 4_4_2 achieved above 90% DIR. However, 4_4_2 generated above 18% MasterPrints. The results of each combination of thresholding and thinning approach for the identification and zero MasterPrint detection test on NIST sd302d dataset is presented in Table 6.6. The highest DIR observed here was 92.38% by 5_1_2 while producing 16.81% MasterPrints.

Table 6.2: Results on CrossMatch Sample DB dataset. δ_0- DIR in zero MasterPrint generation test, MP- MasterPrints generated in identification test.

Binding approach	δ	F	δ_0	MP	Υ
1_1_1 (%)	84.8	6.6	62.21	4.8	8.6
1_1_2 (%)	77.9	7.7	61.8	7.94	14.4
1_1_3 (%)	92.65	3.8	75.65	9.65	3.55
1_1_4 (%)	90.12	3.86	78.4	9.12	6.02
1_2_1 (%)	80.39	8.26	62.32	8.39	11.35
1_2_2 (%)	81.15	5.34	60.84	15.15	13.51
1_2_3 (%)	91.52	5.7	73.29	21.12	2.78
1_2_4 (%)	91.89	2.45	75.62	12.89	5.66
1_3_1 (%)	79.95	7.95	60.4	9.95	12.1
1_3_2 (%)	70.29	9.41	55.21	6.29	20.3
1_3_3 (%)	84.31	6.34	62.87	8.31	9.35
1_3_4 (%)	80.88	4.46	65.48	8.89	14.66
1_4_1 (%)	87.25	4.55	60.98	17.25	8.2
1_4_2 (%)	68.8	7.12	55.43	18.87	24.08
1_4_3 (%)	89.69	2.61	72.64	9.69	7.7
1_4_4 (%)	92.16	3.06	74.7	9.16	4.78

Table 6.3: Results on FVC2002 DB1_A dataset. δ_0- DIR in zero MasterPrint generation test, MP- MasterPrints generated in identification test.

Binding approach	δ	F	δ_0	MP	Υ
2_1_1 (%)	85.2	4.18	68.7	8.2	10.62
2_1_2 (%)	82.06	6.84	69.57	12.06	11.1
2_1_3 (%)	79.35	2.99	60.58	7.35	17.66
2_1_4 (%)	85.88	5.48	62.74	15.88	8.64
2_2_1 (%)	69.61	2.29	55.49	17.61	28.1
2_2_2 (%)	79.85	5.27	62.74	19.85	14.88
2_2_3 (%)	75.88	3.48	60.8	15.88	20.64
2_2_4 (%)	77.11	8.87	65.71	17.11	14.02
2_3_1 (%)	68.05	3.27	55.96	6.05	28.68
2_3_2 (%)	89.71	4.54	75.21	9.71	5.75
2_3_3 (%)	78.69	8.93	65.84	5.69	12.38
2_3_4 (%)	89.12	7.13	62.09	17.12	3.75
2_4_1 (%)	82.75	6.85	70.75	12.75	10.4
2_4_2 (%)	91.13	6.14	73.25	21.13	2.73
2_4_3 (%)	90.31	4.54	78.09	12.37	5.15
2_4_4 (%)	79.84	4.16	64.86	7.84	16

The entries from Table 6.2--6.6 demonstrated that the DIR ranges between $62\% - 93\%$. But the average DIR in the original approach was 93.8%. Thus, the DIR performance of the

Table 6.4: Results on FVC2002 DB2_A dataset. δ_0- DIR in zero MasterPrint generation test, MP- MasterPrints generated in identification test.

Binding approach	δ	F	δ_0	MP	Υ
3_1_1 (%)	81.26	4.06	75.21	9.26	14.68
3_1_2 (%)	81.8	7.41	78.95	10.31	10.79
3_1_3 (%)	85.29	7.21	74	8.29	7.5
3_1_4 (%)	85.05	7.28	62.54	8.05	7.67
3_2_1 (%)	88.53	2.73	60.27	11.53	8.74
3_2_2 (%)	67.33	5.22	51.06	12.5	27.45
3_2_3 (%)	85.05	7.29	71.98	15.05	7.66
3_2_4 (%)	84.31	6.53	62.7	14.31	9.16
3_3_1 (%)	92.16	3.2	80.29	12.16	4.64
3_3_2 (%)	91.79	5.54	76.32	17.79	2.67
3_3_3 (%)	70.51	7.18	62.98	19.51	22.31
3_3_4 (%)	79.26	8.95	59.21	9.26	11.79
3_4_1 (%)	93.32	3.94	80.9	16.32	2.75
3_4_2 (%)	79.51	7.19	60.77	19.51	13.3
3_4_3 (%)	91.67	4.81	74	11.67	3.52
3_4_4 (%)	89.3	6.03	65.74	9.3	4.67

Table 6.5: Results on NIST sd302b dataset. δ_0- DIR in zero MasterPrint generation test, MP- MasterPrints generated in identification test.

Binding approach	δ	F	δ_0	MP	Υ
4_1_1 (%)	92.85	3.37	74.65	4.85	3.78
4_1_2 (%)	85.75	3.65	60.9	9.75	10.6
4_1_3 (%)	65.2	4.07	52.49	15.2	30.73
4_1_4 (%)	62.01	2.94	49.65	6.01	35.05
4_2_1 (%)	91.1	5.11	79.58	9.1	3.79
4_2_2 (%)	81.75	6.61	67.25	13.75	11.64
4_2_3 (%)	62.01	1.83	50.2	12.01	36.16
4_2_4 (%)	62.25	3.08	52.36	12.25	34.67
4_3_1 (%)	74.57	7.41	56.8	11.57	18.02
4_3_2 (%)	74.02	8.53	65.21	22.02	17.45
4_3_3 (%)	84.8	6.88	72.95	8.8	8.32
4_3_4 (%)	88.97	6.33	76.32	18.97	4.7
4_4_1 (%)	81.52	8.97	74.68	15.53	9.51
4_4_2 (%)	90.28	6.29	78.98	18.28	3.43
4_4_3 (%)	78.19	6.83	62.4	17.19	14.98
4_4_4 (%)	83.33	6.26	60.35	13.39	10.41

original method has been reduced by more than 11% on average due to varying combinations of thresholding and thinning methods. Moreover, the DIR in the zero MasterPrint detection

Table 6.6: Results on NIST sd302d dataset. δ_0- DIR in zero MasterPrint generation test, MP- MasterPrints generated in identification test.

Binding approach	δ	F	δ_0	MP	Υ
5_1_1 (%)	82.17	6.07	70.25	12.11	11.76
5_1_2 (%)	92.38	6.19	79.28	16.81	1.43
5_1_3 (%)	73.82	2.12	62.47	3.82	24.06
5_1_4 (%)	67.91	2.31	51.64	7.96	29.78
5_2_1 (%)	82.35	7.3	69.4	8.35	10.35
5_2_2 (%)	83.06	8.21	67.24	9.06	8.73
5_2_3 (%)	67.01	2.32	50.85	7.01	30.67
5_2_4 (%)	88.58	6.54	69.4	8.97	4.88
5_3_1 (%)	85.78	7.2	60.36	15.78	7.02
5_3_2 (%)	87.09	5.45	65.32	17.05	7.46
5_3_3 (%)	83.63	7.02	59.21	13.63	9.35
5_3_4 (%)	71.96	3.55	57.39	11.96	24.49
5_4_1 (%)	72.55	8.37	60.9	17.55	19.08
5_4_2 (%)	91.1	5.22	78.35	15.12	3.68
5_4_3 (%)	84.17	5.67	68.21	14.85	10.16
5_4_4 (%)	73.43	8.09	64.85	23.45	18.48

test was also lowered by nearly 10% during the investigation. However, the average FAR has decreased by 0.6% compared with Joshi et al. [143] work. The average rejection rate also reduced by 7.16%. The percentage of MasterPrint generated during the investigation ranges between $3.82\% - 23.45\%$, whereas for the original method, the range lies within $0.1\% - 2.03\%$.

The investigation witnessed a 14 times increase in the percentage of MasterPrints generated, while the average percentage of MasterPrints generated for the experiments was more than 11% compared to the original paper. These statistics thus demonstrated that the accuracy and MasterPrint mitigation performance of the Joshi et al. [143] method significantly reduce while using different preprocessing schemes. The entries for DIR (δ), FAR (F), DIR while generating zero MasterPrint (δ_0), RR (Υ), and the percentage of MasterPrints generated during the investigations ranges between $62.01 - 93.32\%$, $1.83 - 9.41\%$, $50.2 - 80.9\%$, $1.43 - 36.16\%$, and $3.82 - 23.45\%$, respectively. Thus, the results from Table 6.2--6.6 for identification and zero MasterPrint detection test appear diversely distributed for each of the five parameters under consideration. Furthermore, no binding approach have shown remarkable results invariably on several datasets. Hence, these performance measures do not form

a concrete base to attribute any pair, specific thresholding or thinning method as preferable over others.

6.4.2 CMC and Watchlist ROC curve performance

The CMC and Watchlist ROC curves for each preprocessing combination on individual datasets are depicted on the left and right portion of Figure 6.4--6.8, respectively. The highest and average rank-10 DIR achieved during the investigation was 98.6% and 87.35%, respectively. However, the original method achieved 100% DIR on each dataset till rank-2. It also demonstrated that the DIR did not improve beyond rank-3. In the case of Watchlist ROC plots, ideally, the curves are expected to deviate marginally for DIR and show significant variations on FAR. But the plots from Figure 6.4 to Figure 6.8 demonstrated that the identification rate reduced considerably compared to FAR. Moreover, the average DIR in zero MasterPrint generation test, δ_0, has reduced from 76.14% in the original work to 65.96% during the investigations. Thus, delivered a degraded accuracy of more than 10%. The plots thus showed that the average identification accuracy was reduced by more than 12% when diverse combinations of pre-processing methods were employed in Joshi et al. [143]

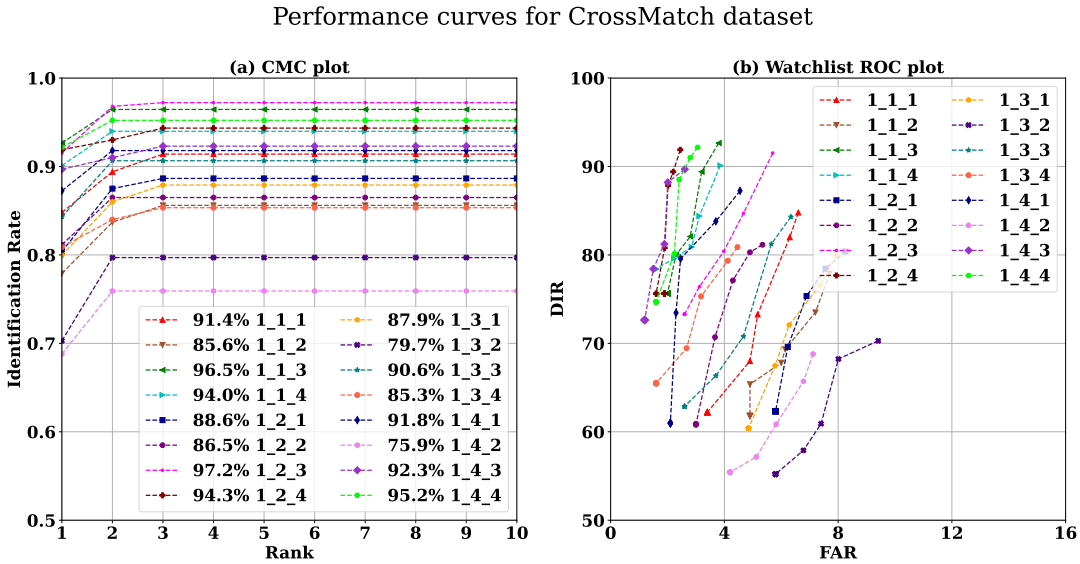


Figure 6.4: Cumulative Matching Characteristic (CMC) curves and Watchlist ROC plots for CrossMatch Sample DB dataset.

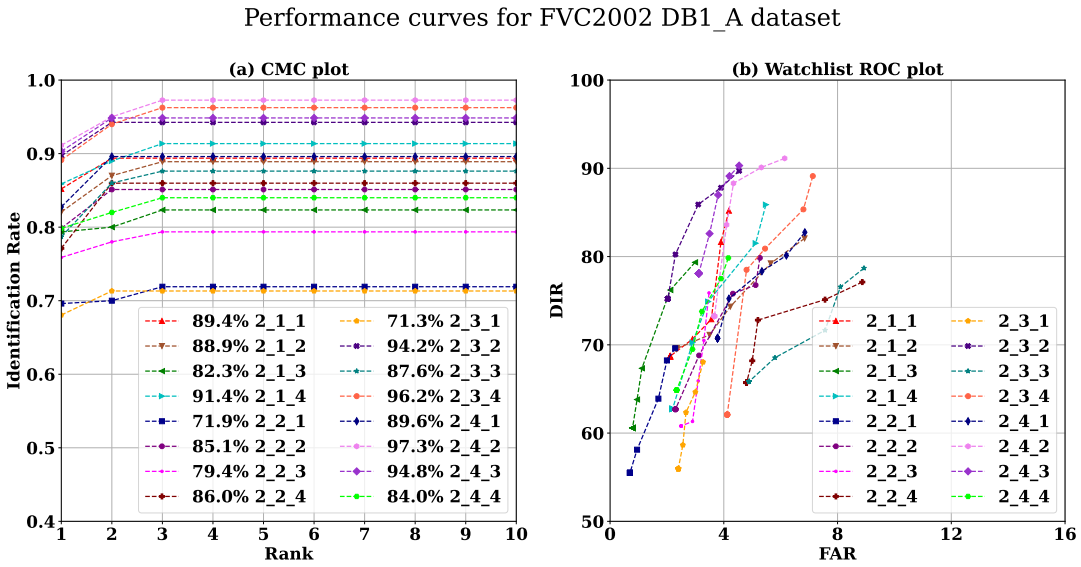


Figure 6.5: Cumulative Matching Characteristic (CMC) curves and Watchlist ROC plots for FVC2002 DB1_A dataset.

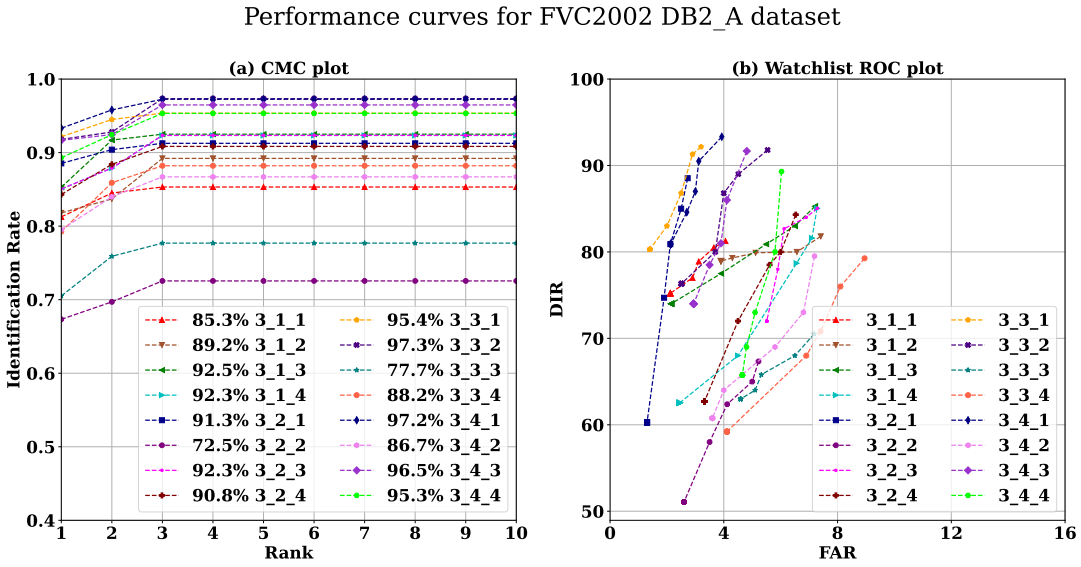


Figure 6.6: Cumulative Matching Characteristic (CMC) curves and Watchlist ROC plots for FVC2002 DB2_A dataset.

work. However, it is not possible to label any single thresholding or thinning method, or their specific combination as best or worst due to their inconsistent performance in CMC and Watchlist ROC plots.

