Chapter 5

Soft Biometrics and Privacy

5.1 Introduction

Soft biometric traits are certain human characteristics which provide some information about the individual but lack the distinctiveness and permanence to sufficiently differentiate any two individuals [33]. Typical examples of such traits include height, weight, skin color, eye color, age, ethnicity, gender etc. When such pieces of information are used in combination with primary biometric traits in a multimodal framework, the overall performance of the biometric system potentially improves [154].

All the soft biometric properties are generally categorized in *face*, *body* or *accessory* groups. As evident from the names, a soft biometric trait is grouped under one of these categories according to its association to the human face, the human body or external accessories respectively. In addition to these rudimentary classifications, the

Soft Trait	Category	Nature of Value	Permanence	Distinctiveness	Subject Perception
Skin Color	Face	Continuous	Medium	Low	Medium
Hair Color	Face	Continuous	Medium	Medium	Medium
Eye Color	Face	Continuous	High	Medium	Medium
Beard	Face	Binary	Low	Medium	Low
Moustache	Face	Binary	Low	Medium	Low
Facial Measures	Face	Continuous	High	Medium	Medium
Facial Shapes	Face	Discrete	High	High	High
Facial Feature Measures	Face	Continuous	High	High	Medium
Facial Feature Shapes	Face	Discrete	High	High	High
Ethnicity	Face	Discrete	High	Medium	Medium
Marks	Face/Body	Discrete	High	Medium	Low
Gender	Face/Body	Binary	High	Low	Low
Age	Face/Body	Continuous	Low	Medium	Medium
Height	Body	Continuous	Medium	Medium	Medium
Weight	Body	Continuous	Low	Medium	Medium
Gait	Body	Continuous	Medium	Medium	High
Body Measures	Body	Continuous	Medium	Medium	Medium
Body Shape	Body	Discrete	Medium	Medium	Medium
Clothes Color	Accessory	Discrete	Low	Medium	Medium
Glasses	Accessory	Binary	Low	Low	Low

TABLE 5.1: Soft biometric traits and their associated properties

soft biometric traits are also adjudicated based on the properties of *nature of value*, *permanence*, *distinctiveness* and *subject perception*.Nature of value refers to the fact that whether any particular trait is continuous or discreet, permanence indicates the ability of the trait to remain invariant with time, distinctiveness signifies the degree of variation of the trait between subjects and subject perception points to the ability of humans to unambiguously recognize the specific trait. All major soft biometric traits with their associated properties are shown in Table 5.1 [155].

It is noticeable that the importance or significance of a particular soft trait depends upon the nature of application it is used for.

5.1.1 Applications of Soft Biometrics

Investigations into the potential uses of soft biometric traits have been undertaken comparatively recently. The first problem addressed in this domain was designing appropriate fusion techniques for primary and soft biometric traits. The Bayesian decision theoretic framework [33] was the first proposed fusion model, which subsequently has become a standard approach. The authors used the combined soft characteristics of gender, ethnicity and height in [156], thereby reporting an improvement of 6.1% in the GAR over standalone fingerprint recognition systems. Another similar study regarding the same soft labels was performed in [157], but this time employing both fingerprints and face data as the primary biometric characteristics. A performance improvement of almost 10% was observed here. More recently, a set of soft characteristics consisting of gender, blood group, weight and height was fused with ear data for newborn babies [158]. Ailisto et al. [159] obtained comparable success using body weight and fat measurements to improve the basic fingerprint recognition. Herein, the total error rate of 62 test subjects was reduced from 3.9% to 1.5%. The use of uncommon soft biometric characteristics was observed in [160] and subsequently implemented in [16]. The central idea of the authors revolved around identifying people using their facial marks like scars, moles, acne and wrinkles. These marks were described as prominently localized regions on the face and subsequently blob detectors based on the Laplacian of Gaussian were used for their detection.

In another direction, works like [161, 162] dealt with automatic extraction of soft biometric features from videos. Mining of these characteristics was studied in details by Dantcheva et.al [155] wherein a set of soft labels based on nine semantic characteristics were proposed. These characteristics were mainly based on facial data (e.g. beard, glasses etc.), body measures (torso and legs) and color of clothes. In another work with similar objectives [163], the authors proposed an approach for weight computation and height estimation from video frames in an unobtrusive manner. Their technique essentially utilized elements of body measurements via Artificial Neural Networks (ANN). In [164], the authors predicted some soft biometric labels categorized under pedestrian attributes (e.g. gender, age, clothing, hair color etc) by utilizing a Multi-Label Convolutional Neural Network (MLCNN). The novelty in their approach lay in the fact that they did not assume independence of attributes during their predictions, but rather developed their framework in an integrated form. Effects of facial soft biometric features on the recognition accuracy of some existing state-of-the-art face recognition algorithms (Local Region PCA and Cohort Linear Discriminant Analysis) were recently studied in [165]. The results from this exclusively facial biometric based study, however, indicate a modest increase in the overall accuracy rates. The soft traits which the authors fused with the existing algorithms primarily consisted of gender, race, eye color, hair color and eyebrow.

The inclusion of soft biometrics was also considered to address the problem of recognition at a distance. The most popular primary characteristic which is used for identification-at-a-distance is the face [166]. However, such system suffers much in

term of accuracy mainly due to the poor sensor quality of most CCTV cameras. To enhance the overall performance, Tome et.al [167] extracted and combined 23 soft biometric features from the captured camera images. The results indicate a comprehensive improvement over normal face identification schemes based on Viola Jones and FaceSDK (a commercial system). The classification and labeling tasks of the aforementioned 23 unique characteristics were originally formulated in [168, 169]. This same categorization was later used in [170] as a basis for semantic biometrics where the patterns and structures within a physical description of biometric data were exploited. This was done for accurate prediction of occluded and erroneous data. With more robust ambient settings, Denman et.al [171] demonstrated the utility of soft biometrics for measuring operational information. Specifically, they developed their techniques to adapt in a multi-camera surveillance network (e.g. airports). The concept of 'average soft biometrics' was at the core of their designs wherein average histograms were obtained by summing the histograms of the individual subject biometrics. The soft traits which they extracted in their work consisted of the color model (head, torso and legs) and height of a person.

An interesting application of soft biometrics was studied and analyzed by Niinuma et.al [172]. Their problem was based on the fact that most existing computer and network systems authenticate a user only at the initial login session. As such, there exists a high chance that an impostor may get access to the system if the genuine user forgets to log out or takes a break from the system. To resolve such an unwanted situation, the authors suggested a framework for continuous user authentication that

primarily uses soft biometric characteristics (e.g., color of user's clothing and facial skin). In a different domain, some works regarding imputation techniques associated with soft biometrics have been researched. Missing soft biometric features can be predicted utilizing the structure within human appearance via these statistical techniques. Keeping these principles in mind, the authors proposed a system which utilized a similarity score based on the techniques alike to the k-nearest neighbor (kNN) classifiers [170]. Adjeroh et al. [173] also studied imputation and correlation in human appearance, but they used continuous data focusing on measurements of the human body. A benchmark testing suite comprising of a database based on keystrokes and soft biometric traits was built in [174]. This study was carried out to enhance and explore the authentication systems based on keystroke dynamics. Finally, a theoretical analysis on the reliability of soft biometric systems was performed in [175]. The authors centered their study in those situations where identification error occurs due to different subjects sharing similar soft biometrics. In addition, they also provided asymptotic bounds for interpreting their statistical model. This work establishes a useful mathematical framework for predicting the benefits associated with using a greater number of soft biometric traits, albeit subjected to some constraints.

An exhaustive study of the various aspects of soft biometrics is efficiently compiled by Dantcheva et.al in [176].

5.1.2 The Privacy Paradigm

Traditionally, privacy has been used as a metric for measuring the level of uncertainty of information corresponding to an individual within a database. Privacy preservation was first described by Dalenius [177] as the guarantee that an adversary learns nothing extra about any target if the adversary gains access to published data. Regarding databases, public and private attributes are generally modeled as random variables having a specific joint probability distribution. The privacy of an individual remains intact (i.e. there is no privacy loss) if the disclosure of the associated public attributes provides no additional information about the corresponding private attributes. In a probabilistic sense, it can be stated that the conditional probability of the private attribute should remain as high as possible after an adversary observes the public attributes. Standardizing metrics for measuring privacy levels has also been a challenging task. Conventionally, privacy has been accounted for in an information theoretic way. The uncertainty about a piece of undisclosed information is related to its information content. The information content of a source S is measured by its entropy H which is defined as-

$$H(S) = \sum_i p_i \log \frac{1}{p_i}$$

where p_i is the probability with which a character s_i is emitted from the source S.

Let a random variable representing the sensitive data of a user be denoted by X_{prv}. Similarly, let another random variable X_{pub} characterize data of the same user which is available publicly (i.e. it is accessible by the adversary). Furthermore, let's assume that X_{prv} and X_{pub} are correlated by a joint probability distribution function $p_{(X_{prv},X_{pub})}(y,x)$ where $\forall(y,x)|y \in X_{prv}$ and $x \in X_{pub}$. Under such a simple scenario, the privacy of the individual (e) can be quantified as -

$$e = \frac{1}{n} H(X_{prv}|X_{pub})$$

where n is the total number of records and $H(X_{prv}|X_{pub})$ represents the conditional entropy (equivocation) of X_{prv} given X_{pub}. This parameter e accurately captures the essence of privacy as it represents the leftover entropy of the private data on the disclosure of the associated public data. An equivalent metric privacy risk (R) [178] has been defined as the mutual information between the public and private random variables. Thus,

$$R = \frac{1}{n}[I(X_{prv}; X_{pub})] = \frac{1}{n}[H(X_{prv}) - H(X_{prv}|X_{pub})]$$

Both the parameters privacy risk (R) and privacy (e) are complimentary and essentially capture the same notion. However in this work, the privacy would be defined by the equivocation e.