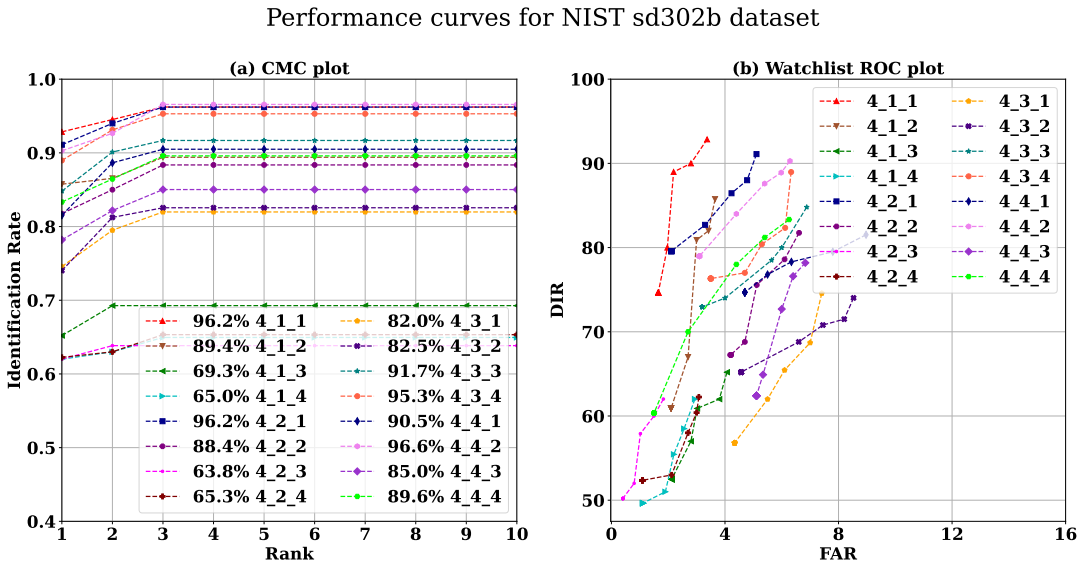


Figure 6.7: Cumulative Matching Characteristic (CMC) curves and Watchlist ROC plots for NIST sd302b dataset.

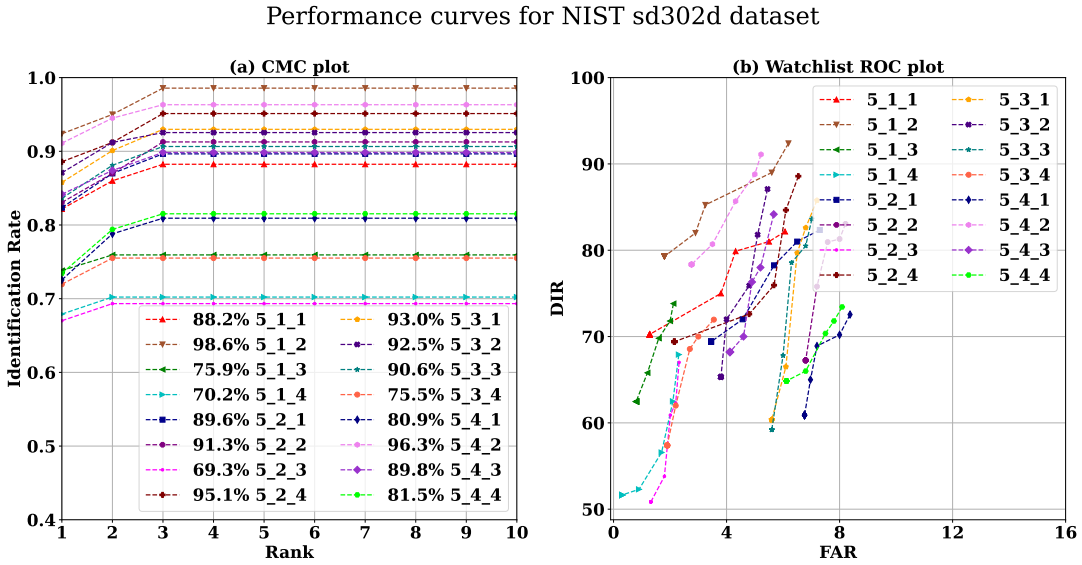


Figure 6.8: Cumulative Matching Characteristic (CMC) curves and Watchlist ROC plots for NIST sd302d dataset.

6.5 Discussion

The investigation illustrated that the partial fingerprint identification and MasterPrint mitigation method presented by Joshi et al. [143] delivered low performance on crucial parameters

when a variety of thresholding and thinning methods were employed in place of the original approaches. The combination of Bernsen's local image thresholding and K3M thinning algorithm produced above 90% DIR on three datasets, namely, DB1_A, NIST sd302b, and sd302d. But the approach generated more than 15% MasterPrints during the same experiment. Another binding of Bernsen's local image thresholding and Hilditch thinning algorithm delivered greater than 90% DIR on FVC datasets while generating more than 11% MasterPrints. Thus, the DIR and the percentage of MasterPrint generated by various preprocessing combinations on five benchmark datasets showed that no pairing has consistently performed better over the other methods. Moreover, the entries from Table 6.2 to Table 6.6 also confirm that suggesting a specific thresholding or thinning method as appropriate for partial fingerprint identification when experimented with given datasets, is not feasible. Based on the analysis, we deduced that the deviation in the results were contributed by five factors, namely, the thresholding method, thinning method, false minutiae removal approach, the sensor type, and the resolution of the fingerprint datasets. Precise thresholding and thinning of fingerprint ridges are the driving factors for accurate minutiae detection. The literature on image processing includes several approaches to binarize and thin grayscale images. All the thinning and thresholding approaches from the literature are incompatible for high-security applications, such as, anonymous user identification through a fingerprint. Conclusively, the results confirmed that high-security applications, and user identification systems employing biometric traits of an individual are greatly influenced by the choices made in the preprocessing stage.

Jabeen and Khan [170] proposed a hybrid algorithm for false minutiae and boundary elimination. The algorithm removes false minutiae from a thinned binarized fingerprint image arising due to bridges, spikes, and ridge breaks. Xiao and Raafat [171] presented a false minutiae detection and elimination method. The authors represented false minutiae using structural and statistical approaches. Kim et al. [167] method for false minutiae removal employed in the experimentation, and other similar approaches used thinned images for false minutiae detection and elimination. Hence, an inappropriate binarized and thinned fingerprint image is the agent behind the improper functioning of any false minutiae removal scheme.

Although there is no lower bound on the count of minutiae matched to decide if the

fingerprints are matched, some legal procedures accept at least an $8-17$ minutiae match for evidence [13]. An average minutiae density on a 500 dpi fingerprint is estimated to be 0.246 $minutiae/mm^2$ [172]. Such statistics should be employed to confirm that accurate minutiae detection and matching was carried out by a novel method proposed in future involving thresholding and thinning methods in the preprocessing phase. A robust novel fingerprint identification method should be experimented with fingerprint datasets employing diverse sensor types. A fingerprint identification system can generate MasterPrints if the fingerprint preprocessing approaches are not precisely tested on several datasets using different sensors.

As future scope of this work, the feasibility of robust partial fingerprint identification and MasterPrint mitigation method for poor quality latent fingerprints could be analysed. Also, studying multiple preprocessing schemes to prove its practicability can also be undertaken. As experimental results showed that preprocessing methods highly affect an identification system's accuracy, the initial goal can be experimenting with different preprocessing and false minutiae removal approaches on poor quality partial and latent fingerprint datasets acquired using dissimilar sensor types.

6.6 Summary of the chapter

The MasterPrint vulnerability makes a partial fingerprint identification system susceptible to presentation attacks. This chapter presented an investigation using sixteen combinations of thresholding and thinning methods on a partial fingerprint identification system to study their impact on the vital parameters, such as DIR and percentage of MasterPrints generated. The results demonstrated that the partial fingerprint identification and MasterPrint mitigation approach had delivered diminished performance that did not match with its original version. Hence, to prove the robustness and feasibility in practical application of a partial fingerprint identification system as a high-security person identification or access control system, firstly, it should be experimented with fingerprint datasets captured from diverse sensor types with poor and good quality images with varying resolutions.

Chapter 7

Conclusion and Future Scope of Work

MasterPrint vulnerability may not be regarded as a serious threat to conventional or contactless fingerprint biometric systems fingerprints using high-resolution full fingerprints for high-security applications. However, handheld devices like smartphones with small-size fingerprint sensors and offering less stringent security requirements require coping with security threats from presentation attacks using MasterPrints. The main objectives of the research in this thesis include introducing a comprehensive threat model depicting sixteen vulnerable attack points and addressing the MasterPrint vulnerability. The investigation on the impact of using various combinations of thresholding and thinning methods demonstrated that MasterPrint generation largely depends on the choice made during the preprocessing step. The MasterPrint vulnerability in the partial fingerprint identification systems was addressed using two novel minutiae-based local feature extraction approaches. This chapter provides the concluding remarks on the work carried out in this thesis. In the beginning, Section 7.1 presents the summary of outcomes from contributing chapters in this thesis. Subsequently, Section 7.2 highlights the possibilities to further extend the research in future.

7.1 Summary of contributions

This section presents the contributions made through the research work provided in this thesis. The following subsections give an overview of the outcomes from the thesis chapters.

7.1.1 The proposed threat model for fingerprint biometric system

The proposed threat model pinpointed sixteen attack points (AP) to highlight the system components vulnerable to various active and passive threats. The model highlighted various existing and potential threats to six components and ten communication links between them. The model depicted the components missing in the existing threat models and classified these components into eight classes to facilitate the security experts and researchers in addressing a common threat at multiple locations within the system. The model included the template protection techniques module and the biometric system controlled application which were missing in the existing models. The model further provided diverse techniques to thwart the attacks on the match-in-database fingerprint biometric system. Thus, the description of the model included the countermeasures to strengthen the security of the biometric system.

7.1.2 Investigating Latent MasterPrint

The MasterPrint vulnerability was studied for a partial fingerprint identification system. Latent fingerprints lifted from a crime scene are also mostly partial and possess unclear ridge patterns. Hence, it was essential to investigate the possibility of Latent MasterPrint. A Latent MasterPrint could be misused against an innocent. Due to reliance of legislative procedure on forensic investigation reports, the AFIS must produce accurate results to ensure only offenders get convicted. The experimental results demonstrated that the Latent MasterPrints exists, and confirmed that addressing MasterPrint vulnerability requires a prompt response from the research community.

7.1.3 MasterPrint mitigation using minutiae-based coordinate system

A novel minutiae-based eight axes coordinate system was introduced in the first attempt to address the MasterPrint vulnerability. The feature extraction employed the binarized and thinned partial fingerprint to create a seventeen element integer-valued unique feature vector. As strict feature matching was imposed during feature vector comparison, a new scheme

was introduced for similarity score computation. The results demonstrated that the proposed approach delivered an 88.78% average identification accuracy while generating 1.27% MasterPrints, on average. The proposed approach thus ensured that accurate user identification through partial fingerprint is possible while reducing the MasterPrint generation significantly.

7.1.4 MasterPrint mitigation employing minutiae geometry

The second method to mitigate MasterPrint vulnerability local minutiae-based geometric constructs were formed to create a feature vector representing each minutia uniquely. The experiments conducted on partial fingerprint datasets cropped from five benchmark full fingerprint datasets proved that local features involving adjacent minutiae alleviate the possibility of MasterPrint generation. The results demonstrated that the proposed approach delivered a 97% average identification accuracy while generating 0.1% MasterPrints, on average. The proposed approach provided high accuracy and generated negligible MasterPrints. Thus, it is suitable for practical use in IoT-based products requiring user identification.

7.1.5 Impact of preprocessing approaches on MasterPrint generation

Minutiae-based accurate feature extraction usually requires a seamlessly blended combination of thresholding and thinning methods in the fingerprint preprocessing stage. However, these methods are implemented independently, and existing literature does not show any evidence of the impact of choosing a diverse combination of these methods. Hence, an investigation involving four thresholding and four thinning approaches was conducted to study the performance on MasterPrint generation and identification accuracy. The observations proved that a novel approach must be rigorously tested using different preprocessing techniques to understand their collective effect on the system performance.

7.2 Future research directions

As research is a never-ending process, the contributions made in this thesis also has certain diverse objectives to be explored in future. The following paragraphs provide an insight into the future scope to extend the outcomes of this thesis.

1. The proposed threat model targeted the match-in-database fingerprint biometric system vulnerabilities. However, smart-card based biometric systems are gaining popularity over remote server-based biometric authentication systems. Hence, the proposed threat model requires an extension towards identifying and addressing threats to template-on-card (ToC), match-on-card (MoC), and system-on-card (SoC) biometric systems. The feasibility of presenting a single threat model applicable to a larger class of biometric systems may be explored in future.

2. The investigation in this thesis confirmed the possibility of Latent MasterPrint. Hence, devising a novel latent fingerprint identification approach extracting local features can be a problem statement for future work. The investigations should employ more latent fingerprint datasets and user-assisted good quality fingerprint datasets to validate the robustness of the approach.

3. The proposed methods for MasterPrint mitigation is highly dependent on the underlying binarization and thinning methods used in the preprocessing stage. Moreover, the partial fingerprint dataset was cropped from reasonably good quality full fingerprint benchmark datasets. Hence, the approaches can be tested in the future by employing diverse preprocessing methods on poor and latent fingerprint datasets.

4. It is possible to improve the performance of the proposed method further by employing a neural network and machine learning scheme, such as Generative Adversarial Networks (GANs), to produce higher quality images. However, a Least Squares Generative Adversarial Networks (LSGANs) employs least-square error minimisation and has demonstrated generating better quality images compared to regular GANs [173]. The samples outlying from the decision boundary are penalised by the least-square ap-

proach and pull them near the decision boundary [174]. Hence, LSGAN can be utilised in future during the preprocessing phase to improve the overall system performance.

Bibliography

[1] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001. [Online]. Available: https://doi.org/10.1147/sj.403.0614

[2] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 113:1–113:17, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1155/2008/579416

[3] E. Marasco and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems," *ACM Computing Surveys*, vol. 47, no. 2, pp. 28:1–28:36, 2014. [Online]. Available: https://doi.org/10.1145/2617756

[4] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction*. Boston, MA: Springer US, 2011, pp. 1–49. [Online]. Available: https://doi.org/10.1007/978-0-387-77326-1_1

[5] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security & Privacy*, vol. 1, no. 2, pp. 33–42, 2003. [Online]. Available: https://doi.org/10.1109/MSECP.2003.1193209

[6] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Introduction*. London: Springer London, 2009, pp. 1–56. [Online]. Available: https://doi.org/10.1007/978-1-84882-254-2_1

[7] A. Roy, N. D. Memon, and A. Ross, "Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2013–2025, 2017. [Online]. Available: https://doi.org/10.1109/TIFS.2017.2691658

[8] A. K. Jain and J. Feng, "Latent fingerprint matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 88–100, Jan. 2011. [Online]. Available: https://doi.org/10.1109%2Ftpami.2010.59

[9] A. Sankaran, M. Vatsa, and R. Singh, "Multisensor optical and latent fingerprint database," *IEEE Access*, vol. 3, pp. 653–665, 2015.

[10] A. K. Jain, K. Nandakumar, and A. Nagar, "Fingerprint template protection: From theory to practice," in *Security and Privacy in Biometrics*, 2013, pp. 187–214. [Online]. Available: https://doi.org/10.1007/978-1-4471-5230-9_8

[11] A. K. Jain, S. S. Arora, L. Best-Rowden, K. Cao, P. S. Sudhish, A. Bhatnagar, and Y. Koda, "Giving infants an identity: Fingerprint sensing and recognition," in *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development, ICTD 2016, Ann Arbor, MI, USA, June 03 - 06, 2016*, 2016, p. 29. [Online]. Available: http://doi.acm.org/10.1145/2909609.2909612

[12] S. Greenberg, M. Aladjem, and D. Kogan, "Fingerprint image enhancement using filtering techniques," *Real-Time Imaging*, vol. 8, no. 3, pp. 227 – 236, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077201401902839

[13] S. Z. Li and A. K. Jain, Eds., *Encyclopedia of Biometrics, Second Edition*. Springer US, 2015. [Online]. Available: https://doi.org/10.1007/978-1-4899-7488-4

[14] Z. Akhtar, A. Hadid, M. Nixon, M. Tistarelli, J. L. Dugelay, and S. Marcel, "Biometrics: In search of identity and security (q a)," *IEEE MultiMedia*, vol. PP, no. 99, pp. 1–1, 2017.

[15] M. D. Marsico, "Biometric recognition errors," in *Encyclopedia of Cryptography, Security and Privacy*. Springer Berlin Heidelberg, 2021, pp. 1–5. [Online]. Available: https://doi.org/10.1007/978-3-642-27739-9_1656-1

[16] Ratha N. K.,Connell J. H.,Bolle R. M., "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, p. , 2001. [Online]. Available: http://www.cedar.buffalo.edu/~govind/CSE717/papers/

[17] S. Li and A. C. Kot, "Attack using reconstructed fingerprint," in *2011 IEEE International Workshop on Information Forensics and Security, WIFS 2011, Iguacu Falls, Brazil, November 29 - December 2, 2011*. IEEE Computer Society, 2011, pp. 1–6. [Online]. Available: https://doi.org/10.1109/WIFS.2011.6123151

[18] A. Rozsa, A. E. Glock, and T. E. Boult, "Genetic algorithm attack on minutiae-based fingerprint authentication and protected template fingerprint systems," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2015, Boston, MA, USA, June 7-12, 2015*. IEEE Computer Society, 2015, pp. 100–108. [Online]. Available: https://doi.org/10.1109/CVPRW.2015.7301325

[19] R. Cappelli, M. Ferrara, and D. Maltoni, "Minutia cylinder-code: A new representation and matching technique for fingerprint recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2128–2141, 2010. [Online]. Available: https://doi.org/10.1109/TPAMI.2010.52

[20] M. Ferrara, D. Maltoni, and R. Cappelli, "Noninvertible minutia cylinder-code representation," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1727–1737, 2012. [Online]. Available: https://doi.org/10.1109/TIFS.2012.2215326

[21] Biometric System Laboratory, DISI - University of Bologna, "Minutia cylinder - code sdk." [Online]. Available: http://biolab.csr.unibo.it

[22] A. Pashalidis, "Simulated annealing attack on certain fingerprint authentication systems," in *2013 BIOSIG - Proceedings of the 12th International Conference of Biometrics Special Interest Group, Darmstadt, Germany, September 4-6, 2013*, ser. LNI, A. Brömme and C. Busch, Eds., vol. P-212. GI, 2013, pp. 63–74. [Online]. Available: https://dl.gi.de/20.500.12116/17692

[23] A. Juels and M. Sudan, "A fuzzy vault scheme," *IACR Cryptology ePrint Archive*, vol. 2002, p. 93, 2002. [Online]. Available: http://eprint.iacr.org/2002/093

[24] A. Kholmatov and B. A. Yanikoglu, "Realization of correlation attack against the fuzzy vault scheme," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, San Jose, CA, USA, January 27, 2008*, ser. SPIE Proceedings, vol. 6819. SPIE, 2008, p. 68190O. [Online]. Available: https://doi.org/10.1117/12.766861

[25] P. Mihailescu, A. Munk, and B. Tams, "The fuzzy vault for fingerprints is vulnerable to brute force attack," in *BIOSIG 2009 - Proceedings of the Special Interest Group on Biometrics and Electronic Signatures, 17.-18. September 2009 in Darmstadt, Germany*, ser. LNI, A. Brömme, C. Busch, and D. Hühnlein, Eds., vol. P-155. GI, 2009, pp. 43–54. [Online]. Available: https://dl.gi.de/20.500.12116/23190

[26] C. Roberts, "Biometric attack vectors and defences," *Computer & Security*, vol. 26, no. 1, pp. 14–25, 2007. [Online]. Available: https://doi.org/10.1016/j.cose.2006.12.008

[27] J. Galbally, "Anti-spoofing, fingerprint databases," in *Encyclopedia of Biometrics, Second Edition*, 2015, pp. 79–86. [Online]. Available: https://doi.org/10.1007/978-1-4899-7488-4_9115

[28] J. Wayman, "Technical testing and evaluation of biometric identification devices," in *Biometrics: Personal Identification in Networked Society*. Boston, MA: Springer US, 1996, ch. 17, pp. 345–368. [Online]. Available: https://doi.org/10.1007/0-306-47044-6_17

[29] Bartlow N., Cukic B., "The vulnerabilities of biometric systems - an integrated look and old and new ideas," *Technical report*, 2005.

[30] D. H. Nabil, K. Benatchba, M. Koudil, and A. Bouridane, "Threats models on biometric systems: A comparative study," in *Fourth International Conference on Computational Aspects of Social Networks, CASoN 2012, Sao Carlos, Brazil, November 21-23, 2012*. IEEE, 2012, pp. 186–191. [Online]. Available: https://doi.org/10.1109/CASoN.2012.6412400

[31] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia, "An evaluation of indirect attacks and countermeasures in fingerprint verification systems," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1643–1651, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016786551100105X

[32] U. Uludag and A. K. Jain, "Attacks on biometric systems: a case study in fingerprints," in *Security, Steganography, and Watermarking of Multimedia Contents VI, San Jose, California, USA, January 18-22, 2004, Proceedings*, 2004, pp. 622–633. [Online]. Available: https://doi.org/10.1117/12.530907

[33] M. Hendre, S. Patil, and A. Abhyankar, "Biometric recognition robust to partial and poor quality fingerprints using distinctive region adaptive SIFT keypoint fusion," *Multimedia Tools and Applications*, vol. 81, no. 12, pp. 17 483–17 507, Mar. 2022. [Online]. Available: https://doi.org/10.1007/s11042-021-11686-2

[34] K. Cao, E. Liu, L. Pang, J. Liang, and J. Tian, "Fingerprint matching by incorporating minutiae discriminability," in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, Oct. 2011. [Online]. Available: https://doi.org/10.1109/ijcb.2011.6117537

[35] P. Bontrager, A. Roy, J. Togelius, N. D. Memon, and A. Ross, "Deepmasterprints: Generating masterprints for dictionary attacks via latent variable evolution*," in *9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2018, Redondo Beach, CA, USA, October 22-25, 2018*. IEEE, 2018, pp. 1–9. [Online]. Available: https://doi.org/10.1109/BTAS.2018.8698539

[36] A. Roy, N. D. Memon, J. Togelius, and A. Ross, "Evolutionary methods for generating synthetic masterprint templates: Dictionary attack in fingerprint recognition," in *2018 International Conference on Biometrics, ICB 2018, Gold Coast, Australia, February 20-23, 2018*. IEEE, 2018, pp. 39–46. [Online]. Available: https://doi.org/10.1109/ICB2018.2018.00017

[37] E. Marasco and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems," *ACM Computing Surveys*, vol. 47, no. 2, pp. 1–36, Jan. 2015. [Online]. Available: https://doi.org/10.1145/2617756

[38] U. Uludag, A. Ross, and A. Jain, "Biometric template selection and update: a case study in fingerprints," *Pattern Recognition*, vol. 37, no. 7, pp. 1533–1542, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320304000081

[39] A. Jain, Y. Chen, and M. Demirkus, "Pores and ridges: Fingerprint matching using level 3 features," in *18th International Conference on Pattern Recognition (ICPR'06)*. IEEE, 2006. [Online]. Available: https://doi.org/10.1109/icpr.2006.938

[40] A. Ross, A. Jain, and J. Reisman, "A hybrid fingerprint matcher," *Pattern Recognition*, vol. 36, no. 7, pp. 1661–1673, Jul. 2003. [Online]. Available: https://doi.org/10.1016/s0031-3203(02)00349-7

[41] A. Ross, "Information fusion in fingerprint authentication," Ph.D. dissertation, Ph.D. dissertation, Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, 2003, unpublished thesis.

[42] A. Ross, S. Shah, and J. Shah, "Image versus feature mosaicing: a case study in fingerprints," in *SPIE Proceedings*, P. J. Flynn and S. Pankanti, Eds. SPIE, Apr. 2006. [Online]. Available: https://doi.org/10.1117/12.666278

[43] A. Roy, N. Memon, and A. Ross, "MasterPrint attack resistance: A maximum cover based approach for automatic fingerprint template selection," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, Sep. 2019. [Online]. Available: https://doi.org/10.1109/btas46853.2019.9186010

[44] S. Marcel, M. S. Nixon, and S. Z. Li, Eds., *Handbook of Biometric Anti-Spoofing - Trusted Biometrics under Spoofing Attacks*, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2014. [Online]. Available: https://doi.org/10.1007/978-1-4471-6524-8

[45] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: Challenges and research opportunities," *IEEE Security & Privacy*, vol. 13, no. 5, pp. 63–72, 2015. [Online]. Available: https://doi.org/10.1109/MSP.2015.116

[46] J. Galbally, R. Cappelli, A. Lumini, G. G. de Rivera, D. Maltoni, J. Fiérrez, J. Ortega-Garcia, and D. Maio, "An evaluation of direct attacks using fake fingers generated from ISO templates," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 725–732, 2010. [Online]. Available: https://doi.org/10.1016/j.patrec.2009.09.032

[47] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial "gummy" fingers on fingerprint systems," *Datenschutz und Datensicherheit*, vol. 26, no. 8, 2002.

[48] J. Galbally, R. Cappelli, A. Lumini, D. Maltoni, and J. Fiérrez-Aguilar, "Fake fingertip generation from a minutiae template," in *19th International Conference on Pattern Recognition (ICPR 2008), December 8-11, 2008, Tampa, Florida, USA*, 2008, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ICPR.2008.4761456

[49] H. Choi, R. Kang, K. Choi, and J. Kim, "Aliveness detection of fingerprints using multiple static features," *World Academy of Science, Engineering and Technology*, vol. 2, Jan. 2007.

[50] R. P. Sharma and S. Dey, "Fingerprint liveness detection using local quality features," *The Visual Computer*, vol. 35, no. 10, pp. 1393–1410, 2019. [Online]. Available: https://doi.org/10.1007/s00371-018-01618-x

[51] L. Ghiani, A. Hadid, G. L. Marcialis, and F. Roli, "Fingerprint liveness detection using local texture features," *IET Biometrics*, vol. 6, no. 3, pp. 224–231, 2017. [Online]. Available: https://doi.org/10.1049/iet-bmt.2016.0007

[52] Z. Xia, R. Lv, Y. Zhu, P. Ji, H. Sun, and Y. Shi, "Fingerprint liveness detection using gradient-based texture features," *Signal, Image and Video Processing*, vol. 11, no. 2, pp. 381–388, 2017. [Online]. Available: https://doi.org/10.1007/s11760-016-0936-z

[53] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *Audio- and Video-Based Biometric Person Authentication, Third International Conference, AVBPA 2001 Halmstad, Sweden, June 6-8, 2001, Proceedings*, 2001, pp. 223–228. [Online]. Available: https://doi.org/10.1007/3-540-45344-X_32

[54] S. Naraharisetty. Transfer encryption. [Online]. Available: https://www.2brightsparks.com/resources/articles/transfer-encryption.html

[55] I. Anshel, D. Atkins, D. Goldfeld, and P. E. Gunnells, "Walnutdsa(tm): A quantum resistant group theoretic digital signature algorithm," *IACR Cryptology ePrint Archive*, vol. 2017, p. 58, 2017. [Online]. Available: http://eprint.iacr.org/2017/058

[56] J. Waddle and D. A. Wagner, "Towards efficient second-order power analysis," in *Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings*, 2004, pp. 1–15. [Online]. Available: https://doi.org/10.1007/978-3-540-28632-5_1

[57] F. Standaert, "Introduction to side-channel attacks," in *Secure Integrated Circuits and Systems*, 2010, pp. 27–42. [Online]. Available: https://doi.org/10.1007/978-0-387-71829-3_2

[58] J. Galbally, S. Carballo, J. Fiérrez, and J. Ortega-Garcia, "Vulnerability assessment of fingerprint matching based on time analysis," in *Biometric ID Management and Multimodal Communication, Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, September 16-18, 2009. Proceedings*, 2009, pp. 285–292. [Online]. Available: https://doi.org/10.1007/978-3-642-04391-8_37

[59] K. Gandolfi, C. Mourtel, and F. Olivier, "Electromagnetic analysis: Concrete results," in *Cryptographic Hardware and Embedded Systems - CHES 2001, Third International Workshop, Paris, France, May 14-16, 2001, Proceedings*, no. Generators, 2001, pp. 251–261. [Online]. Available: https://doi.org/10.1007/3-540-44709-1_21

[60] J.-S. Coron, A. Greuet, and R. Zeitoun, "Side-channel masking with pseudo-random generator," in *Advances in Cryptology – EUROCRYPT 2020*. Springer International Publishing, 2020, pp. 342–375. [Online]. Available: https://doi.org/10.1007/978-3-030-45727-3_12

[61] S. Yang and I. M. Verbauwhede, "A secure fingerprint matching technique," in *Proceedings of the 2003 ACM SIGMM Workshop on Biometrics Methods and Applications*, ser. WBMA '03. New York, NY, USA: ACM, 2003, pp. 89–94. [Online]. Available: http://doi.acm.org/10.1145/982507.982524

[62] A. Ross, J. Shah, and A. K. Jain, "From template to image: Reconstructing fingerprints from minutiae points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 544–560, 2007. [Online]. Available: https://doi.org/10.1109/TPAMI.2007.1018