5.2 Motivation and Overview

In a soft biometric based fusion framework, the soft features of the users get stored alongside the primary ones in a central database. In this study, the issues of privacy risks upon the leakage of such information have been addressed and subsequently a privacy preserving framework for the same has been proposed. Abstractly, it can be understood that the privacy issues of the users emerge due to the high degree of correlation that exists between their soft biometric information and other external databases. This ultimately increases the net amount of information associated with the users. [1].

The present work is divided into two parts. The first part tries to theoretically quantify the privacy levels (or the loss in privacy) on the leakage of a soft biometric database by proposing a consistent formal model. The principal factor that increases the privacy risks associated with such databases is the various linkages (attribute based) which exist between soft biometric and micro databases. These links assist an adversary in performing various cross matching or correlation based attacks in between the databases. The privacy risks further increase when the adversary possesses some auxiliary background information about either any particular targeted user or about the entire database as a whole. This work attempts to quantify a user's privacy guarantees by providing a theoretical framework under the various possible attack scenarios. As per knowledge of the author, there exists no work in

[1] Zero leaked information translates to the highest level of privacy whereas complete disclosure of information relates to zero privacy

the literature which attempts such an evaluation. In a sense, this work can be generalized as the assessment of privacy risks due to the presence of correlation among two databases. The study is motivated by the pioneering work of [179], in which the authors gave an information theoretical analysis of the trade-off between utility and privacy in databases.

The second part of the work deals with the construction of a privacy preserving soft biometric based multimodal framework. The privacy guaranteeing notion *differential privacy* [180] has been adapted in the realm of biometric recognition systems for such a construction. For implementing differential privacy, a normal biometric recognition system has been initially modeled as an interactive query response based framework. This modified structure is termed as a Query Based Biometric System (QBBS). Furthermore, the Laplacian noise mechanism has been adapted which adds external noise to the values of the soft biometric data. This random noise addition serves as a transformation on the original data (say Y) obtained during the enrollment phase. The same amount of noise is subsequently added to the soft biometric data captured during the verification process (say $Y*$). These two values are then matched using a modified Bayesian decision theory based probabilistic matching technique [33], thereby producing a final decisive score. Thus the enrollment and the verification data are essentially matched in a transformed domain without using a pre-defined cryptographic key. This feature automatically diminishes much of the problems and overheads related to usage of such keys (e.g. key generation and key distribution). Interestingly, the application of differential privacy principles not only provides the

guaranteed level of privacy but also safeguards against various other possible attacks in the soft biometric framework.

Part I: Formal Model

Construction

5.3 Important Observations

This section presents the various aspects of a generic soft biometric database, especially the similarities which they share with a standard micro database.

5.3.1 Soft Biometric vs. Micro Databases

A typical soft biometric database differs from a micro database in several aspects. Firstly, the soft databases are not distributed in the public domain for any research purposes. The reason for this stems from the fact that the sole purpose of a soft biometric database is to store the records of enrolled users such that they (i.e. the records) can be conveniently retrieved during the authentication phase. On the other hand, conventional micro databases are distributed to external communities either for some research purpose or due to some pre-defined agreements. Another primary difference between the two databases arises regarding the accuracy of the data which they store. Since disclosure of private information directly accounts for privacy breach for the users, micro databases are subjected to some data sanitization mechanisms prior to their distribution. These statistical mechanisms either perturb or generalized the data based on appropriate requirements. These principles are collectively known as Privacy Preserving Data Publishing (PPDP) in the literature [181]. On the contrary, soft biometric databases are stored in their original form without any modification, since any altercation in the database could profoundly affect the performance of the entire recognition system. Despite all these differences,

a soft biometric database can be considered identical to a micro database if the database gets leaked. Such exposure can become perilous for the users since an adversary could acquire their private data in the original form, rather than in any modified version.

5.3.2 Generic Soft Biometric Database Schema

A typical soft biometric database consists of a user enrollment ID I(for uniquely identifying a user) and its associated soft biometric attributes ($\mathcal{K}$). Adhering to the conventional terminologies of a database [178], a generic soft biometric database schema can be assumed to contain the following types of attributes-

- **Identifiers**- These attributes unambiguously identify the respondents. Typical examples of these in micro databases include *SSN number* and *passport number*. These attributes are either removed or encrypted prior to distribution due to the high privacy risks associated with them. In the case of soft biometric databases, identifiers correspond to the unique ID I of the enrolled users.

- **Key attributes**- Key attributes are those properties which can be linked or combined with external sources or databases to re-identify a respondent. Typical examples of such attributes include *age* and *gender*. In the case of soft biometric databases, the set of key attributes is denoted by $\mathcal{K}_{pub}$. Here

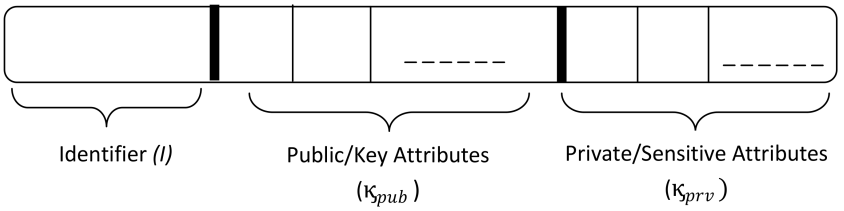


FIGURE 5.1: The soft biometric database schema.

pub stands for public since these attributes do not pose any privacy threat to the users unless when linked with external sources of data.

- **Sensitive attributes**- These attributes contain the most critical data of the users; maintaining their confidentiality is the primary objective of any security scheme. Obtaining unauthorized information about these properties is considered as a direct breach of privacy. Examples of these type of attributes include *medical diagnosis*, *political affiliation* and *salary*. Regarding soft biometric databases, the set of sensitive attributes are denoted by $\mathcal{K}_{prv}$. Here *prv* stands for private data.

The general soft biometric database schema is diagrammatically illustrated in Figure 5.1.

By the definitions, $\mathcal{K} = (\mathcal{K}_{pub} \cup \mathcal{K}_{prv})$. It should be noted that a soft biometric attribute can either be a key attribute (public) or a sensitive one (private), but not both (i.e. $\mathcal{K}_{pub} \cap \mathcal{K}_{prv} = \phi$). The choice of a soft attribute falling into one of these two categories depends upon the need of the user. For example, if a user considers his weight as a public attribute then *weight* falls in the set $\mathcal{K}_{pub}$. Otherwise if the same user considers his weight to be a confidential piece of information, it is included

in the set $\mathcal{K}_{prv}$. As demonstrated in later sections, both of these scenarios have been considered while designing the theoretical framework depicting privacy levels.

5.3.3 Attribute Based Correlation - Practical Instances

With the proliferation of biometric systems, a large number of users get enrolled in various biometric databases. However with the advent of the data-centric world, there exists a very high chance that the same users have also enrolled and provided their personal data in other services. These other services may typically include medical and hospital records, insurance records, voter list records and normal census data. Among all these micro databases, some databases get distributed to the outside world. As mentioned before, this distribution of data mainly occurs either due to the data publishing policies of some organizations or for general research purposes. Typical examples of this type of situation include the hospitals in California which publish their patient records on the web [182] or the *Adult Data Set* of UCI machine learning repository [183] which is heavily used in the research community. Since the actual number or percentage of such users are not available (their explicit identities are suppressed owing to privacy factors), focus has been made upon the attributes present in these databases. These attributes are commonly termed as demographical values. Most frequently occurring demographics across various databases include name (suppressed in some cases), date of birth (or age), gender, race, and address. However, this inter-relation between various databases results in some serious threats for the users. For instance, Sweeney demonstrated that she

Soft Traits	*Height, Gender, Race*	*Gender, Blood Group, Weight, Height*	*Weight*, Fat Measurements	*Set of 23 body measurements*	*Age, Gender*	*Weight*, Clothes Color, Eye Color + 6 more
Ref.	[157]	[158]	[159]	[167]	[174]	[155]

TABLE 5.2: Attributes used in soft biometrics databases.

Domain	Hospital Records	Healthcare	Census	Epidemiology	Anthropometry	Govt. Issued ID Card	Research
Demographic Attributes	*DOB, Gender*, Blood Group	*Age, DOB, Gender, Race Weight*	ZIP, *DOB, Gender, Race*	*Height, Weight, Age*, BMI	*Height, Weight, Age, Gender*	*Height, Weight, Age*, BMI	*Age, Gender, Race*, Workclass, Country
Ref.	[185]	[186]	[187]	[188]	[189]	[190]	[183]

TABLE 5.3: Attributes of external micro databases.

could re-identify the medical record of William Weld (governor of Massachusetts) using only his date of birth, gender, and ZIP code [184].

A majority of the soft biometric databases contain demographic values of the enrolled users as well. This ultimately results in increased privacy risks for the users since their biometric data can be easily linked with the other databases discussed previously. Some of the sets of overlapping demographic values are represented in a tabular form in Table 5.2 (soft biometrics) and Table 5.3 (micro databases). The most common attributes are marked in *italics*. This similarity of attributes provides a validation of the practical scenarios which have been considered in this work.