[63] S. Li and A. C. Kot, "Privacy protection of fingerprint database," *IEEE Signal Processing Letters*, vol. 18, no. 2, pp. 115–118, 2011. [Online]. Available: https://doi.org/10.1109/LSP.2010.2097592

[64] P. C. van Oorschot, "Revisiting software protection," in *Information Security, 6th International Conference, ISC 2003, Bristol, UK, October 1-3, 2003, Proceedings*, 2003, pp. 1–13. [Online]. Available: https://doi.org/10.1007/10958513_1

[65] H. Chang and M. J. Atallah, "Protecting software code by guards," in *Security and Privacy in Digital Rights Management, ACM CCS-8 Workshop DRM 2001, Philadelphia, PA, USA, November 5, 2001, Revised Papers*, 2001, pp. 160–175. [Online]. Available: https://doi.org/10.1007/3-540-47870-1_10

[66] P. C. van Oorschot, A. Somayaji, and G. Wurster, "Hardware-assisted circumvention of self-hashing software tamper resistance," *IEEE Transactions on Dependable and Secure Computing*, vol. 2, no. 2, pp. 82–92, 2005. [Online]. Available: https://doi.org/10.1109/TDSC.2005.24

[67] M. Martinez-Diaz, J. Fiérrez, J. Galbally, and J. Ortega-Garcia, "An evaluation of indirect attacks and countermeasures in fingerprint verification systems," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1643–1651, 2011. [Online]. Available: https://doi.org/10.1016/j.patrec.2011.04.005

[68] M. Berthier, Y. Bocktaels, J. Bringer, H. Chabanne, T. Chouta, J. Danger, M. Favre, and T. Graba, "Studying potential side channel leakages on an embedded biometric comparison system," *IACR Cryptology ePrint Archive*, vol. 2014, p. 26, 2014. [Online]. Available: http://eprint.iacr.org/2014/026

[69] K. Tiri, D. D. Hwang, A. Hodjat, B. Lai, S. Yang, P. Schaumont, and I. Verbauwhede, "A side-channel leakage free coprocessor IC in 0.18μm CMOS for embedded aes-based cryptographic and biometric processing," in *Proceedings of the 42nd Design Automation Conference, DAC 2005, San Diego, CA, USA, June 13-17, 2005*, W. H. J. Jr., G. Martin, and A. B. Kahng, Eds. ACM, 2005, pp. 222–227. [Online]. Available: https://doi.org/10.1145/1065579.1065639

[70] N. S. N.V., "Nxp smartmx family," https://www.nxp.com/docs/en/brochure/75017515.pdf, 2014.

[71] Y. He, J. Tian, L. Li, H. Chen, and X. Yang, "Fingerprint matching based on global comprehensive similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 850–862, 2006. [Online]. Available: https://doi.org/10.1109/TPAMI.2006.119

[72] A. Ross, A. K. Jain, and J. Reisman, "A hybrid fingerprint matcher," *Pattern Recognition*, vol. 36, no. 7, pp. 1661–1673, 2003. [Online]. Available: https://doi.org/10.1016/S0031-3203(02)00349-7

[73] M. A. Medina-Pérez, A. Gutiérrez-Rodríguez, and M. García-Borroto, "Improving fingerprint matching using an orientation-based minutia descriptor," in *Progress*

in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 14th Iberoamerican Conference on Pattern Recognition, CIARP 2009, Guadalajara, Jalisco, Mexico, November 15-18, 2009. Proceedings, ser. Lecture Notes in Computer Science, E. Bayro-Corrochano and J. Eklundh, Eds., vol. 5856. Springer, 2009, pp. 121–128. [Online]. Available: https://doi.org/10.1007/978-3-642-10268-4_14

[74] A. K. Jain, A. Ross, and S. Prabhakar, "Fingerprint matching using minutiae and texture features," in *Proceedings of the 2001 International Conference on Image Processing, ICIP 2001, Thessaloniki, Greece, October 7-10, 2001*. IEEE, 2001, pp. 282–285. [Online]. Available: https://doi.org/10.1109/ICIP.2001.958106

[75] M. Tico and P. Kuosmanen, "Fingerprint matching using an orientation-based minutia descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1009–1014, 2003. [Online]. Available: https://doi.org/10.1109/TPAMI.2003.1217604

[76] N. Poh, R. Wong, and G. L. Marcialis, "Toward an attack-sensitive tamper-resistant biometric recognition with a symmetric matcher: A fingerprint case study," in *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management, CIBIM 2014, Orlando, FL, USA, December 9-12, 2014*. IEEE, 2014, pp. 175–180. [Online]. Available: https://doi.org/10.1109/CIBIM.2014.7015460

[77] T. Zhang, J. Tian, Y. He, and X. Yang, "Fingerprint alignment using similarity histogram," in *Audio-and Video-Based Biometrie Person Authentication, 4th International Conference, AVBPA 2003, Guildford, UK, June 9-11, 2003 Proceedings*, ser. Lecture Notes in Computer Science, J. Kittler and M. S. Nixon, Eds., vol. 2688. Springer, 2003, pp. 854–861. [Online]. Available: https://doi.org/10.1007/3-540-44887-X_99

[78] M. El-Abed, P. Lacharme, and C. Rosenberger, "Security EvaBio: An analysis tool for the security evaluation of biometric authentication systems," in *5th IAPR International Conference on Biometrics, ICB 2012, New Delhi, India, March 29 - April 1, 2012*, 2012, pp. 460–465. [Online]. Available: https://doi.org/10.1109/ICB.2012.6199793

[79] A. Adler, "Biometric system security," in *Handbook of Biometrics*, A. Jain, P. Flynn, and A. Ross, Eds. Boston, MA: Springer US, 2008, ch. 19, pp. 381–402. [Online]. Available: https://doi.org/10.1007/978-0-387-71041-9_19

[80] A. Hadid, N. W. D. Evans, S. Marcel, and J. Fiérrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20–30, 2015. [Online]. Available: https://doi.org/10.1109/MSP.2015.2437652

[81] J. Feng, A. K. Jain, and A. Ross, "Detecting altered fingerprints," in *20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010*, 2010, pp. 1622–1625. [Online]. Available: https://doi.org/10.1109/ICPR.2010.401

[82] J. Shelton, K. S. Bryant, S. Abrams, L. Small, J. Adams, D. Leflore, A. Alford, K. Ricanek, and G. V. Dozier, "Genetic & evolutionary biometric

security: Disposable feature extractors for mitigating biometric replay attacks," in *Proceedings of the Conference on Systems Engineering Research, CSER 2012, St. Louis, MO, USA, March 19-22, 2012*, 2012, pp. 351–360. [Online]. Available: https://doi.org/10.1016/j.procs.2012.01.072

[83] M. Dürmuth, D. Oswald, and N. Pastewka, "Side-channel attacks on fingerprint matching algorithms," in *Proceedings of the 6th International Workshop on Trustworthy Embedded Devices, TrustED@CCS 16, Vienna, Austria, October 28, 2016*, 2016, pp. 3–13. [Online]. Available: http://doi.acm.org/10.1145/2995289.2995294

[84] A. A. Ross, J. Shah, and A. K. Jain, "Toward reconstructing fingerprints from minutiae points," in *SPIE Proceedings*, A. K. Jain and N. K. Ratha, Eds. SPIE, Mar. 2005. [Online]. Available: https://doi.org/10.1117/12.604477

[85] A. K. Jain, A. Ross, and U. Uludag, "Biometric security: Challenges and solutions," in *13th European Signal Processing Conference, EUSIPCO 2005, Antalya, Turkey, September 4-8, 2005*, 2005, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/7078369/

[86] M. Martinez-Diaz, J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia, and J. A. Siguenza, "Hill-climbing and brute-force attacks on biometric systems: A case study in match-on-card fingerprint verification," in *Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology*, Oct. 2006, pp. 151–159.

[87] B. Tams, "Attacks and countermeasures in fingerprint based biometric cryptosystems," *CoRR*, vol. abs/1304.7386, 2013. [Online]. Available: http://arxiv.org/abs/1304.7386

[88] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, Jun. 2006.

[89] N. Singla, M. Kaur, and S. Sofat, "Automated latent fingerprint identification system: A review," *Forensic Science International*, vol. 309, p. 110187, 2020.

[90] K. N. Win, K. Li, J. Chen, P. Fournier-Viger, and K. Li, "Fingerprint classification and identification algorithms for criminal investigation: A survey," *Future Generation Computer Systems*, vol. 110, pp. 758–771, 2020.

[91] S. Yoon, K. Cao, E. Liu, and A. K. Jain, "LFIQ: latent fingerprint image quality," in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems, BTAS 2013, Arlington, VA, USA, September 29 - October 2, 2013*. IEEE, 2013, pp. 1–8.

[92] Himanshi, A. Kaur, and A. Verma, "Latent fingerprint recognition using hybridization approach of partial differential equation and exemplar inpainting," *Indian Journal of Science and Technology*, vol. 9, no. 45, Dec. 2016.

[93] M. D. Garris and R. M. McCabe, "NIST Special Database 27:," National Institute of Standards and Technology, Tech. Rep., 2000. [Online]. Available: https://doi.org/10.6028/nist.ir.6534

[94] A A Paulino and A K Jain and Jianjiang Feng, "Latent fingerprint matching: Fusion of manually marked and derived minutiae," in *2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images*. IEEE, Aug. 2010. [Online]. Available: https://doi.org/10.1109/sibgrapi.2010.17

[95] U. U. Deshpande, V. S. Malemath, S. M. Patil, and S. V. Chaugule, "Latent fingerprint identification system based on a local combination of minutiae feature points," *SN Computer Science*, vol. 2, no. 3, Apr. 2021.

[96] R. P. Krish, J. Fierrez, D. Ramos, F. Alonso-Fernandez, and J. Bigun, "Improving automated latent fingerprint identification using extended minutia types," *Information Fusion*, vol. 50, pp. 9–19, Oct. 2019.

[97] D. Blackburn, C. Miles, B. Wing, and K. Shepard, "Biometric testing and statistics," National Science and Technology Council (NSTC), Washington, D.C., Tech. Rep., 2006. [Online]. Available: http://www.nws-sa.com/biometrics/testing/BioTestingAndStats.pdf

[98] B. DeCann and A. Ross, "Relating ROC and CMC curves via the biometric menagerie," in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems, BTAS 2013, Arlington, VA, USA, September 29 - October 2, 2013*. IEEE, 2013, pp. 1–8.

[99] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, S. Janet, and K. Ko, "User's Guide to NIST Biometric Image Software (NBIS)," *NIST Interagency/Internal Report (NISTIR) - 7392*, 2007. [Online]. Available: https://doi.org/10.6028/NIST.IR.7392

[100] A. Sankaran, M. Vatsa, and R. Singh, "Latent fingerprint matching: A survey," *IEEE Access*, vol. 2, pp. 982–1004, 2014.

[101] "Fingerprint Verification Competition (FVC2002)," 2002, http://bias.csr.unibo.it/fvc2002/databases.asp.

[102] "Crossmatch Sample DB Dataset," 2020, https://neurotechnology.com/download/CrossMatch_Sample_DB.zip.

[103] G. Fiumara, P. Flanagan, J. Grantham, K. Ko, K. Marshall, M. Schwarz, E. Tabassi, B. Woodgate, and C. Boehnen, "National Institute of Standards and Technology Special Database 302: Nail to Nail Fingerprint Challenge," National Institute of Standards and Technology, Technical Note 2007, Aug. 2018.

[104] W. Lee, S. Cho, H. Choi, and J. Kim, "Partial fingerprint matching using minutiae and ridge shape features for small fingerprint scanners," *Expert Systems with Applications*, vol. 87, pp. 183–198, 2017. [Online]. Available: https://doi.org/10.1016/j.eswa.2017.06.019

[105] A. Kumar, "Introduction to trends in fingerprint identification," in *Contactless 3D Fingerprint Identification*. Springer International Publishing, 2018, pp. 1–15. [Online]. Available: https://doi.org/10.1007/978-3-319-67681-4_1

[106] A. Kumar and Y. Zhou, "Contactless fingerprint identification using level zero features," in *CVPR 2011 WORKSHOPS*. IEEE, Jun. 2011. [Online]. Available: https://doi.org/10.1109/cvprw.2011.5981823

[107] H. Kaur, D. Koundal, and V. Kadyan, "Image fusion techniques: A survey," *Archives of Computational Methods in Engineering*, vol. 28, no. 7, pp. 4425–4447, Jan. 2021. [Online]. Available: https://doi.org/10.1007/s11831-021-09540-7

[108] P. Vijayaprasad, M. N. Sulaiman, N. Mustapha, and R. Rahmat, "Partial fingerprint recognition using support vector machine," *Information Technology Journal*, vol. 9, no. 4, pp. 844–848, May 2010. [Online]. Available: https://doi.org/10.3923/itj.2010.844.848

[109] J. Khodadoust and A. M. Khodadoust, "Partial fingerprint identification for large databases," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 19–34, 2018. [Online]. Available: https://doi.org/10.1007/s10044-017-0665-0

[110] O. Zanganeh, N. Bhattacharjee, and B. Srinivasan, "Partial fingerprint alignment and matching through region-based approach," in *Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia, MoMM 2015, Brussels, Belgium, December 11-13, 2015*, 2015, pp. 275–284. [Online]. Available: https://doi.org/10.1145/2837126.2837132

[111] T. Jea and V. Govindaraju, "A minutia-based partial fingerprint recognition system," *Pattern Recognition*, vol. 38, no. 10, pp. 1672–1684, 2005. [Online]. Available: https://doi.org/10.1016/j.patcog.2005.03.016

[112] G. P. Arada and E. P. Dadios, "Partial fingerprint identification through checkerboard sampling method using ann," in *TENCON 2012 IEEE Region 10 Conference*, Nov. 2012, pp. 1–6.

[113] T.-Y. Jea, V. S. Chavan, V. Govindaraju, and J. K. Schneider, "Security and matching of partial fingerprint recognition systems," in *Biometric Technology for Human Identification*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, A. K. Jain and N. K. Ratha, Eds., vol. 5404, Aug. 2004, pp. 39–50.

[114] F. Zeng, S. Hu, and K. Xiao, "Research on partial fingerprint recognition algorithm based on deep learning," *Neural Computing & Applications*, vol. 31, no. 9, pp. 4789–4798, 2019. [Online]. Available: https://doi.org/10.1007/s00521-018-3609-8

[115] S. Mil'shtein, A. Pillai, A. Shendye, C. Liessner, and M. Baier, "Fingerprint recognition algorithms for partial and full fingerprints," in *2008 IEEE Conference on Technologies for Homeland Security*, May 2008, pp. 449–452.

[116] F. Zhang, S. Xin, and J. Feng, "Combining global and minutia deep features for partial high-resolution fingerprint matching," *Pattern Recognition Letters*, vol. 119, pp. 139–147, 2019. [Online]. Available: https://doi.org/10.1016/j.patrec.2017.09.014

[117] G. Bae, H. Lee, S. Son, D. Hwang, and J. Kim, "Secure and robust user authentication using partial fingerprint matching," in *IEEE International Conference on Consumer Electronics, ICCE 2018, Las Vegas, NV, USA, January 12-14, 2018*, 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICCE.2018.8326078

[118] A. Aravindan and A. S. M., "Robust partial fingerprint recognition using wavelet SIFT descriptors," *Pattern Analysis and Applications*, vol. 20, no. 4, pp. 963–979, 2017. [Online]. Available: https://doi.org/10.1007/s10044-017-0615-x

[119] M. R. Flores, G. A. Torres, G. G. García, and M. Á. G. Licona, "Fingerprint verification using computational geometry," *DYNA*, vol. 83, no. 195, pp. 128–137, Feb. 2016. [Online]. Available: https://doi.org/10.15446/dyna.v83n195.46323

[120] Y. Chen and A. K. Jain, "Dots and incipients: Extended features for partial fingerprint matching," in *2007 Biometrics Symposium*, Sep. 2007, pp. 1–6.