5.4 Model Construction

This section is devoted to the development of the privacy assessment model.

5.4.1 Assumptions

A few assumptions have been made in this work regarding the distribution and correlation of attributes in a generic soft biometric database. These suppositions assist in developing a formal and consistent mathematical model. A majority of these assumptions have been already justified and subsequently used in previous works [179].

Firstly, a soft biometric database is modeled as a collection of m observations (rows) generated by a memoryless source whose outputs are independently and identically distributed (i.i.d). Additionally, the rows of the database is a collection of correlated attributes (corresponding to a user) that is generated according to its probability of occurrence from a well-defined source. Theoretically, it can also be inferred the data in the database (which is empirical) to be the statistical distribution of the source; this approximation grows better as m grows.

One of the primary objectives of this study is to quantify privacy due to correlation based attacks. For this purpose it is assumed that linking based attacks are performed on the basis of only public attributes. Thus public attributes serve as the source of correlation attacks for an adversary, whereas the final objective of the

adversary is to obtain information about the private attributes (thereby decreasing the corresponding privacy levels). Lastly, it is assumed that micro databases are available to the adversary in a sanitized form. This condition is both practical and consistent with previous discussions.

Some assumptions are also made on the adversary. The adversary is considered to possess some auxiliary background information regarding either any particular targeted user or the entire group of users in the database. For example, the adversary may know whether or not a user had participated in a database. A worst case example assumes that the adversary has knowledge about the unique identifier of a user. This assumption enables us not only to take into consideration the various possibilities of privacy breach but also makes the model more generic.

5.4.2 Categories of Privacy Loss

As already discussed, the attributes in a soft biometric database can be categorized as either private ($\mathcal{K}_{prv}$) or public ($\mathcal{K}_{pub}$). Under such circumstances and previous assumptions, privacy can be breached in the following three instances -

- **Case 1:**[$\mathcal{K} = \mathcal{K}_{prv}$] - In this case all the attributes in the soft biometric database are private. Thereby a leakage in the database would result in the revelation of all the sensitive information corresponding to a user. However, the adversary would not be able to link this soft biometric database to other

external databases since there are no public attributes present in it. This case is very unlikely in practical scenarios.

- **Case 2:**[$\mathcal{K} = \mathcal{K}_{pub}$] - The second scenario corresponds to the case when all the attributes present in the soft biometric database are public. The only option of the adversary, in this case, would be to obtain information about a user by linking $\mathcal{K}_{pub}$ with external databases. The severity of these linking attacks increases with the amount of background information available to the adversary.

- **Case 3:**[$\mathcal{K} = \mathcal{K}_{prv} \cup \mathcal{K}_{pub}$] - The third and final case accounts for the situation when the soft biometric database consists of both private and public attributes. This is the riskiest situation since, in addition to directly obtaining the sensitive information through the private attributes, the adversary can additionally link the public attributes with external data sources thereby obtaining the maximum amount of information about the user. This case also represents the majority of practical scenarios.

5.4.3 Micro Database Model

All the notations associated with a generic micro database DB are defined in this section. Let K denote the number of attributes in the database and $\mathcal{K}$ be the set representing these K attributes. Also, let X_k be a random variable denoting the k^{th} attribute, $k = \{1, 2, ...K\}$. The set of these K random variables are denoted by

$X_{\mathcal{K}}$; thus $X_{\mathcal{K}} = (X_1, X_2, ...X_K)$. It is assumed that $X_{\mathcal{K}}$ takes values from a finite set $\mathcal{X}_{\mathcal{K}}$. Let DB consist of n independent observations (i.e. rows) which follow a joint probability distribution -

$$p_{X_{\mathcal{K}}}(x_{\mathcal{K}}) = p_{X_1 X_2 ... X_K}(x_1, x_2, ...x_K)$$

The above dependency captures the correlation between the K attributes. However in accordance with previous works, independence among the n rows are assumed.

Let $\mathcal{K}_{prv}$ and $\mathcal{K}_{pub}$ represent private and public attributes in the database respectively. To reiterate, $(\mathcal{K}_{pub} \cup \mathcal{K}_{prv}) = \mathcal{K}$ and $(\mathcal{K}_{pub} \cap \mathcal{K}_{prv}) = \phi$. Furthermore, the set of their corresponding random variables are denoted by $X_{\mathcal{K}_{prv}}$ and $X_{\mathcal{K}_{pub}}$ respectively. Thus,

$$X_{\mathcal{K}_{prv}} = \{X_k\}_{k \in \mathcal{K}_{prv}} \quad \text{and} \quad X_{\mathcal{K}_{pub}} = \{X_k\}_{k \in \mathcal{K}_{pub}}$$

Accordingly, $X_{\mathcal{K}} = (X_{\mathcal{K}_{prv}} \cup X_{\mathcal{K}_{pub}})$.

Finally, a sanitization mechanism of DB is denoted through an encoding function F_E which maps DB to a set of indices $J = \{1, 2, ...M\}$ and a set of associated output sanitized databases SDB. Here M denotes the number of sanitized databases. Thus,

$$F_E : DB \rightarrow J, \{SDB_k\}_{k=1}^{M}$$

5.4.4 Soft Biometric Database Model

The soft biometric database and its associated parameters are marked with an asterisks sign ($*$). Otherwise, the notations are relatively similar to that of the micro database model. Let the soft biometric database itself be denoted by DB^*, the number of attributes by K^S and the set of such attributes by $\mathcal{K}^S$. As observed previously, many attributes overlap amid micro and soft biometric databases. Let the number of these common attributes be denoted by K^* and the corresponding set of common attributes by $\mathcal{K}^*$. Thus, $\mathcal{K}^* = (\mathcal{K}^S \cap \mathcal{K})$.

The reason for extracting only common attributes is due to the assumption that an adversary can perform cross correlation based attacks only on the basis of them. If an attribute is unique to any database, then it serves no purpose for the linking based attacks. Let X_k be a random variable denoting the k^{th} common attribute of $DB^*, k \in \{1, 2, ...K^*\}$; also let the set of such random variables be denoted by $X_{\mathcal{K}^*}$ such that $X_{\mathcal{K}^*} = (X_1, X_2, ...X_{K^*})$. Let DB^* consist of m independent rows ($m \neq n$) which follow a joint probability distribution -

$$p_{X_{\mathcal{K}^*}}(x_{\mathcal{K}^*}) = p_{X_1 X_2, ...X_{K^*}}(x_1, x_2, , ...x_{K^*})$$

This distribution function captures the attribute wise correlations for the soft biometric database. These attributes are next divided according to the three cases discussed previously. Let the private and public attributes be denoted by $\mathcal{K}^*_{prv}$ and

$\mathcal{K}^*_{pub}$ respectively. Moreover let their associated random variables be represented by $X_{\mathcal{K}^*_{prv}}$ and $X_{\mathcal{K}^*_{pub}}$. Thus,

$$X_{\mathcal{K}^*_{prv}} = \{X_k\}_{k \in \mathcal{K}^*_{prv}} \quad \text{and} \quad X_{\mathcal{K}^*_{pub}} = \{X_k\}_{k \in \mathcal{K}^*_{pub}}$$

Hence the three cases are formulated as -

1. **Private :** All the attributes in this case are private. Hence,

   $$\mathcal{K}^*_{prv} = \mathcal{K}^* \quad \text{and} \quad \mathcal{K}^*_{pub} = \phi$$

   Accordingly, $X_{\mathcal{K}^*_{prv}} = \{X_k\}_{k \in \mathcal{K}^*}$ and $X_{\mathcal{K}^*_{pub}} = \phi$.

2. **Public :** All the attributes in this case are public. Hence,

   $$\mathcal{K}^*_{pub} = \mathcal{K}^* \quad \text{and} \quad \mathcal{K}^*_{prv} = \phi$$

   Accordingly, $X_{\mathcal{K}^*_{pub}} = \{X_k\}_{k \in \mathcal{K}^*}$ and $X_{\mathcal{K}^*_{prv}} = \phi$.

3. **Mixed:** In this case there is a mixture of private and public attributes. Hence,

   $$\{\mathcal{K}^*_{prv} \cup \mathcal{K}^*_{pub}\} = \mathcal{K}^*$$

   Accordingly, $X_{\mathcal{K}^*_{prv}} = \{X_k\}_{k \in \mathcal{K}^*_{prv}}$ and $X_{\mathcal{K}^*_{pub}} = \{X_k\}_{k \in \mathcal{K}^*_{pub}}$.

5.4.5 Auxiliary Background Information

In addition to the public information available, an adversary can also utilize any related background information. Incorporating this aspect gives more power to the adversary for mining sensitive information about individuals. The most famous example of this notion can be described by the commonly referred Terry Gross's height example [180]. It basically states that supposing height was considered sensitive information, an adversary possessing the background knowledge that "Terry Gross is two inches shorter than the average Lithuanian women", can accurately calculate Terry Gross's height from a statistical database containing average heights of woman of different nationalities. In the case of soft biometric aided systems, this piece of sensitive information (i.e. height of Terry Gross) can be conveniently computed by running some simple arithmetic operations on a biometric database containing the heights of several Lithuanian women.

For this model, the side information is modeled as a K length sequence and denoted by Z and Z^* corresponding to the micro and soft databases respectively. Thus,

$$Z, Z^* \in (Z_1, Z_2, ...Z_K)$$

where $Z_i, (1 \leq i \leq K)$ takes values from a finite set $\mathcal{Z}$.

The side information must be correlated with both the main and soft biometric databases to be meaningful. These correlations are denoted by the joint probability

distribution functions $p_{X_{\mathcal{K}}Z}(x_{\mathcal{K}}, z)$ and $p_{X^*_{\mathcal{K}}Z^*}(x^*_{\mathcal{K}}, z^*)$ for micro and soft databases respectively.

This completes the construction of the database models and their associated dependencies. The process of quantifying privacy achievable through this model is defined next.

5.5 Quantified Privacy Levels

Privacy of an individual is quantified as the remaining entropy of their sensitive information, given that an adversary has access to some correlated public information. In this case, the sensitive information corresponds to the private attributes, while the public data relates to both the public attributes and auxiliary side information for that individual. An important observation can be made in this context. Since the effects on privacy upon the leakage of a soft biometric database is captured, one can consider it (i.e. the soft biometric database) completely as public information. Intuitively, this reduces the privacy of the individual since the adversary now possesses much more information which is correlated with the sensitive attributes.

The main principle for formulating the absolute privacy of a user involves two steps. In the first step, the privacy levels are considered when the adversary deals only with the main micro database. In such a case the privacy is given by the equivocation (e) as -

$$e = \frac{1}{n} H(X_{\mathcal{K}_{prv}} | J, Z) \geq E$$

where e is lower bounded by E. The maximum value of e occurs when (J, Z_n) reveals no information about $X_{\mathcal{K}_{prv}}$, i.e. when $X_{\mathcal{K}_{prv}}$ is independent of both J and Z_n. Privacy in that case equates to the entropy of $X_{\mathcal{K}_{prv}}$, i.e. $H(X_{\mathcal{K}_{prv}})$.

Since the sanitized micro database is made public, the quantity e serves as the minimum amount of information which the adversary possesses prior to the leakage of the soft biometric database DB^*. Moreover since e is a function of $(X_{\mathcal{K}_{prv}}, J, Z_n)$ it can be represented as a random variable X_e where,

$$X_e : (X_{\mathcal{K}_{prv}}, J, Z_n) \rightarrow [E, H(X_{\mathcal{K}_{prv}})]$$

Subsequently, the stage 1 privacy is quantified for all of these three distinct cases as-

1. **Private Attributes:**

   Here the privacy of an individual is denoted by e_1. Thus,

   $$e_1 = \frac{1}{m} H(X_e | X_{\mathcal{K}^*_{prv}}, Z^*) \geq E_1$$

   where E_1 represents a general lower bound on e_1.

2. **Public Attributes:**

   In this case, the privacy of an individual is denoted by e_2. Thus,

$$e_2 = \frac{1}{m} H(X_e | X_{\mathcal{K}^*_{pub}}, Z^*) \geq E_2$$

   where E_2 represents a general lower bound on e_2.

3. **Mixed Attributes:**

   For this case, the privacy of an individual is denoted by e_3. Thus,

$$e_3 = \frac{1}{m} H(X_e | X_{\mathcal{K}^*}, Z^*) \geq E_3$$

   where E_3 represents a general lower bound on e_3.

For all three cases, the maximum value of privacy occurs when the soft biometric database DB^* is uncorrelated (i.e. independent) to both the main database DB and the background information Z^*. This statement alternatively signifies that a particular user U (whose record is present in DB) had not participated during the enrollment procedure of the soft biometric database in the first place and the adversary possesses no related background information about U. Under such circumstances, the adversary gains no additional information upon observing the leaked database DB^*, thus preserving the original privacy level e.

Part II: Providing Privacy Guarantees

5.6 Background Requisites

This section discusses some basic concepts which are essential for understanding the intuition behind the approach for constructing the proposed privacy preserving framework.

5.6.1 Underlying Soft Biometrics Fusion Framework

This work employs the modified Bayesian decision theory based probabilistic fusion framework introduced by [33] as the underlying mechanism. The enrollment procedure of this Bayesian framework is similar to that of a normal biometric system. In this first phase, the values of a primary trait and more than one soft traits are extracted from users. Formally speaking, let the users be represented as

$$U = (U_1, U_2, ...U_c)$$

where c represents the total number of users. Let X represent the primary biometric value and $Y = (Y_1, Y_2, ...Y_n)$ be the set of both discrete and continuous soft biometric values associated with each user. Here n denotes the total number of soft biometric values corresponding to each user. These values get stored in a database after some pre-processing steps. In all the subsequent constructions, i would index the users (U) and j would index their soft biometric attributes (Y). Thus, $i = \{1, 2, ...c\}$ and

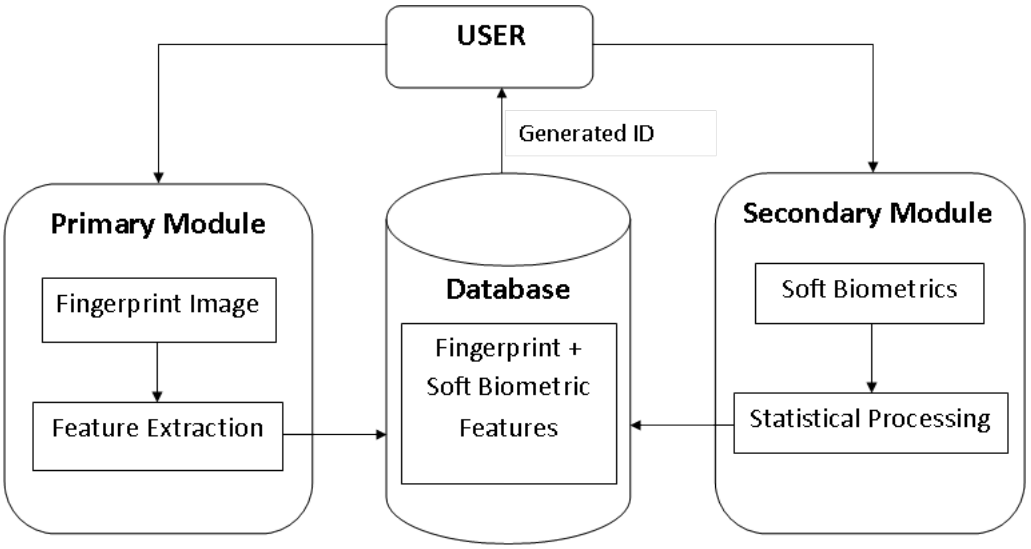


FIGURE 5.2: Enrollment process in the Bayesian decision theoretic framework.

$j = \{1, 2, ...n\}$. The enrollment procedure is diagrammatically illustrated in Figure 5.2.

The verification/identification procedure is somewhat different from normal biometric systems though. During this phase, a combined score based on the matching values of both the primary and the secondary system is generated. The final decision is taken according to this calculated score. The combination of the two modalities takes place in the form of a multi-leveled Bayesian approach. The verification process is depicted in Figure 5.3. Formally speaking, let the final matching score for user i be represented as $P(U_i|X, Y)$. The probability is calculated as -

$$P(U_i|X, Y) = \frac{P(Y|U_i)P(U_i|X)}{\sum_{i=1}^{c} P(Y|U_i)P(U_i|X)} \tag{5.1}$$

If the Y values are assumed to be independent, Equation 5.1 can be rewritten as:

$$P(U_i|X,Y) = \frac{P(Y_1|U_i) \times P(Y_2|U_i)...P(Y_n|U_i) \times P(U_i|X)}{\sum_{i=1}^{c} P(Y_1|U_i) \times P(Y_2|U_i)...P(Y_n|U_i) \times P(U_i|X)} \tag{5.2}$$

On assigning variable weights to each modality and taking logarithms of both sides, the final form of Equation 5.2 becomes -

$$g_i(X,Y) = a_0 \times \log P(U_i|X) + a_1 \times \log P(Y_1|U_i) + ...a_n \times \log P(Y_n|U_i) \tag{5.3}$$

Here, $g_i(X,Y)$ represents the final decisive score for the user, $P(U_i|X)$ represents the prior probability that the user is genuine given the primary biometric based score and $P(Y_j|U_i)$ represent the prior probabilities of soft biometric traits of user i during the verification phase given the true values of the same traits obtained during the enrollment phase. $P(Y_j|U_i)$ can also be represented as $P(Y*|Y)$ where $Y* = (Y_1*, Y_2*, ...Y_n*)$ corresponds to the observed soft biometric values during verification.