[121] G. Fang, S. N. Srihari, H. Srinivasan, and P. Phatak, "Use of ridge points in partial fingerprint matching," in *Biometric Technology for Human Identification IV*, S. Prabhakar and A. A. Ross, Eds., vol. 6539, International Society for Optics and Photonics. SPIE, 2007, pp. 115 – 123. [Online]. Available: https://doi.org/10.1117/12.718941

[122] C.-H. Chang, J.-H. Lin, and I. Her, "New minutiae-matching method based on partial fingerprints," *Journal of Imaging Science and Technology*, vol. 56, pp. 1–10, Jan. 2012.

[123] Q. Gao, "Schemes of utilizing partial fingerprints for user verification," in *American Society for Engineering Education (ASEE), Mid Atlantic Conference, Newark, DE (April 20-21, 2012)*, 2012, pp. 1–7.

[124] F. Chen, M. Li, and Y. Zhang, "A fusion method for partial fingerprint recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 6, 2013. [Online]. Available: https://doi.org/10.1142/S0218001413560090

[125] O. Zanganeh, B. Srinivasan, and N. Bhattacharjee, "Partial fingerprint matching through region-based similarity," in *2014 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2014, Wollongong, New South Wales, Australia, November 25-27, 2014*, 2014, pp. 1–8. [Online]. Available: https://doi.org/10.1109/DICTA.2014.7008121

[126] W. Zhou, J. Hu, S. Wang, I. R. Petersen, and M. Bennamoun, "Partial fingerprint indexing: a combination of local and reconstructed global features," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 10, pp. 2940–2957, 2016. [Online]. Available: https://doi.org/10.1002/cpe.3600

[127] J. Qin, S. Tang, C. Han, and T. Guo, "Partial fingerprint matching via phase-only correlation and deep convolutional neural network," in *Neural Information Processing - 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part VI*, 2017, pp. 602–611. [Online]. Available: https://doi.org/10.1007/978-3-319-70136-3_64

[128] S. Boujnah, S. Jaballah, A. B. Khalifa, and M. L. Ammari, "Person's identification with partial fingerprint based on a redefinition of minutiae features," in *15th IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2018, Aqaba, Jordan, October 28 - Nov. 1, 2018*, 2018, pp. 1–5. [Online]. Available: https://doi.org/10.1109/AICCSA.2018.8612884

[129] U. U. Deshpande, V. S. Malemath, S. M. Patil, and S. V. Chaugule, "CNNAI: A convolution neural network-based latent fingerprint matching using the combination of nearest neighbor arrangement indexing," *Frontiers Robotics AI*, vol. 7, p. 113, 2020. [Online]. Available: https://doi.org/10.3389/frobt.2020.00113

[130] F. Zhang, X. Chen, and X. Zhang, "Parallel thinning and skeletonization algorithm based on cellular automaton," *Multimedia Tools and Applications*, vol. 79, no. 43-44, pp. 33 215–33 232, 2020. [Online]. Available: https://doi.org/10.1007/s11042-020-09660-5

[131] G. Limei, Z. Yingbin, and H. Duan, "A fingerprint minutiae extraction method in quantum thinned binary image," *International Journal of Theoretical Physics*, vol. 60, no. 5, pp. 1883–1894, May 2021. [Online]. Available: https://doi.org/10.1007/s10773-021-04807-y

[132] S. Kim, D. Lee, and J. Kim, "Algorithm for detection and elimination of false minutiae in fingerprint images," in *Audio- and Video-Based Biometric Person Authentication, Third International Conference, AVBPA 2001 Halmstad, Sweden, June 6-8, 2001, Proceedings*, 2001, pp. 235–240. [Online]. Available: https://doi.org/10.1007/3-540-45344-X_34

[133] D. Peralta, M. Galar, I. Triguero, D. Paternain, S. García, E. Barrenechea, J. M. Benítez, H. Bustince, and F. Herrera, "A survey on fingerprint minutiae-based local matching for verification and identification: Taxonomy and experimental evaluation," *Information Sciences*, vol. 315, pp. 67–87, 2015. [Online]. Available: https://doi.org/10.1016/j.ins.2015.04.013

[134] A. M. Bazen and S. H. Gerez, "Fingerprint matching by thin-plate spline modelling of elastic deformations," *Pattern Recognition*, vol. 36, no. 8, pp. 1859–1867, 2003. [Online]. Available: https://doi.org/10.1016/S0031-3203(03)00036-0

[135] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008, similarity Matching in Computer Vision and Multimedia. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314207001555

[136] K. H. Kuban and W. M. Jwaid, "A novel modification of SURF algorithm for fingerprint matching," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 6, pp. 1570–1581, 2018.

[137] G. Fang, S. Srihari, H. Srinivasan, and P. Phatak, "Use of ridge points in partial fingerprint matching," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6539, Apr. 2007.

[138] L. Nanni and A. Lumini, "Descriptors for image-based fingerprint matchers," *Expert Systems and Applications*, vol. 36, no. 10, pp. 12 414–12 422, 2009. [Online]. Available: https://doi.org/10.1016/j.eswa.2009.04.041

[139] S. Mathur, A. Vjay, J. Shah, S. Das, and A. Malla, "Methodology for partial fingerprint enrollment and authentication on mobile devices," in *International Conference on Biometrics, ICB 2016, Halmstad, Sweden, June 13-16, 2016*, 2016, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ICB.2016.7550093

[140] D. Bradley and G. Roth, "Adaptive thresholding using the integral image," *Journal of Graphics Tools*, vol. 12, no. 2, pp. 13–21, 2007. [Online]. Available: https://doi.org/10.1080/2151237X.2007.10129236

[141] U. U. Deshpande and V. S. Malemath, "MINU-EXTRACTNET: automatic latent fingerprint feature extraction system using deep convolutional neural network," in *Recent Trends in Image Processing and Pattern Recognition - Third International Conference, RTIP2R 2020, Aurangabad, India, January 3-4, 2020, Revised Selected Papers, Part I*, ser. Communications in Computer and Information Science, K. C. Santosh and B. Gawali, Eds., vol. 1380. Springer, 2020, pp. 44–56. [Online]. Available: https://doi.org/10.1007/978-981-16-0507-9_5

[142] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds., 2015, pp. 2017–2025. [Online]. Available: https://proceedings.neurips.cc/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html

[143] M. Joshi, B. Mazumdar, and S. Dey, "Mitigating MasterPrint vulnerability by employing minutiae geometry," *Journal of Electronic Imaging*, vol. 31, no. 1, pp. 1 – 20, 2022. [Online]. Available: https://doi.org/10.1117/1.JEI.31.1.013026

[144] M. Sonka, V. Hlavác, and R. Boyle, *Image Processing, Analysis and Machine Vision (3. ed.)*. Thomson, 2008.

[145] P. Stathis, E. Kavallieratou, and N. Papamarkos, "An evaluation technique for binarization algorithms," *Journal of Universal Computer Science*, vol. 14, no. 18, pp. 3011–3030, 2008. [Online]. Available: https://doi.org/10.3217/jucs-014-18-3011

[146] S. H. Shaikh, K. Saeed, and N. Chaki, "Performance benchmarking of different binarization techniques for fingerprint-based biometric authentication," in *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013, Milkow, Poland, 27-29 May 2013*, ser. Advances in Intelligent Systems and Computing, R. Burduk, K. Jackowski, M. Kurzynski, M. Wozniak, and A. Zolnierek, Eds., vol. 226. Springer, 2013, pp. 237–246. [Online]. Available: https://doi.org/10.1007/978-3-319-00969-8_23

[147] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[148] W. Niblack, *An introduction to digital image processing*. Prentice Hall International, 1988.

[149] B. J., "Dynamic thresholding of grey-level images," in *Proc. of the 8th Int. Conf. on Pattern Recognition*, Paris, France, 1986, pp. 1251–1255.

[150] K. Buch, H. Kuno, M. M. Qureshi, B. Li, and O. Sakai, "Quantitative variations in texture analysis features dependent on mri scanning parameters: A phantom model," *Journal of Applied Clinical Medical Physics*, vol. 19, no. 6, pp. 253–264, 2018. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/acm2.12482

[151] N. B. Rais, M. S. Hanif, and I. A. Taj, "Adaptive thresholding technique for document image analysis," in *8th International Multitopic Conference, 2004. Proceedings of INMIC 2004.*, 2004, pp. 61–66.

[152] K. Khurshid, I. Siddiqi, C. Faure, and N. Vincent, "Comparison of niblack inspired binarization methods for ancient documents," in *Document Recognition and Retrieval XVI, part of the IS&T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 20-22, 2009. Proceedings*, ser. SPIE Proceedings, K. Berkner and L. Likforman-Sulem, Eds., vol. 7247. SPIE, 2009, p. 72470U. [Online]. Available: https://doi.org/10.1117/12.805827

[153] N Senthilkumaran and S Vaithegi, "Image segmentation by using thresholding techniques for medical images," *Computer Science and Engineering: An International Journal*, vol. 6, no. 1, pp. 1–13, Feb. 2016. [Online]. Available: https://doi.org/10.5121/cseij.2016.6101

[154] Q. Li, X. Bai, and W. Liu, "Skeletonization of gray-scale image from incomplete boundaries," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008. [Online]. Available: https://doi.org/10.1109/icip.2008.4711895

[155] P. Bontrager, A. Roy, J. Togelius, N. Memon, and A. Ross, "DeepMasterPrints: Generating MasterPrints for dictionary attacks via latent variable evolution," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, Oct. 2018. [Online]. Available: https://doi.org/10.1109/btas.2018.8698539

[156] M. Nazarkevych, S. Dmytruk, V. Hrytsyk, O. Vozna, A. Kuza, O. Shevchuk, Y. Voznyi, I. Maslanych, and V. Sheketa, "Evaluation of the effectiveness of different image skeletonization methods in biometric security systems," *International Journal of Sensors, Wireless Communications and Control*, vol. 11, no. 5, pp. 542–552, Jun. 2021. [Online]. Available: https://doi.org/10.2174/2210327910666201210151809

[157] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984. [Online]. Available: https://doi.org/10.1145/357994.358023

[158] C. J. Hilditch, "Linear skeletons from square cupboards," in *Machine Intelligence 4*, B. Meltzer and D. Michie, Eds. Edinburgh University Press, 1969, p. 403.

[159] P. K. Saha, G. Borgefors, and G. S. di Baja, "A survey on skeletonization algorithms and their applications," *Pattern Recognition Letters*, vol. 76, pp. 3–12, 2016. [Online]. Available: https://doi.org/10.1016/j.patrec.2015.04.006

[160] F. W. M. Stentiford and R. G. Mortimer, "Some new heuristics for thinning binary handprinted characters for OCR," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 13, no. 1, pp. 81–84, 1983. [Online]. Available: https://doi.org/10.1109/TSMC.1983.6313034

[161] K. Saeed, M. Tabedzki, M. Rybnik, and M. Adamski, "K3M: A universal algorithm for image skeletonization and a review of thinning techniques," *International Journal of Applied Mathematics and Computer Science*, vol. 20, no. 2, pp. 317–335, 2010. [Online]. Available: https://doi.org/10.2478/v10006-010-0024-4

[162] M. Tabedzki, K. Saeed, and A. Szczepanski, "A modified K3M thinning algorithm," *International Journal of Applied Mathematics and Computer Science*, vol. 26, no. 2, pp. 439–450, 2016. [Online]. Available: https://doi.org/10.1515/amcs-2016-0031

[163] K. Saeed, M. Rybnik, and M. Tabedzki, "Implementation and advanced results on the non-interrupted skeletonization algorithm," in *Computer Analysis of Images and Patterns, 9th International Conference, CAIP 2001 Warsaw, Poland, September 5-7, 2001, Proceedings*, ser. Lecture Notes in Computer Science, W. Skarbek, Ed., vol. 2124. Springer, 2001, pp. 601–609. [Online]. Available: https://doi.org/10.1007/3-540-44692-3_72

[164] K. Saeed, "Text and image processing: Non-interrupted skeletonization," in *Proceedings of the 1st International IEEE Conference on Circuits, Systems, Comunications and Computers—IEEE-CSCC'01*. Society Press, 2001, pp. 350–354.

[165] J. Yu and Y. Li, "Improving hilditch thinning algorithms for text image," in *Proceedings of the 2009 International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government*, ser. EEEE '09. USA: IEEE Computer Society, 2009, p. 76–79. [Online]. Available: https://doi.org/10.1109/EEEE.2009.44

[166] S. Yokoi, J. Toriwaki, and T. Fukumura, "Topological properties in digitized binary pictures," *Systems Computers Controls*, vol. 4, pp. 32–39, 1973.

[167] S. Kim, D. Lee, and J. Kim, "Algorithm for detection and elimination of false minutiae in fingerprint images," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001, pp. 235–240. [Online]. Available: https://doi.org/10.1007/3-540-45344-x_34

[168] D. Blackburn, C. Miles, B. Wing, and K. Shepard, "Biometric testing and statistics," National Science and Technology Council (NSTC), Washington, D.C., Tech. Rep., 2006. [Online]. Available: http://www.nws-sa.com/biometrics/testing/BioTestingAndStats.pdf

[169] B. DeCann and A. Ross, "Relating ROC and CMC curves via the biometric menagerie," in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems, BTAS 2013, Arlington, VA, USA, September 29 - October 2, 2013*. IEEE, 2013, pp. 1–8. [Online]. Available: https://doi.org/10.1109/BTAS.2013.6712705

[170] S. Jabeen and S. A. Khan, "A hybrid false minutiae removal algorithm with boundary elimination," in *2008 IEEE International Conference on System of Systems Engineering*. IEEE, Jun. 2008. [Online]. Available: https://doi.org/10.1109/sysose.2008.4724177

[171] Q. Xiao and H. Raafat, "Fingerprint image postprocessing: A combined statistical and structural approach," *Pattern Recognition*, vol. 24, no. 10, pp. 985–992, Jan. 1991. [Online]. Available: https://doi.org/10.1016/0031-3203(91)90095-m

[172] S. Pankanti, S. Prabhakar, and A. K. Jain, "On the individuality of fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1010–1025, 2002. [Online]. Available: https://doi.org/10.1109/TPAMI.2002.1023799

[173] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds., 2014, pp. 2672–2680. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

[174] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "On the effectiveness of least squares generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2947–2960, 2019. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2872043

[175] D. Glaroudis, A. Iossifides, and P. Chatzimisios, "Survey, comparison and research challenges of IoT application protocols for smart farming," *Computer Networks*, vol. 168, p. 107037, 2020. [Online]. Available: https://doi.org/10.1016/j.comnet.2019.107037

[176] S. Munirathinam, "Industry 4.0: Industrial internet of things (IIOT)," in *Advances in Computers*. Elsevier, 2020, pp. 129–164.

[177] G. J. Joyia, , R. M. Liaqat, A. Farooq, and S. Rehman, "Internet of medical things (IOMT): Applications, benefits and future challenges in healthcare domain," *Journal of Communications*, 2017.

[178] M. J. Farooq and Q. Zhu, "On the secure and reconfigurable multi-layer network design for critical information dissemination in the internet of battlefield things (IoBT)," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2618–2632, Apr. 2018.

[179] R. Zheng, H. Wang, and J. Zhao, "A unified management framework for EIoT systems based on metadata and event detection," *IEEE Access*, vol. 7, pp. 112 629–112 638, 2019.

[180] A. Ross, S. Banerjee, and A. Chowdhury, "Security in smart cities: A brief review of digital forensic schemes for biometric data," *Pattern Recognition Letters*, vol. 138, pp. 346–354, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865520302555

[181] C. C. Sobin, "A survey on architecture, protocols and challenges in IoT," *Wireless Personal Communications*, vol. 112, no. 3, pp. 1383–1429, Jan. 2020.

[182] D. Orme, "Can biometrics secure the internet of things?" *Biometric Technology Today*, vol. 2019, no. 5, pp. 5–7, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969476519300682

[183] R. Srinivas, "10 iot security incidents that make you feel less secure," 2020. [Online]. Available: https://cisomag.eccouncil.org/10-iot-security-incidents-that-make-you-feel-less-secure/

[184] A. Maiti and M. Jadliwala, "Smart light-based information leakage attacks," *GetMobile: Mobile Computing and Communications*, vol. 24, no. 1, pp. 28–32, Aug. 2020.

[185] I. Shumailov, L. Simon, J. Yan, and R. Anderson, "Hearing your touch: A new acoustic side channel on smartphones," *CoRR*, vol. abs/1903.11137, 2019. [Online]. Available: http://arxiv.org/abs/1903.11137

[186] Gemalto, "A safer internet of things : Gemalto's guide to making the internet of things a safe place to connect," 2016. [Online]. Available: https://www.thalesgroup.com/sites/default/files/gemalto/iot-security-ebook.PDF

[187] Y.-Y. Leong and Y.-C. Chen, "Cyber risk cost and management in IoT devices-linked health insurance," *The Geneva Papers on Risk and Insurance - Issues and Practice*, vol. 45, no. 4, pp. 737–759, May 2020.

[188] P. Aufner, "The IoT security gap: a look down into the valley between threat models and their implementation," *International Journal of Information Security*, vol. 19, no. 1, pp. 3–14, Jun. 2019.

[189] A. Sengupta and S. Kundu, "Guest editorial securing IoT hardware: Threat models and reliable, low-power design solutions," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 12, pp. 3265–3267, Dec. 2017.

[190] K. Chen, S. Zhang, Z. Li, Y. Zhang, Q. Deng, S. Ray, and Y. Jin, "Internet-of-things security and vulnerabilities: Taxonomy, challenges, and practice," *Journal of Hardware and Systems Security*, vol. 2, no. 2, pp. 97–110, May 2018.

[191] I. Makhdoom, M. Abolhasan, J. Lipman, R. P. Liu, and W. Ni, "Anatomy of threats to the internet of things," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1636–1675, 2019.

[192] A. K. Sikder, G. Petracca, H. Aksu, T. Jaeger, and A. S. Uluagac, "A survey on sensor-based threats to internet-of-things (iot) devices and applications," 2018. [Online]. Available: https://arxiv.org/abs/1802.02041

[193] F. Meneghello, M. Calore, D. Zucchetto, M. Polese, and A. Zanella, "IoT: Internet of threats? a survey of practical security vulnerabilities in real IoT devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8182–8201, Oct. 2019.

[194] B. Klare, A. A. Paulino, and A. K. Jain, "Analysis of facial features in identical twins," in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, Oct. 2011.

[195] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters*, vol. 79, pp. 80–105, Aug. 2016.

[196] P. Punithavathi and S. Geetha, "Partial dct-based cancelable biometric authentication with security and privacy preservation for iot applications," *Multimedia Tools and Applications*, vol. 78, no. 18, p. 25487–25514, Sep. 2019. [Online]. Available: https://doi.org/10.1007/s11042-019-7617-1

[197] O. Olazabal, M. Gofman, Y. Bai, Y. Choi, N. Sandico, S. Mitra, and K. Pham, "Multimodal biometrics for enhanced IoT security," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, Jan. 2019.

[198] M. Shayan, M. Naser, and G. Hossein, "IoT-based anonymous authentication protocol using biometrics in smart homes," in *2019 16th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC)*. IEEE, Aug. 2019.

[199] A. Barros, D. Rosario, P. Resque, and E. Cerqueira, "Heart of IoT: ECG as biometric sign for authentication and identification," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, Jun. 2019.

[200] Q. Zhang, "Deep learning of electrocardiography dynamics for biometric human identification in era of IoT," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, Nov. 2018.

[201] S. Lee, J. Sa, H. Cho, and D. P. and, "Energy-efficient biometrics-based remote user authentication for mobile multimedia iot application," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 12, pp. 6152–6168, Dec. 2017.

[202] Z. Guo, N. Karimian, M. M. Tehranipoor, and D. Forte, "Hardware security meets biometrics for the age of IoT," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, May 2016.

[203] L. Janik, D. Chuda, and K. Burda, *SGFA: A Two-Factor Smartphone Authentication Mechanism Using Touch Behavioral Biometrics*. New York, NY, USA: Association for Computing Machinery, 2020, p. 35–42. [Online]. Available: https://doi.org/10.1145/3407982.3408021

[204] M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen, "Sensor-based continuous authentication of smartphones' users using behavioral biometrics: A contemporary survey," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 65–84, Jan. 2021.

[205] Y. Wang, C. Wu, K. Zheng, and X. Wang, "Improving reliability: User authentication on smartphones using keystroke biometrics," *IEEE Access*, vol. 7, pp. 26 218–26 228, 2019.

[206] H. Hamidi, "An approach to develop the smart health using internet of things and authentication based on biometric technology," *Future Generation Computer Systems*, vol. 91, pp. 434–449, Feb. 2019.

[207] A. Buriro, B. Crispo, and M. Conti, "AnswerAuth: A bimodal behavioral biometric-based user authentication scheme for smartphones," *Journal of Information Security and Applications*, vol. 44, pp. 89–103, Feb. 2019.

[208] S. Banerjee, C. Chunka, S. Sen, and R. S. Goswami, "An enhanced and secure biometric based user authentication scheme in wireless sensor networks using smart cards," *Wireless Personal Communications*, vol. 107, no. 1, p. 243–270, Jul. 2019. [Online]. Available: https://doi.org/10.1007/s11277-019-06252-x

[209] J. Cui, R. Sui, X. Zhang, H. Li, and N. Cao, "A biometrics-based remote user authentication scheme using smart cards," in *Cloud Computing and Security*. Springer International Publishing, 2018, pp. 531–542.

[210] D. W. Chadwick, R. Laborde, A. Oglaza, R. Venant, S. Wazan, and M. Nijjar, "Improved identity management with verifiable credentials and FIDO," *IEEE Communications Standards Magazine*, vol. 3, no. 4, pp. 14–20, Dec. 2019.

[211] R. Laborde, A. Oglaza, S. Wazan, F. Barrere, A. Benzekri, D. W. Chadwick, and R. Venant, "A user-centric identity management framework based on the w3c verifiable credentials and the FIDO universal authentication framework," in *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, Jan. 2020.

Appendix A

Biometric-based Secure Authentication for IoT Enabled Devices and Applications

Smart connected consumer devices employing IoT as the backbone are becoming a part of our day-to-day life. These products are fascinating to everyone but bear a dark side of becoming a threat to the consumers and the vendors. The users and the smart devices manufacturers are losing finances as well as confidential data to the adversary mainly due to employing insecure authentication methods. There are several reported incidences of such security breach in IoT enabled systems. The IoT industry finds it difficult to cope up with the fast developments and innovations in lightweight protocols, hardware devices, and authentication mechanisms developed explicitly for IoT based products. Furthermore, they are reluctant to embed these innovations with a fear that the consumers may not find the new product budget-friendly. Smart cities, smart homes, smart cars, smart grids, etc. are gaining attention from various sections of the society without any knowledge of the vulnerabilities that are associated with these revolutionary systems. This work considers the IoT system from the perspectives of a consumer, the vendor, and a researcher to figure out the present scenario, and give future directions to the authentication related security issues in IoT subsystems.

The device-to-user authentication while accessing connected consumer devices are the areas in IoT systems that need serious attention from the biometric research community and the biometric industry. Also, the automated payment system implemented in smart consumer products poses a threat from malicious attackers. Even though there are alliances to brainstorm the specific problem and standardization of protocols for IoT infrastructure, there is

slow growth in incorporating the most secure, and cost-effective solution to the security issues in IoT. We must understand the vulnerabilities and loopholes in IoT infrastructure and correctly design mitigation techniques to build a robust system. This work provides a precise investigation of the current scenario to integrate biometric authentication in IoT applications and systems along with the required techniques to mitigate software and hardware-level vulnerabilities in these systems. This work reviews the current research outcomes in this direction and pinpoints their pros and cons while implementing them into the future IoT products. It is also required to decide the best biometric modality and its implementation mechanisms that are convenient and pocket friendly to the consumers. In addition, this work also discusses various biometric traits from IoT perspective and suggest the best modalities for such systems.

A.1 Internet-of-Things (IoT) impacting our livelihood

The Internet has played a significant role in recent innovations and advancements in information technology. Internet-of-Things are referred to as a network of smart hardware devices (things) communicating with each other using the Internet to collectively provide specific functionality for individuals, industries, and organizations. With the advent of numerous IoT applications, human lives are more comfortable as such applications employ intelligent programmable consumer products to interact with people around them and amongst themselves. A few years back, what seemed to be an electronic device has now become a smart device by employing IoT as a backbone. The present-day world is witnessing the impact of IoT products and systems on everyday life as humans are increasingly becoming more reliant on the Internet for routine activities. Some sectors that employ IoT applications comprise healthcare, infrastructure, agriculture, logistics, manufacturing, automation industries, and many others [175].

A smart speaker wakes us up with our favorite music in the morning; a smart wristband keeps the record of our health, and suggests a healthy diet; a smart assistant reminds us about the daily schedule and appointments; the list continues to grow. With the advent of smart cities, facilities, such as a smart home, smart grid, smart cars, smart street lights and

traffic signals, smart meters, and smart TVs, are influencing human lives in some or the other way. The vendor provides a smart phone app for smart consumer appliances to connect and control the device conveniently through the Internet. Such devices employ sensors, whereas the smart phone app performs local analytics. In case the reports of the analytics need to be shared with multiple parties, the system uses a cloud-based server for analytics. A smart wristband gives health updates using different sensors to the smart phone applications. In such cases, local analytics in the application is sufficient. A health monitoring system for an older person living alone may require the data processing at the cloud server since a physician, a caretaker, or a relative may be sharing the reports.

The automation and transportation industry extensively uses radio-frequency identification (RFID) tags. The tracking of assets, paying toll charges, and managing inventory are some more application areas of these tags. These tags contain digital information and require a reader device to collect this information through radio waves. Near-Field-Communication (NFC) tags offer low-range communication over a low-speed channel. The vendors employ NFC tags for product labeling and electronic payment purposes. The application of IoT to the specific field is named accordingly. For instance, industrial IoT (IIoT) [176], Internet of Medical Things (IoMT) [177], Internet of Battlefield Things (IoBT) [178], Environmental Internet of Things (EIoT) [179] are some examples of explicit IoT applications. The IoT based home and consumer appliances include HomePod, HomeKit, Siri, Apple Watch from Apple Inc., Amazon's Echo and Alexa, Google's Nest and Google Home. Figure A.1 provides a glimpse of various IoT-enabled applications and services in diverse areas.

A.2 IoT ecosystem

We should be aware of the functioning inside a typical IoT system to figure out the vulnerabilities and threats associated with it. Figure A.2 shows major components in the IoT ecosystem. The elements and communication channels in the IoT system can be roughly classified into six categories, namely the environment, the things directly interacting with the environment, the low-range LAN communication protocols employed for intranet data transfer, the local control centre, i.e., the IoT gateways, the Internet, the remote servers, and

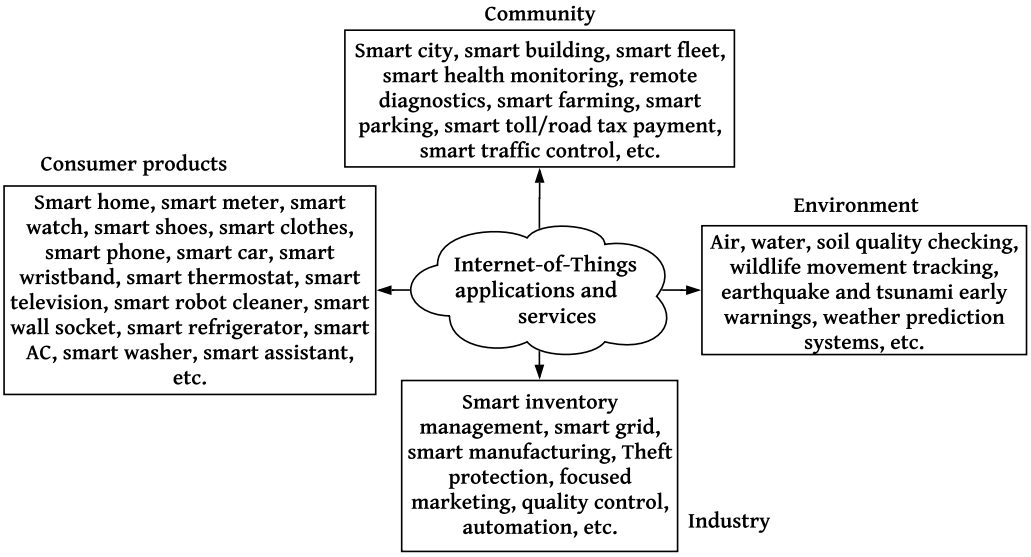


Figure A.1: IoT-enabled applications and services

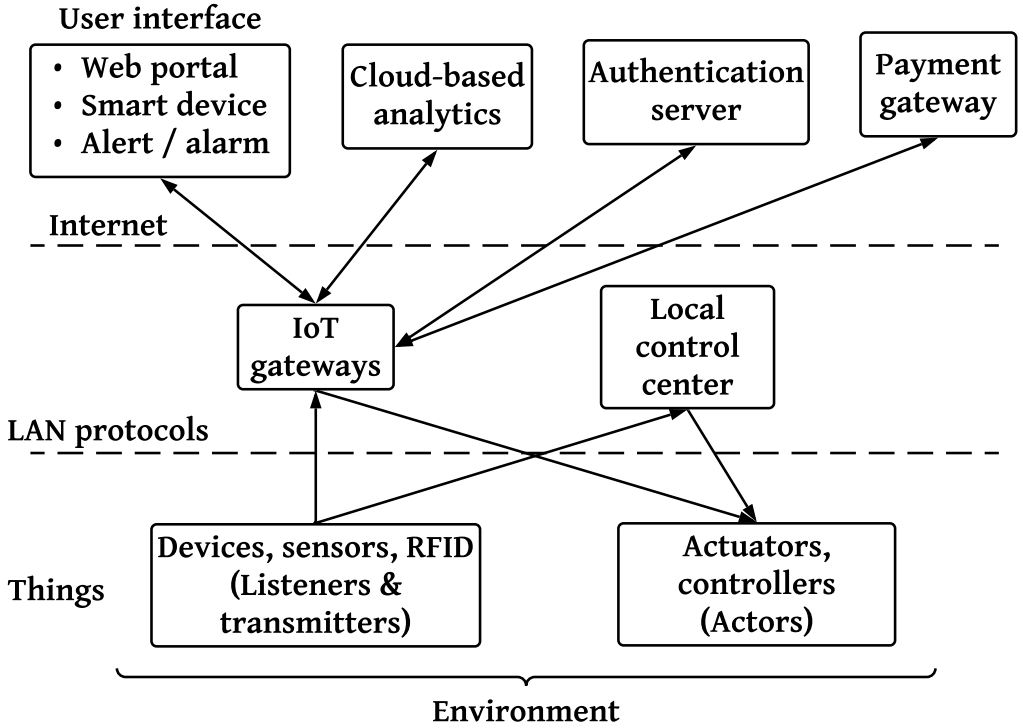


Figure A.2: Major components in the IoT ecosystem

user interfaces.