In this work, the Bayesian framework has been slightly altered to suit the objective. In the original construction, calculation of the soft biometric based conditional probabilities (i.e. $P(Y^*|Y)$) demands that either the traits be discrete, or possess some underlying distributions. But real life continuous soft biometric data (e.g. height,

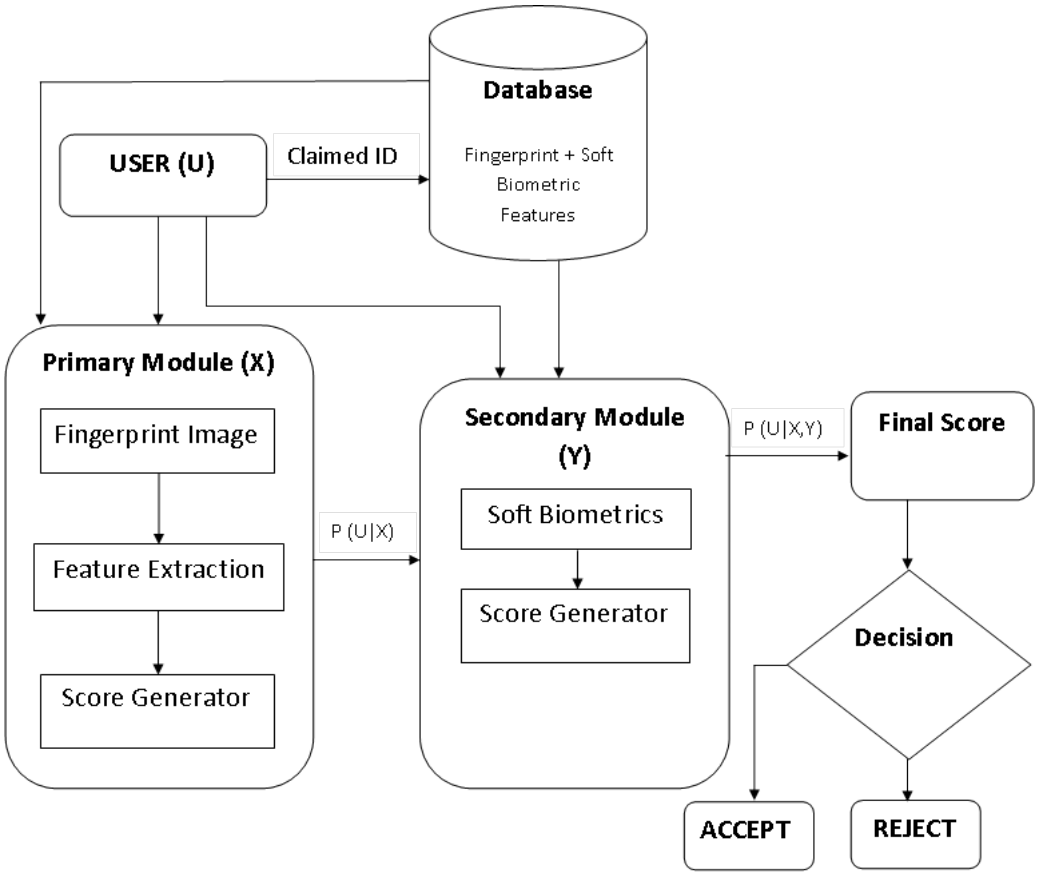


FIGURE 5.3: Verification process in the Bayesian decision theoretic framework.

weight etc.) seldom follow any specific distribution. Hence a Gaussian function has been used to approximate the value of $P(Y^*|Y)$. A Gaussian function is of the form:

$$f(x) = a \times \exp\left(-\frac{(x-b)^2}{2c^2}\right)$$

where a = height of the curve's peak, b = position of the center of the peak and c = standard deviation.

This function has been used with fixed parameters $a = 1$, $b(mean) = Y$ and $c(sd) = 1$. The motivation in using this function lies in its widespread use in statistics and

probability theory for representing real-valued random variables whose distributions are not known [191]. The value of the height (a) is fixed to 1 due to the requirement of a probabilistic score between 0 and 1. This function fits well into the scheme since real-life continuous values have been for the experimental purposes. Thus by using this approach, all the values of $P(Y^*|Y)$ were calculated and subsequently fitted into the final form of Equation 5.3.

5.6.2 Differential Privacy Foundations

Differential privacy is a perturbation type (in contrast to the generalization schemes like k-anonymity, l- diversity etc.) data modification scheme that changes the values of the original database. This mechanism usually works in a query response framework wherein, instead of an individual having direct accessibility to the main database (Db), a series of queries are fired on it to get appropriate responses. These responses are then subjected to some statistical analysis so as to retrieve some information about the individuals participating in the database. If an adversary is able to find out some unwarranted crucial information about any individual after analyzing the responses, the privacy of that particular individual gets breached. Dwork [180] first introduced the notion of differential privacy to counter such a scenario. The intuition behind this novel idea was that the participants would feel safe if they had the assurance that the response to any query fired in the database would remain the same whether or not they submitted their data.

Definition 1 (ϵ differential privacy [180]). A randomized function K gives ϵ-differential privacy if for all datasets D_1 and D_2 differing on at most one element, and all $S \subseteq Range(K)$,

$$Pr[K(D_1) \in S] \leq exp(\epsilon) \times Pr[K(D_2) \in S]$$

where, Pr is a probability distribution over the randomized function.

Differential privacy is achieved by introducing a random amount of noise in the output responses obtained from the database. The magnitude of this noise depends upon the largest change that a single participant can effect on the responses. This quantity is termed as $L1$ sensitivity or Global sensitivity (Δf).

Definition 2 (Global sensitivity (Δf) [180]). For any function $f : D \rightarrow R^d$, the $L1$-sensitivity of f is

$$\Delta f = max_{D_1,D_2} ||f(D_1) - f(D_2)||$$

for all D_1, D_2 differing in at most one element.

Differential privacy is usually enforced by adding noise which is properly calibrated to the output of the responses. The noise is sampled from a Laplace distribution with probability density function:

$$Laplace(x, \lambda) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

where λ is determined by both Δf and desired privacy controlling parameter ϵ.

Theorem 1 (Laplace Noise mechanism [192]). For any function $f : D \rightarrow R^d$, the mechanism

$$Laplace(D, f, \epsilon) = f(D) + [L_1(\lambda), L_2(\lambda), ..., L_d(\lambda)]$$

gives ϵ-differential privacy if $\lambda = \Delta f/\epsilon$ and $L_i(\lambda)$ are i.i.d. Laplace random variables.

A comprehensive survey about differential privacy techniques and its diverse applications is efficiently compiled in [193].

A parameter which is adversely affected by the implementation of differential privacy is the utility factor. As intuition suggests, a greater privacy level leads to a larger amount of noise in the system and a consequent decrease in the eventual utility of the database. Due to this fact and some other opinions, studies such as [194] and [195] have questioned the viability of this procedure for numeric queries. However in this work, the differential privacy principle serves more as a data transformation scheme rather than a data masquerading technique. The probabilistic comparison between the data obtained during enrollment and verification phases is done after perturbing the data (in both the phases) with the same amount of noise. Because of this fact, the privacy preserving framework does not affect the utility of the recognition

system (which is represented by the ROC curve of the system in this case). It was also reasoned in [194] that the Laplace noise based mechanism does not satisfy differential privacy for numeric data sets. However in doing so, they considered the situation that an adversary can work upon a new data set that differs from the original database by exactly one record and the missing observation is the most influential observation for a particular query. But for this case the risk is negligible since such a juxtaposition is very rare in biometric recognition systems.

5.7 Framework Development

5.7.1 Query Based Biometric System (QBBS)

A differential privacy mechanism encompasses a query response framework. Hence a basic biometric authentication system has been aptly modified to suite such a framework. This sub-system is termed as a Query Based Biometric System (QBBS). In this modified design, a new module termed as a Query Module (QM) is introduced in the basic framework such that it is integrated with the central database into one unit. As such, the database cannot be anyhow separately accessed (during data read operation) without bypassing the QM. The QM basically serves as a functioning block for communicating and retrieving data from the database. The matcher module in a QBBS is also slightly modified to facilitate in the computation of the

final trait values from the individual response fragments. This modified framework is shown diagrammatically in Figure 5.4.

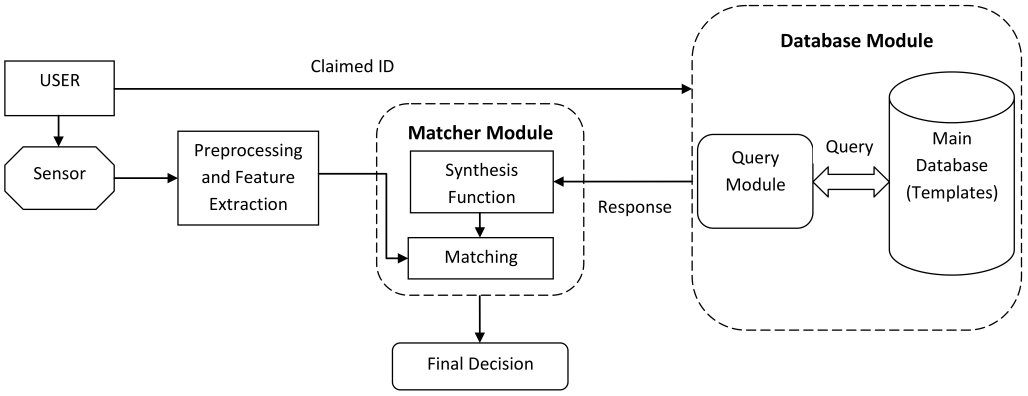


FIGURE 5.4: Basic structure of a QBBS

An enrollment phase in QBBS works identically to a normal biometric system. A successfully enrolled user submits his biometric entries into the system, which then get stored in the database. The QM is simply bypassed in this phase. However during the authentication phase, the biometric ID of a querying user is used by the QM to probe into the database for retrieving data related to the corresponding ID. For doing so, the QM structures some appropriate queries and fires them into the database. The responses from these queries are collected and some statistical operations (synthesizing function) are performed on them for revealing the final form of the required data. Before initiating any biometric enrollment session, it is necessary to specify a fixed query structure (i.e. both the number and type of queries) suited to the framework. This requirement is essential so as to build an accurate and correct synthesizing function. An important point to note here is that the queries in a QBBS can be issued by either a legitimate user or an adversary.