The things in an IoT ecosystem comprise the hardware-based embedded systems that listen and respond to their environment. The interface to the environment comprises sen-

sors (acting like human nerves), RFID tags, controllers, and actuators (acting like human muscles). In general, sensors exist for capturing physical observable, such as temperature, ambient light, humidity, dust, fire, motion, smoke, color, water, etc. [180]. These sensors continuously collect the data for the desired parameter and, send it to the IoT gateway and the local control system without any delay. Sensors, typically, use LAN protocols for transmitting data. RFID tags use a similar approach to send smart bar-codes embedded on them through radio frequency technology. An RFID reader can sense, collect, and read the information from the tag. In both these cases, the communication is one-way. Hence, these devices are referred to as *listeners* and *transmitters*. Furthermore, certain things in the IoT system intend to respond to the environment based on collected data or the instructions received over the Internet termed as *actuators* and *controllers*. They act by following the pre-programmed actions based on the commands received. These devices can only receive the data through the LAN protocols and usually do not respond to the source of information. Such devices are termed as *actors*. The communication within the local IoT subsystem among things, gateways, and local control centers use low-range wireless protocols. Some of these protocols providing low-range communication include Bluetooth, ZigBee, Wi-Fi, BLE, NFC, etc.

IoT gateways provide the communication link between the local IoT subsystem and the remote cloud-based servers, user interfaces, and data analytics [181]. These devices can also employ local data analysis for instant decision making on the incoming data stream from the sensors. The data analysis and control capabilities of the IoT gateways can sometimes be shared using an explicit local control centre. Local network management, system diagnostics, device configuration management are some additional functions of IoT gateways and local control centers. In a typical IoT environment, the remote user authentication and data analytics servers connect with the IoT infrastructure with the help of the TCP/IP based Internet. The system usually provides a user interface in the form of a web portal, smart phone application, and alert/alarm application for the customers and the system administrators to process the raw data, and visualize the data analysis. A dedicated authentication server performs security checks such as user authentication. Some IoT applications such as smart parking, food ordering, paying toll charges, etc. and periodic subscription-based services also employ a payment gateway.

Despite so many features and utilities of IoT systems, such systems also face vulnerabilities and threats associated with various components and communication channels in the entire IoT ecosystem. Experts in multiple applications have reported several incidences of financial losses due to credit and debit card frauds in recent times [182]. In IoT systems users also faced threat to data confidentiality and data integrity. Security breaches in IoT other than financial frauds comprise insecure mobile interface, privacy issues, insecure cloud and web interfaces, etc. Such threats must be addressed and monitored closely to understand the risk intensity and accordingly provide mitigation mechanisms against them.

A.3 Classification of IoT-powered applications and services

In the present-day world, a considerable number of IoT-enabled consumer products and services are in use. With the emerging evolution of IoT technology, we can find several types of such devices around us. We classify these smart devices into seven categories based on their working mechanism, as follows.

1. *Sensing only* : The sensing only IoT devices consists of sensors located at different places to collect and send the data to the server. The server stores the data and a system administrator retrieves the data whenever needed and presents the data in its original form. Smart cities monitoring temperature, oxygen, humidity, dust, and carbon monoxide uses sensor-enabled infrastructure and display the air quality parameter values periodically to let the citizens know the best and worst colonies in the city. Health monitoring applications for a bed-ridden patient at home or hospital can also use sensors to record various health parameters, such as pulse rate, blood pressure, temperature, and other critical health parameters, and a display device shows variations in such parameters.

2. *Sensing and real-time local analysis* : IoT systems wherein the sensors collect data from the environment and perform real-time analysis of the data at the local gateway fall under this category. Weather prediction systems employ various sensors to collect data about temperature, humidity, and other environmental factors, and perform local data analysis to predict a rainfall, forecast temperature variations in the locality. A

vehicle maintenance system uses sensors to monitor the condition of critical spare parts and shares the analysis report of the respective vehicle with driver and maintenance staff.

3. *Sensing and cloud-based data analytics* : Occasionally, data analysis can occur at a centralized cloud server for sensors deployed in a geographically large area, such as a state or a country. The train track monitoring system uses sensors at a short distance over the entire tracks in a country. An unexpected event on the tracks may lead to a major accident or interrupt the rail traffic for a longer duration until the damage gets repaired. Hence, to divert the trains to an alternate route and stop trains approaching the faults, cloud-analytics embedded in the system collect the data, and provide reports to different authorities at the station and onboard train staff based on real-time locations of trains.

4. *Sensing and analysis of data with automated control* : The IoT systems can use rule-based automatic control mechanisms in case of an alarming situation. In smart lighting and home systems, the actuators play a significant role in responding to the environment based on rules and instructions according to the underlying algorithm. The electrical appliances switches ON/OFF based on the presence of an individual sensed by the motion detectors. Consequently, the air conditioning system can adjust the temperature based on the number of individuals present in a room perceived by the CO_2 level inside the room.

5. *Sensing and analysis of data with manual control* : It may not always be possible for the system itself to deal with every unexpected situation and thus needs manual intervention. The sensors and the analytics can show a damaged node within a network that needs replacement or repairing urgently. In another scenario, a smart phone app can also provide an interface to display a probable malicious node and allow the consumer or administrator to either restart the node or stop transmitting the sensed data. Suppose, the system detects unusual data collected at the gateway or cloud-server, and the system cannot locate the hacked sensor. In such incidences, the network requires manual intervention.

6. *Sensing and analysis of data with automatic and manual control* : We usually provide intelligence to the smart devices so that they can respond quickly to certain situations. But they are still not intelligent enough to differentiate and respond to a more severe event that is within their capability. Smart cameras installed at the traffic signals can automatically zoom at the registration numbers of the vehicles of the traffic rule violators and send images to the server. But in case there is an accident, the traffic police seating at the control room can control the camera to zoom in at the offender of the incident than allowing the cameras to operate normally.

7. *Smart devices (artificial intelligence (AI) + machine learning (ML) + rule-based and manual control)* : The smart consumer appliance usually embeds the latest hardware and software technology such as intelligent sensors, artificial intelligence, and machine learning algorithms to provide state-of-the-art experience to the consumer. These devices learn, adapt, and perform better with increasing consumer interactions. They can place orders and pay on behalf of the consumer, automatically sense the consumer mood and play songs accordingly, read the news, receive calls, set reminders, and many more.

A.4 IoT security breach

Smart TV, printers, smart security cameras are vulnerable to a zero-day attack [183]. Bitdefender discovered that the Ring Doorbell cameras from Amazon were allowing hackers to access a user's Wi-Fi as well as other devices connected through the Wi-Fi. It was also possible to mount a distributed denial-of-service (DDoS) attack via Blink XT2 security camera systems from Amazon [183]. Additionally, the cameras provided access to footage from the camera and the audio output. Check Point researchers demonstrated that fax machines are vulnerable to hacking using the fax number and the telephone line [183]. They also illustrated the exploitation of a security bug in HP all-in-one printers during a conference [183]. The FBI stated that the camera and microphone integrated into most of the smart TV manufacturers could be hacked to control volume and even channels [183]. Researchers also illustrated the information leakage based attacks on smart light controlled using infrared [184].

A news channel in the US reported an incidence of hacking a smart home and a thermostat of a couple [183]. Researchers from academic institutions developed malware to steal confidential data from a smart phone via a hacked microphone from the device [185]. They also demonstrated the acoustic side-channel attack on a touch-screen device which unveils everything a user types. Smart coffee machines possess high-security risks as their vendors give minimal efforts on its security aspects while designing. The apps for these machines are vulnerable to reveal information about the consumer's bank and cards [183]. It is also observed that LAN printers pose a risk for cyber attacks into the organizations [183]. Researchers hacked the smart speaker, Amazon Echo, during a live demonstration at a security conference [183]. As per Trend Micro researchers, IoT-based cyber attacks are possible via Internet-connected gas stations [183].

During a demonstration by the researchers from the University of Central Florida, a Nest Learning thermostat was hacked within fifteen seconds when the hackers were allowed to access the device physically [186]. The hacker then employed the thermostat to expose the Wi-Fi credentials, spy the consumer, and attack other devices connected with the same wireless network [186]. It was also possible to geographically track the movement of fitness trackers by employing a customized Raspberry Pi [186]. Further, experts successfully managed to send spam and phishing emails using an Internet-connected refrigerator [186]. There was also an incident of the hacking in-flight entertainment system of an aeroplane [186]. A professor reported that a hi-tech train signaling system was prone to hacking, prompting severe consequences in UK [186]. It was also claimed that more than 0.45 million connected vehicles are vulnerable to intrusion attacks if the attacker gets access to their IP addresses [186].

The insurance firms employing IoT-enabled services and devices possess a high risk of cyber attacks [187]. More than 34% of the present-day Internet traffic accounts for the cyber threat. An average financial loss due to IoT related hacks crossed 8 million USD in 2019 [187]. Most of the security breaches reported in the past are the consequences of some critical misconceptions about IoT security, e.g., constrained IoT things poses no risk to the system as they transfer unimportant raw data. However, authentication and authorization mechanisms are sufficient for securing the entire system, and it is enough to employ threat detection mechanism as a device reset can stop such attempt immediately [187].

A.5 IoT threat model and mitigation approaches

Wireless Sensor Network (WSN) forms the basis, and the Internet provides the backbone for any IoT-based smart devices and services. Hence, the vulnerabilities these technologies exhibit may also apply to the IoT infrastructure. Denial-of-Service (DoS), Hello flooding, Sybil, and Sinkhole attacks usually target a WSN [188]. The wireless network connecting the sensors, actuators, IoT gateways in an IoT infrastructure may possess similar risk. The proprietary protocols employed in such systems may not be secure and robust enough to resist all the probable existing attacks on the network. A bug in a smart device such as a web camera and a smart speaker may become a threat to the entire system using the same infrastructure.

The incidences of IoT security breaches indicate that the IoT enabled consumer appliances, smart devices, and services are still vulnerable to attacks, such as hacking, data interception, etc. As there are several components and communication channels connecting them, the chances to get exposed to any attack attempt remains high. So, we should identify all the possibilities for a probable attack and categorize them such that a set of mitigation techniques can protect and secure the IoT infrastructure. Based on the major components involved in an IoT ecosystem discussed in Figure A.2, we propose an IoT threat model with ten vulnerabilities. Figure A.3 shows the proposed threat model, and the naming convention V_1 to V_{10} depicts the ten vulnerabilities.

The class of threats associated with listeners and transmitters fall under vulnerabilities V_1. The adversaries hack these devices and divert the traffic to the fraudulent server. He may also inject some malicious contents into the network to harm the infrastructure. A *hardware Trojan* is a type of deliberate insertion into hardware design [189]. Usually, an act of a rogue designer or vendor can lead to such a security breach in the system. He can also mount a denial-of-service (DoS) attack through the IoT nodes. These memory-constrained devices are also vulnerable to side-channel attacks. The attacker analyzes computation time and electric emission to collect confidential information such as encryption key or authentication secrets. As these constrained devices are usually battery-powered, it is not feasible to provide additional software or hardware just for security motives.

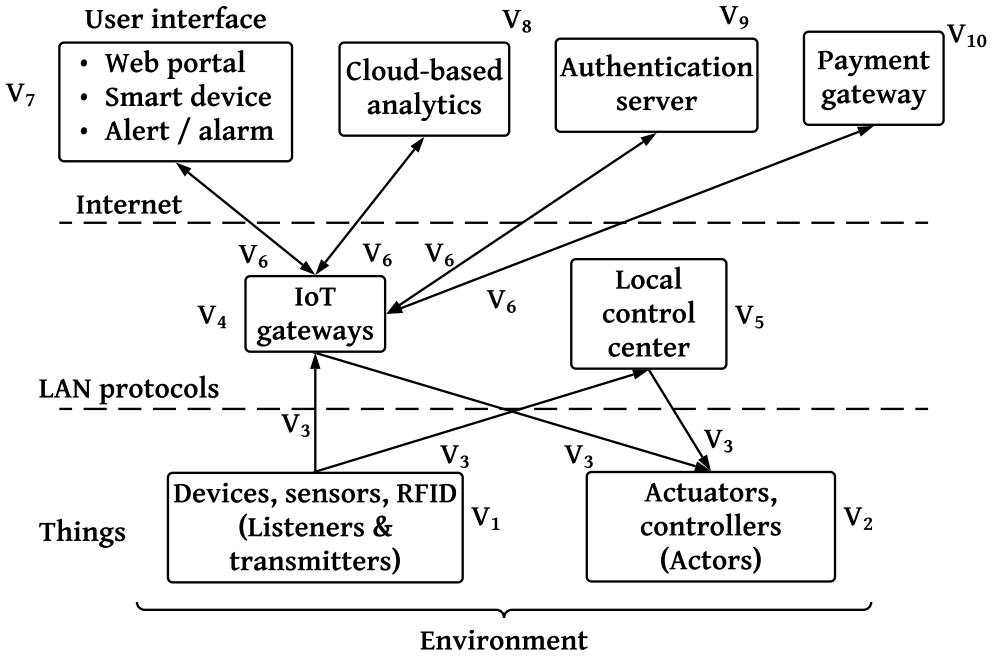


Figure A.3: IoT threat model

The possible attacks at nodes that directly interact with the IoT environment, such as actuators and controllers, fall under vulnerabilities V_2. The hacker targets devices, such as actuators and controllers, responsible for acting on behalf of the user or the automated control mechanism in the IoT system to perform an unintended activity. The adversary often employs tools such as oscilloscope, logic analyzer, and ChipWhisperer to figure out the vulnerabilities in the target node. Usually, such incidences are noticeable, and the attacker has very little to achieve in the long term except destroying or damaging the components on the IoT infrastructure. Hence, an intelligent adversary invests least time and effort on such attempts. However, security experts must give necessary consideration to probable threats at such nodes.

Typically, the nodes in the IoT infrastructure communicate via technologies, such as Wi-Fi, fiber, Ethernet, ZigBee, 3G, 4G, Bluetooth LE, etc. We categorize the threats to these communication media as vulnerabilities V_3. The adversary can retrieve secret information or perform traffic analysis by eavesdropping over the communication medium. Once the attacker obtains some confidential data, he uses it to mount a replay attack at a later time. Packet flooding in such LAN is a form of denial-of-service attack. If the hacker succeeds

in controlling the IoT nodes, he can mount several other attacks, such as man-in-the-middle (MITM) and Sybil attack, as the entire network will be at his fingertips [190]. Consequently, he can exploit the vulnerabilities in the communication protocol and the nodes to destroy or damage the whole system.

In a distributed denial-of-service (DDoS) attack scenario, a set of geographically dispersed computers target different nodes within a network infrastructure to ultimately deny any services to the authorized users. During a Sybil attack in a WSN, the adversary employs the compromised node to mislead a victim node by presenting multiple identities. The victim node, in turn, executes the same instruction or operation redundantly. Sniffer instruments are employed to collect network-related information, communication patterns, physical locations of various wireless access points, and the protocols used in the network [190]. A Sinkhole attack targets the network infrastructure in the IoT system. Here, a malicious node gathers data from its nearby nodes and bypasses all other communication links without any hint to the system [190].

Usually, the IoT gateways possess high processing power compared to other nodes in the LAN, which is sufficient enough to execute critical and intensive applications. This empowerment provides more opportunities for the attacker to succeed in mounting an attack. IoT gateways located at the intersection of LAN and the Internet vulnerable to software or hardware level threat becomes an entry point into the entire system. IoT gateways are susceptible to data leakage and topology disclosure. In data leakage, the adversary manages to collect the data from the local storage at a node or by diverting the traffic from the victim node. As every sensor, actuator, and controller device within the network communicates with the gateway, a malicious gateway can consequently disclose the location and identity of these nodes. We label the probable attacks at IoT gateways as V_4. The attacks targeting the local control center within an IoT LAN fall under the category of vulnerabilities V_5. The threats to the IoT gateways also apply to the local center. Additionally, they both can be a possible target of a Trojan horse attempt.

The IoT gateway communicates with various remote servers, cloud-based services, and user interfaces using the Internet. Even though we have a well-established Internet service throughout the world, it is still vulnerable to specific risks and so cannot be labeled as fully secure communication medium. An adversary can execute a DoS attack, man-in-the-middle

(MiTM) attack, eavesdropping, selective forwarding, Sybil attack, channel congestion, and collision attack on Internet services [191]. We categorize all such threats through the Internet traffic under V_6.

The service provider and smart product vendor usually provide a user interface for the consumer and various teams working at the cloud server. Web portals, smart phone apps, and plug-in for alarm or alert mechanism are a few examples of such interfaces. The vulnerabilities associated with them is categorized as V_7. The hacker can use another app or operating system bug to capture the smart phone screen or read the app data from system memory. An insecure web application portal or APIs can disclose secret user information or delete some files.

Almost every IoT application requires data analytics, and the vendors prefer cloud-based services for the same. But the application, in turn, draws in all the risks associated with cloud servers to its environment. Several attacks have been mounted on cloud servers, which comprise flooding attack, cloud malware injection, SQL injection attack, and signature wrapping attack [190]. The adversary tries to control the cloud services by injecting malware, a malicious service instance, or virtual machine on the cloud. He can also modify the XML signature commonly employed by cloud servers for ensuring service integrity. A malicious SQL query code for updating, deleting or reading the database contents also poses a potent threat to cloud infrastructure. The adversary uses the Internet for mounting a DoS attack on the cloud through flooding. Such risks to cloud-analytics servers are categorized as V_8.

An authentication server in an IoT environment ensures that only authorized individuals are allowed to access the system resources and services. Usually, these servers are maintained separately away from the LAN for security issues. Such remote servers are vulnerable to leaking the user credentials, false data injection, eavesdropping, record delete or update, identity theft, password, key or session token disclosure. These types of threats to the remote authentication server are categorized as vulnerabilities V_9.

Smart devices allow consumers to place orders and pay the merchant via smart card whose credentials are stored at the vendor's server. The IoT service provider maintains a payment gateway remotely for providing a secure payment process. The adversary targets these gateways for performing financial frauds. He intercepts the traffic to such servers to collect user credentials and utilizes them for his benefits or places unnecessary orders on be-

half of the smart device. He can also execute DoS, DDoS, false data injection attempts on the gateway. We classify such threats to the payment gateways under the class of vulnerabilities V_{10} in the proposed threat model.

The security mechanisms at different levels in the IoT infrastructure ensure that the system components and communication links connecting them are least vulnerable to known threats. The possibility of building an IoT application or service free from all existing and unforeseen risks cannot be guaranteed. However, the best consumer product or system should be resistant to a large number of most common threats that incurs the least loss in terms of finances, data, and user privacy. As we can broadly categorize the IoT components into three categories, namely hardware, software, and communication medium, we present the countermeasures for the IoT system in these categories to enhance the security and thus build a robust application or service from scratch.

The hardware devices should be tamper-proof and employ code signing to avoid any possibility of a Trojan horse attack. Hardware security features such as ARM TrustZone ensures secure data flow within the devices [192]. Designing ICs with active shields protects them from probable side-channel attacks. An X.509 Digital Certificate should be employed to identify each node. Subsequently, the communication between the nodes should use HTTPS or NTLS protocols. Further, each node can verify the identity of every other trusted node. Additional security measures include embedding a Trusted Platform Module (TPM) device, usage of a PUF (Physical Unclonable Function), randomize instruction execution cycles, using lightweight hardware implementation of a cipher, network segmentation, and device registry.

The strategy to secure the communication channel includes encryption using a hash function [193]. The authentication mechanism should use message authentication codes (MAC), digital signatures, and hash functions. A pseudo-random number generator that satisfies a majority of randomness tests can further enhance the security of communication by generating asymmetric keys and lower the chance of a replay attack. We can reduce the risk of most software-based threats by adapting lightweight intrusion detection methods, software-defined networking (SDN), ensuring software integrity during updates, auditing via log management for each update. Cloud-based IoT systems should implement homomorphic encryption and Cloud Access Security Broker (CASB) for providing security and privacy protection

in cloud-based services. Blockchain technology can also protect against replay attack, ransomware, and malware.

The threats to various communication links, software modules, and hardware components have different intensities in terms of financial and data loss to the consumer and the service provider. Smart devices have touched almost every aspect of human life, and subsequently, we are inviting their probable dangers into our lives, exposing everything we own. Since not every consumer is technical savvy, they believe in every news and reviews available online and form misconceptions. We addressed the possible risks associated with smart devices and also specified the ways to mitigate them. In today's digital world, a consumer excited about smart devices must become smart enough to understand the know-how about every such device in their possession.

A.6 Biometrics for IoT security

The biometric field has evolved into diverse applications over the past fifty years. The system that initially gained popularity as a mere verification or identification system found suitable for various problems. Present day use cases comprising, border security, access control, forensic investigations, controlling child trafficking, monitoring infant vaccination, channelizing government schemes for the underprivileged citizens, home security, university attendance system, banking services such as cash dispenser machines employ biometric systems. These systems found scope in security and privacy applications along with identity management. Present day smart phones, laptops and other handheld devices, smart homes are few recent examples that incorporate biometric recognition. The researchers are now looking forward to figuring out the possibility of integrating biometric into IoT-based devices and systems. Figure A.4 provides a glimpse of biometric-enabled applications.

The advantage of biometric-based user authentication to the IoT system lies in its uniqueness. The user gets relieved from worries of forgetting his credentials for authentication for almost a lifetime. In case his biometrics is stolen by some means we have cancellable biometrics which regenerates another authentication data from the same biometric. Since there are multiple traits that an individual poses, the system can be made more secure by employ-

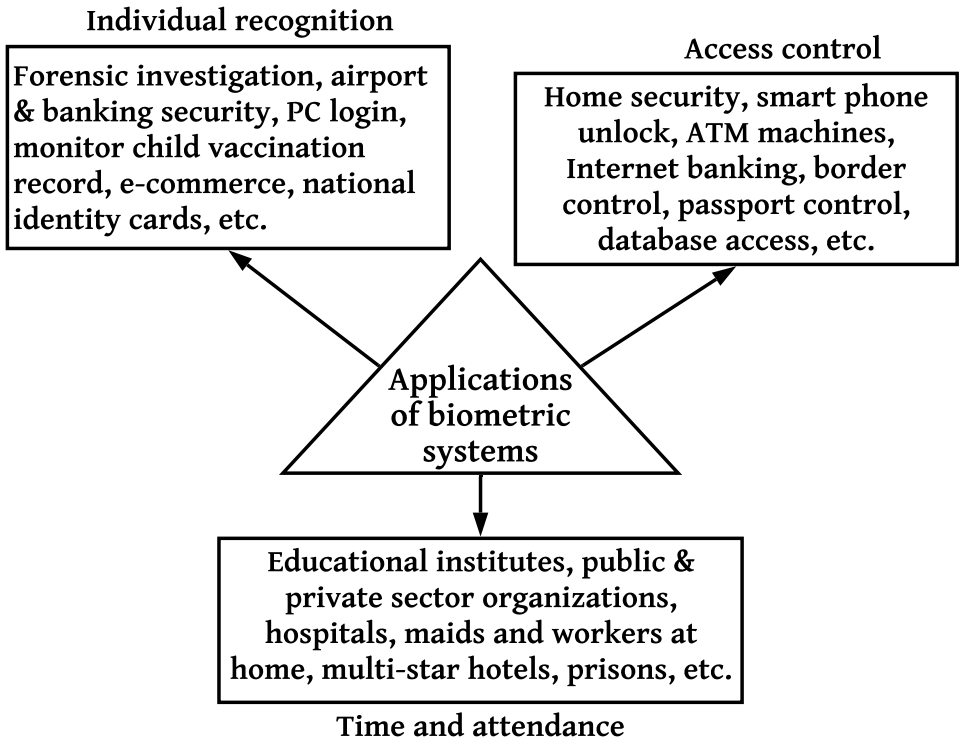


Figure A.4: Applications areas of biometric systems

ing a combination of these traits. Also, the template protection mechanisms for storing the encrypted biometric data on the server assure that the information about the biometric features or patterns of a user is entirely confidential. Hence, even if the attacker gets access to the template, he will find it infeasible to know the biometric details of the owner.

There exists a high possibility to employ biometric-based user authentication for IoT devices and systems. However, as we have multiple characteristics of an individual for unique identification and authentication, we must decide the most feasible trait for a specific application. Facial recognition systems may not differentiate identical twins or in some incidences even face masks can fool such systems [182], [194]. Since social media has entered into our everyday life, we are becoming less hesitant to post our and family photos on such platforms. Hence, we should avoid facial authentication for financial transactions or critical applications. Voice-based authentication may not distinguish a recorded voice and thus fail to provide the expected level of security. We have modalities such as voice, ear, gait, etc. which are still in the experimental stage [182].

The fingerprint of an individual is unique, and till date, no record exists for two individ-

uals with similar fingerprint patterns. Moreover, the research in this domain already crossed fifty years recently, and we have highly accurate fingerprint biometric systems available in the market [195]. Additionally, compared to other modalities, the fingerprint biometric system requires the least assistance from the user, and it is comfortable to operate even for a layperson. Hence, being completely mature, efficient, and convenient among all the characteristics, these systems are employed for user identification and authentication systems at various government, private, and even high-security applications. Fingerprint-based biometric authentication thus wins the race for being the most suitable and practical option for the IoT environment.

A cancelable biometric authentication can help in preserving security and privacy for IoT-based applications [196]. Similarly, the multimodal biometric system proved to render enhanced security and improved accuracy while authorizing an individual for accessing IoT network [197]. The biometric-powered anonymous user authentication scheme performing lightweight operations in IoT infrastructure proves to be efficient [198]. The mechanism is best suited for smart homes, and it assures that the rest of smart devices inside the house remain unaffected even if the hacker breaks into the security of the smart home. The user may choose to use ECG (electrocardiography) as a biometric trait for authentication into the IoT system [199], [200]. The face recognition approach can provide user authentication on smart phone devices [201].

A novel combination of system-level obfuscation method along with electrocardiogram (ECG) and photoplethysmograph (PPG) biometrics solves the issue of unauthorized access to IoT nodes [202]. The approach also helps in preventing any possibility of tampering and reverse engineering. The research also shows the use of behavioral biometrics to provide secure and authorized access to the IoT system components [203], [204]. Also, the keystroke biometrics has demonstrated promising results when targeted for user authentication [205]. We can employ these approaches, especially for smart phone authentication. Smart health care system also needs authorized access exclusively by medical practitioners and family members. The biometric-based system can be a feasible solution in such scenarios [206]. A bimodal biometric authentication provides enhanced security on smart phone devices [207].

The use of the smart card for storing the biometric template mitigates several attacks associated with the templates and requires no additional remote template database server. A sim-

ilar approach can also help in secure and authorized access to the IoT resources [208], [209]. Additionally, if the system employs a SoC, then most of the vulnerabilities associated with the biometric system will be eliminated. The consumer would find it convenient to carry his smartcard and get rid of the several issues related to traditional authentication mechanisms. Biometric-based access control systems would benefit the vendors and buyers in multiple folds. Hence, sooner or later, IoT enabled smart devices and services would offer such solutions to the consumers.

The Fast Identity Online (FIDO) Alliance provides solutions for replacing conventional approaches of identity management (IM) in a more secure, convenient, and feasible manner [210]. The alliance includes more than 200 industry leaders in software and hardware sectors. Their first framework provides smart devices with a password-free authentication. In another protocol, the coalition presents a small hardware token for implementing two-factor authentication. An asymmetric encryption method forms the base for authentication under both approaches. The Verifiable Credentials (VC) data model proposed by the World Wide Web Consortium (W3C) is also a similar initiative to identify better user-centric solutions for the identity ecosystem [211]. These solutions are decentralized digital identity systems and are presently available for real-time use in smart devices and IoT-based user authentication. As there is a profound demand for a standard protocol across various consumer products and services, different market leaders are collaborating with the alliance and incorporating the strong authentication standards as a security mechanism for multiple products and services offered by them.

We can employ a fingerprint-based biometric system over other modalities to solve the problem of user-to-device authentication as well as allowing only authorized individuals to access the data and services concerning vulnerability V_9 from the threat model proposed in Figure A.3. It is also possible that the consumers appreciate such a move as most of them might have already used a fingerprint-based biometric system in the past. We are living in a rapidly progressing world where privacy and security need utmost attention to ensure that the future generation finds it safe to use any smart devices and services. A fingerprint biometric system would meet the requirement of the time and become a solution to authentication-related queries of the industry innovations.

A.7 Summary of the appendix

We are living in a world of information and communication technology (ICT). The advancements and innovations powered by ICT have made a significant impact on our daily routine. We are so connected with this new Internet world that people feel uncomfortable if their Internet goes down even for a few minutes. Hence, Internet service providers (ISP) worldwide are exercising every possible step to provide uninterrupted and high-speed connectivity to their consumers, even in a remote area. The notion of Internet-of-Things is the outcome of a sufficiently mature ICT. This work provided a broad review of the current scenario of Internet-of-Things technology. It covered various types of IoT devices, and services and presented differences between them in terms of their functionalities. This work also mentioned the security issues in these devices and suggested various corrective and mitigation steps against any probable threats to such system. The study also addressed the need for a highly secure authentication mechanism like a biometric system to authorize and authenticate an individual. It emphasized the benefits of employing a fingerprint-based biometric system in an IoT infrastructure. This work provided fine-tuned contents such that the readers understand the significance of every term specific to IoT and biometric systems. In a nutshell, this work introduced the reader with IoT and biometric system from a security perspective and encouraged them to address various threats to them.