Formally speaking, let the identifier of a user be denoted by ID and his biometric entries be denoted by the set B. Here B may be primary biometric values (X), soft biometric values (Y) or both. Moreover, let n be the total number of biometric traits of the user. In this and subsequent sections, i would index B (i.e. $i = \{1, 2, ...n\}$). Thus for each user -

$$B = \{B_1, B_2, ...B_n\}$$

In the verification phase, the QM fires m number of queries corresponding to each n traits. Let Q denote the set of all fired queries for a particular user and q represent the individual queries. In all subsequent sections, j would index these individual queries (i.e. $j = \{1, 2, ...m\}$). Thus -

$$Q = \{Q_1, Q_2, ...Q_n\} \quad \text{and} \quad Q_i = \{q_1, q_2, ...q_m\}$$

Each of these queries is a function of the verifying identity of the user (ID) and the corresponding biometric trait. Thus -

$$Q_i = f(ID, B_i)$$

It is assumed n this model that equal number of queries are fired for each biometric trait. Let R represent the set of all responses obtained from the corresponding

queries and r specify the individual responses. Thus,

$$R = \{R_1, R_2, ...R_n\} \quad \text{and} \quad R_i = \{r_1, r_2, ...r_m\}$$

In this model, the matcher module executes the additional task of performing the synthesizing function on the responses, thereby calculating the true values of the biometric traits. Let the set of these true values be denoted by V. Thus,

$$V = \{V_1, V_2, ...V_n\} \quad \text{where} \quad V_i = f(R_i)$$

Simple examples of such functional operations include addition, subtraction, multiplication and averaging. These computed values are then compared by the matcher module with the biometric values of the user obtained during the verification phase, thus finally generating the matching score.

5.7.2 Permissible Queries

A biometric database is not similar to a standard statistical database partly because it is not released to the outside world for research and analysis purposes. As such, modeling a biometric system as a query response framework limits the type of issuable queries. Since these queries are targeted towards the calculation of the biometric trait values, simple functional queries are permitted. For example, a SQL

query to calculate the sum of all the 'age' attribute values in the database can be structured as -

SELECT SUM(Db.age)

FROM database Db;

The following set of numeric functional queries provides an accurate way for calculating the age of a user having $id = 87$ (id represents identification number of an enrolled user).

Q1 = **SELECT SUM**(Db.age)

FROM database Db;

Q2 = **SELECT SUM**(Db.age)

FROM database Db;

WHERE Db.id **NOT IN** ('87');

Hence age of the user, $V = (Q1 - Q2)$.

For primary biometric modalities (i.e. face, fingerprint etc.), normal selection queries are issued. These queries directly retrieve the data corresponding to the provided ID number of an enrolled user. As evident from these queries, there is no need for combining responses in the matcher module (i.e. synthesizing function is not required) in the case of primary biometric traits. Examples for these queries include (*data* represents the primary trait values like minutiae points or iriscodes) -

Q3 = **SELECT** Db.data

FROM database Db

WHERE Db.id = 87;

These selection queries can also be used to directly retrieve the soft biometric values, but instead the aforementioned multiple functional queries were utilized to achieve the dual purpose of facilitating in the computation of 'global sensitivity' corresponding to the queries and increasing the overall security of the framework. Other query types like count or predicate queries are not allowed since their objectives (counting the total number or the fraction of database entries matching a predefined predicate) does not fulfill any requirements. For example a SQL count query such as the following ($Q4$) will have no functionality in the framework.

Q4 = **SELECT COUNT**(Db.age)

FROM database Db

WHERE Db.id = 87;

5.7.3 Global Sensitivity (Δf) Computation

Global sensitivity estimation for queries is essential for implementing differential privacy since the amount of randomly generated noise directly depends upon this parameter. However this computational procedure requires the knowledge about the upper and lower bounds of the data associated with the issued queries [194]. Although cumbersome for arbitrary statistical databases, this piece of information can be conveniently mined from a biometric database. For this framework, this task

is performed by the Query Module (QM). Upon the issue of a query concerning a particular soft biometric trait, the QM first scans the database and then selects the highest and the lowest values of that particular trait in the database. Let these values be denoted as max_i and min_i. This information is then utilized for the calculation of Δf. Thus,

$$\Delta f_i = \sum_{j=1}^{m} \Delta f_j \quad \text{where} \quad \Delta f_j = f(q_j, max_i, min_i)$$

For example, Δf for query numbers (1) and (2) can be calculated by further issuing the following queries -

Q5 = **SELECT SUM**(Db.age)

FROM database Db;

Q6 = **SELECT SUM**(Db.age)

FROM database Db;

WHERE Db.age **NOT IN** ('max−age')

Q7 = **SELECT SUM**(Db.age)

FROM database Db;

WHERE Db.id **NOT IN** ('87');

Q8 = **SELECT SUM**(Db.age)

FROM database Db;

WHERE Db.age **NOT IN** ('max−age')

AND Db.id **NOT IN** ('87');

Thus, $\Delta f(1) = Q5 - Q6$, $\Delta f(2) = Q7 - Q8$ and finally, $\Delta f(age) = \Delta f(1) + \Delta f(2)$.

This procedure for the calculation of global sensitivity arises from its basic definition. Δf for a query is defined as the worst possible outcome of the query if it is fired on a dataset which differs from the original dataset by only one data value. Corresponding to the queries, this worst possible situation arises when that particular data value corresponds to the maximum value of the soft biometric trait in the database. This is also the principal reason for pre-establishing the query structure for the QBBS.

5.7.4 QBBS Based Privacy Preserving Framework

Using all the elements discussed till now, a privacy preserving soft biometric framework is constructed here. This design is essentially a QBBS where the QM additionally perturbs all the responses by some random Laplacian noise. The magnitude of this noise depends upon the global sensitivity of corresponding soft biometric traits. This specific noise addition mechanism ensures that the overall framework enjoys differential privacy, thus preserving privacy for the enrolled users. The whole design is illustrated in Figure 5.5.

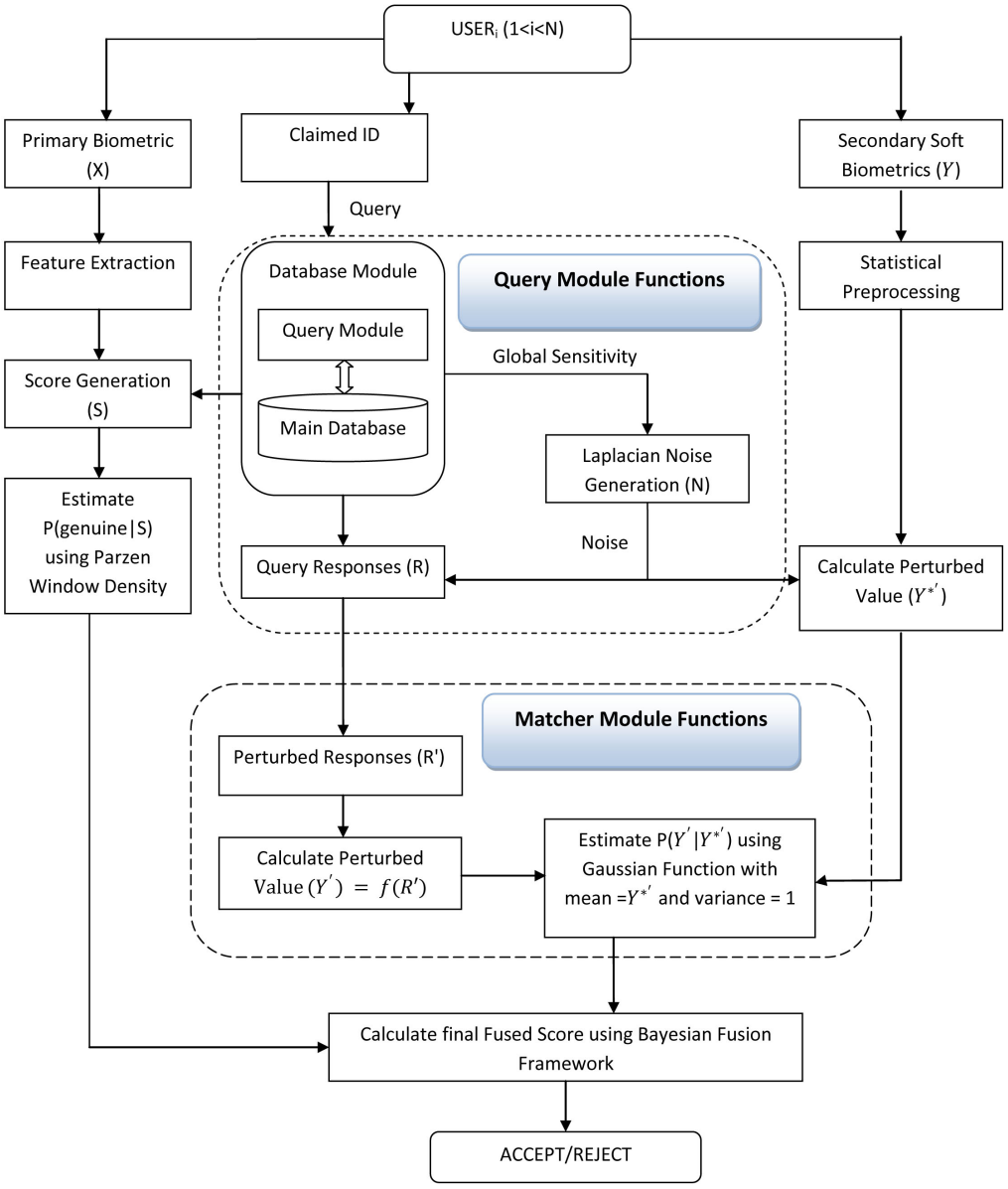


FIGURE 5.5: Proposed privacy preserving QBBS

Formally addressing, let the primary biometric trait of a user be denoted by X, his soft biometric trait values obtained during enrollment by Y, the same extracted during verification by Y^* and the associated identification number of the user by ID. For the entire construction and subsequent analysis, the total number of soft biometric traits would be denoted by n and the number of queries fired for each trait

by m. Also, i will be used to index n whereas j for indexing m. Thus,

$$Y = \{Y_1, Y_2, ...Y_n\}$$

Let the queries fired by the QM be represented by the set Q. The individual queries belonging to a particular trait are denoted by q. Thus -

$$Q = \{Q_1, Q_2, ...Q_n\} \quad \text{and} \quad Q_i = \{q_1, q_2, ...q_m\}$$

As argued previously, $q_j = f(ID, Y_i)$. Let all the received responses be denoted by the set R. The individual responses for a particular trait are denoted by r. Thus -

$$R = \{R_1, R_2, ...R_n\} \quad \text{where} \quad R_i = \{r_1, r_2, ...r_m\}$$

After receiving these m responses corresponding to each soft trait, the QM computes the Global sensitivity (Δf) values for each trait as -

$$\Delta f_i = \{\Delta f_1, \Delta f_2, ...\Delta f_m\} \quad \text{where} \quad \Delta f_j = f(q_j, max_i, min_i)$$

Based on these values and a selected ϵ value, random noise is generated from a Laplace distribution having a probability density function -

$$f(x|\mu,\lambda) = \frac{1}{2\lambda} exp\bigg(- \frac{|x-\mu|}{\lambda} \bigg) \quad \text{where} \quad \mu = 0 \quad \text{and} \quad \lambda = \frac{\Delta f}{\epsilon}$$

.

Let the generated noise corresponding to the queries be denoted by the set N. The individually generated noises (for a particular soft trait) is denoted by ns.Thus -

$$N = \{N_1, N_2, ...N_n\} \quad \text{and} \quad N_i = \{ns_1, ns_2, ...ns_m\}$$

Also let,

$$\phi_i = f(N_i)$$

Here ϕ_i basically represents the net amount of noise introduced in the system w.r.t. the soft biometric trait i. The values of these noise terms directly depend upon the query structure which gets fixed at the beginning of the biometric session. They are successively used to perturb the original responses associated with trait i. Let these masked responses be denoted by r' and the set of all such responses by R'. Thus,

$$R' = \{R'_1, R'_2, ...R'_n\} \quad \text{and} \quad R'_i = \{r'_1, r'_2, ...r'_m\} \quad \text{where} \quad r'_j = (r_j + ns_j)$$

According to Theorem 1, the proposed mechanism provides ϵ - differential privacy to the mapping function $f : Q_i \rightarrow R'_i$.

Let the final estimated values of the original soft biometric traits (i.e. Y) be denoted as Y'. This value is computed in the matcher module by computing a synthesis function on the perturbed responses. Thus,

$$Y_i' = f(R_i')$$

Alternatively, the soft biometric values obtained during the verification phase ($Y*$) are also transformed by trait wise incrementing them with ϕ. Let these altered values be denoted as $Y*'$. Hence,

$$Y*' = Y_i * + \phi_i$$

Finally, the two transformed soft biometric values (i.e. Y' and $Y*'$) are probabilistically matched using the Bayesian decision theoretic fusion framework. The final score for a user is calculated by -

$$g(X,Y) = a_0 \log(P(U|X)) + a_1 \log(P(Y_1 *' |Y_1')) + a_2 \log(P(Y_2 *' |Y_2')) + \\ ....a_n \log(P(Y_n *' |Y_n')) \quad (5.4)$$

In summary, the privacy preserving framework applies the principles of differential privacy to modify the soft biometric data of a user. Since an equal amount of net

Laplacian noise is added to both the soft biometric data (i.e. Y and $Y*$), their matching in the transformed forms (i.e. Y' and $Y*'$) is valid.

5.7.5 Composability

Composability in the domain of differential privacy refers to the assurance that the privacy requirements remain fulfilled even when several outputs (responses) are taken together and subjected to joint analysis. This framework demonstrates that the individual query responses satisfy ϵ-differential privacy, but this section discusses the composability of the overall set of responses. This analysis is important due to the studies done in [196], wherein composition attacks on independent k-anonymizations of intersecting data sets were exhibited.

The framework essentially comprises of a sequence of computations which provide differential privacy in isolation. This type of structures is termed as a sequential composition [197]. Let M be a mechanism for providing ϵ-differential privacy to a random variable X via the Laplace noise addition method.

Theorem 2 (Sequntial Composition [197]). Let each M_i provide ϵ_i-differential privacy. Then the sequence of $M_i(X)$ provides $(\sum_i \epsilon_i)$-differential privacy.

In accordance to Theorem 2, the proposed framework satisfies $(\sum_m \epsilon_m)$-differential privacy for every trait. This happens since each soft biometric trait i consists of m query responses, wherein each of them individually satisfies ϵ- differential privacy.

5.7.6 Additional Advantages

In addition to providing privacy to the users, the proposed framework can mitigate most of the other security threats associated with a normal biometric system. The principal advantages of this framework arise from the fact that the accessibility to the central database is very much restricted in this design. Since the framework is a modified QBBS, the only way to access the database is by issuing queries via the QM. From an adversary's point of view, he/she can only work with the masked responses obtained from the QM to launch any attack. Moreover, issuing the same set of queries multiple times will be of no avail to the adversary since every time the responses will be perturbed by a varying amount of randomly generated noise. The ability of this model to handle other prominent types of attacks is discussed as follows.

- **Linking Attacks**- An adversary obtains only the modified response values from the QM instead of the real values. Under such circumstances, linking with other external databases is not possible since the adversary does not possess accurate trait values.

- **Impersonation Attacks**- The adversary needs to accurately know the soft biometric values of a genuine user to launch this attack. Due to the reasons mentioned above, this attack is also not possible.

- **Attacks at Interconnections**- All the biometric feature values propagate throughout the entire framework in a perturbed (masked) form. Henceforth,

an adversary eavesdropping on the intermediate links between the various modules of the framework would gain very little information about the actual values of the attributes.

5.8 Theoretical Security Analysis

In this section, a security analysis is performed based on the probabilistic advantage that an adversary possesses via the net information that he extracts from the system. From observing the proposed framework, it is apparent that the only way an adversary can accurately reproduce the original trait values is by trait wise subtracting the net amount of noise from the final perturbed values, i.e.

$$Y_i = Y_i' - \phi_i$$

In the worst case, the adversary can intercept all the perturbed responses from the output of the QM and also the net amount of noise (ϕ) for each trait from intermediate transmission links. It can be also assumed that the adversary has knowledge of the operations that are required to be performed on these responses since the query structure remains fixed throughout the system. However the facts that remain hidden from the adversary are the actual correspondence between these responses (i.e. which responses are relevant to which biometric trait) and the relationship between the correct responses and the net noise associated with a particular trait (i.e. which

quanta of noise is required to be subtracted from which combination of responses). Assuming that there are n soft biometric traits of a user and for each trait a total of m responses exists, the total number of ways the adversary can group these responses in sets of cardinality m are -

$$\binom{nm}{m} = \frac{(nm)!}{m!(nm-m)!}$$

After grouping the responses, the adversary needs to perform some operations on them. Assuming these operations to be non-commutative (in the worst case), the total number of possible arrangements becomes -

$$\frac{(nm)!}{m!(nm-m)!} \times m! = \frac{(nm)!}{(nm-m)!}$$

Finally, the adversary needs to map the results from these combinations to the set of ϕ values having cardinality n. Since the actual correspondence between them is not known to the adversary, the total number of combinations that is required in the worst case is given by -

$$\frac{(nm)!}{(nm-m)!} \times n$$

Thus the probability of success for the adversary can be represented as -

$$P[Success] = \frac{1}{\frac{(nm)!}{(nm-m)!} \times n} = \frac{(nm-m)!}{n \times (nm)!}$$

For the simulation purposes, the following values have been considered-

- n=4(height, weight, gender, age)
- m=2 (queries identical to the ones illustrated in Section 5.7.2)
- operations = subtraction (non commutative over real numbers)

Thus,

$$P[Success] \approx 0.0045$$

This success probability of an adversary can be further decreased by structuring more number of functional queries for each soft biometric trait. For constructing 5 queries for each soft trait (i.e. m=5), the success probability becomes -

$$P[Success] \approx 0.0000001343739$$

The security level of the proposed framework can be more rigorously treated in an information theoretic way in terms of the entropy of the system. As discussed in previous sections, entropy is a function which estimates the amount of information

emitted by a source. This metric has usually been utilized for the measurement of security for contemporary biometric cryptosystems (e.g. fuzzy vault) in bits. However, the overall security has been determined in terms of the *minimum entropy* (or worst case entropy) of some public data associated with the biometric framework. This notion of minimum entropy is useful for this purpose since it captures the adversary's best strategy in predicting a random value, i.e. by guessing the most likely value. For the framework, this source of public data is the information that can be potentially extracted by an adversary while intercepting the communication channels in between the system modules. This information (in the worst case) include the perturbed responses and the net amount of noise generated corresponding to each soft trait. The minimum entropy of a source is defined as [83] -

$$H_{\infty}(S) = -\log(max_s Pr[S = s])$$

In this model, the probabilities (i.e. p_i) are the advantages that the adversary possesses for each enrolled user. These advantages are nothing but the brute force success probabilities of the adversary in determining the true values of the soft traits. Thus,

$$p_i = \frac{(nm - m)!}{n \times (nm)!}$$

In is noticeable here that the p_i values are same corresponding to each user. Thus for the framework -

$$H_\infty(S) = -\log\left(\frac{(nm-m)!}{n \times (nm)!}\right)$$

since also,

$$max_s Pr[S = s] = \frac{(nm-m)!}{n \times (nm)!}$$

Here, S is the proposed framework. For test values of n=4 and m=5,

$$H_\infty(S) \approx 22.82 bits$$

Thus the proposed privacy preserving framework also provides adequate level of security for the enrolled users.

5.9 Results and Analysis

This section presents and analyzes the results obtained from experimenting on the framework. It has been assumed for the system to work in a verification mode rather than in an identification mode. Thus during the second phase, an individual requires to produce the unique ID number prior to the biometric measurements.

5.9.1 Database

Fingerprint and face have been considered as the primary biometric characteristics due to their large-spread acceptance and use. Regarding soft biometric traits, four properties namely height, weight, age and gender were chosen.

5.9.1.1 Fingerprint

For fingerprints, the performance of the proposed secure QBBS has been evaluated on FVC 2002-DB1 [144] database. This particular database consists of 8 high quality fingerprint images of 100 individuals, thus accumulating a total of 800 images. These images were captured by the optical sensor TouchView II by Identix, and have a resolution of 500 dpi.

5.9.1.2 Face

For obtaining facial images, the AR Face Database has been used [198]. This publicly available database contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Pertaining to various constraints, a total of 26 distinct images were taken for each individual. The experiments have been performed on a sub-part of this database consisting of first 10 images for 100 individuals (50 men and 50 women). The 10 distinct features corresponding to these 10 images are - Neutral expression, Smile, Anger, Scream, left light on, right light on, all side lights on, wearing sunglasses, wearing sunglasses with left light on, and wearing sunglasses

with right light on. Thus the final face database consisted of 100×10 i.e. 1000 images.

5.9.1.3 Soft Biometrics

The secondary soft biometric characteristics consisted of 4 identifiers - height, weight, age and gender. The source of these soft attribute values is a survey which was conducted separately. The survey was carried out in the Institute of Medical Sciences (IMS), BHU, Varanasi, India. Choosing this location for this survey facilitated in keeping the sampling region both diverse and large. The data was collected corresponding to 'age' values varying from 15 to 70 years. However, the total number of samples in a single age group varied due to unavailability of individuals belonging to all the age groups. The data was also partitioned according to gender specification. In total, this set of information was collected for 700 individuals. However, 100 individuals were randomly selected from this set according to the needs.

5.9.1.4 Fusion Process

All the three databases (i.e. FVC2002-DB1, AR and survey data) were combined to create an artificial dataset of 100 individuals. The fusion process simply consisted of sequentially making a one-to-one association between the three attributes, viz. fingerprint (X), face (Y) and soft labels (Z). Since each of the biometric databases consisted of 100 individuals, this mapping was a bijective function -

$$X \rightarrow Y \rightarrow Z$$

Thus eventually, each participant i was associated with an unique biometric template

-

$$\{fingerprint_i, face_i, height_i, weight_i, age_i, gender_i\}$$

Here $i = 1, 2...100$, $fingerprint_i$ consisted of 8 images, $face_i$ consisted of 10 images, and $height_i$, $weight_i$, age_i, $gender_i$ consisted of single non-zero values.

5.9.2 Framework Analysis

The proposed scheme is progressively analyzed under four distinct precincts. The first section (Section 5.9.2.1) exhibits the usefulness of including more than one biometric trait (i.e. multimodal) in a recognition system whereas the second section (Section 5.9.2.2) deals with the inclusion of soft biometric information in the same system. The third part (Section 5.9.2.3) analyzes the effects of Laplacian noise inclusion in a QBBS and the final portion (Section 5.9.2.4) consists of a through scrutinizing of the proposed secure QBBS. For this framework, the utility factor associated with differential privacy is estimated by the performance of the QBBS itself.

5.9.2.1 Multimodal Biometrics

The feature extraction modules for both fingerprint and face produced a similarity score based on minutiae matching procedure and Euclidean distance measurements between eigenfaces respectively. These scores were then converted into conditional probabilities using the Parzen window density estimation method [199] with a Gaussian window kernel function having width 1. Let the probability values for genuine users and unauthorized users be represented as $P(s|genuine)$ and $P(s|imposter)$ respectively. The value of the posterior probability of a user based upon the score (i.e. $P(genuine|s)$) was obtained by using the formula -

$$P(genuine|s) = \frac{P(s|genuine) \times P(genuine)}{P(s)} \tag{5.5}$$

where,

$$P(s) = P(s|genuine) \times P(genuine) + P(s|imposter) \times P(imposter)$$

Both the genuine and impostor probabilities were assumed to be 0.5 (i.e. $P(genuine)$ $= 0.5$ and $P(impostor) = 0.5$). The final score was calculated by modifying Equation 5.4 as -

$$g(s) = a_{01} \log P(genuine|s_1) + a_{02} \log P(genuine|s_2) \tag{5.6}$$

Here, s_1 represents the score associated with fingerprints and s_2 corresponds to the score associated with facial data. The variables a_{01} and a_{02} were fixed appropriately to represent the contributions of fingerprint and face in the Bayesian framework respectively. These values and their associated biometric trait meanings are shown in Table 5.4. It should be noted that these constants must be subjected to the constraint

$$\sum_{i=0}^{n-1} a_i = 1$$

where n is the total number of biometric traits (including primary and soft).

Primary Trait(s)	a_{01}	a_{02}
Fingerprint	1	0
Face	0	1
Fingerprint + Face (Multimodal)	0.5	0.5

TABLE 5.4: Variation of a values for Primary traits.

Accordingly, experiments were conducted on three primary biometric systems namely: fingerprint, face, and a multi-modal system combining face and fingerprint. The performance of the biometric systems based on these three characteristics is analyzed by plotting their corresponding ROC curves in Figure 5.6. It is evident from these curves that the overall performance of the biometric system gradually gets better in the hierarchical order of face, fingerprint and (face + fingerprint). For example to a FAR of 10%, the GAR of these three modalities are approximately 85%, 95% and 98% respectively. The performance of the face based system drastically decreases after this point, even reaching a GAR of 93% for a very high FAR of 63%. This observation is normal since face based recognition systems usually suffer from many

disadvantages owing to various external constraints like occlusion and background conditions. On the other hand, the multimodal system surpasses the fingerprint based one in all the regions, especially for low acceptance rates. All these observations reaffirm the notion that multimodal frameworks perform better than their unimodal counterparts.

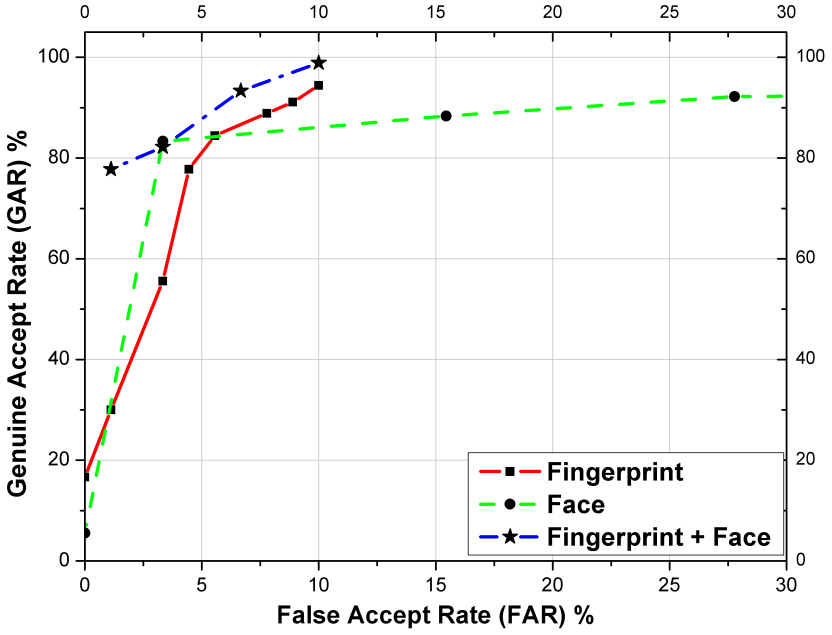


FIGURE 5.6: Performance comparison of biometric systems based on primary traits.

5.9.2.2 Inclusion of Soft Biometrics

The soft biometric information is incorporated into the basic Bayesian framework by manipulating the variables associated with the soft traits. The conditional probabilities of soft biometric attributes were calculated by using the aforementioned Gaussian function based technique. For an individual, lets denote weight by W, height by H, age by A and gender by G. Thus,

$$Y = \{W, H, A, G\} \quad \text{and} \quad Y* = \{W*, H*, A*, G*\}$$

The final score was calculated by modifying Equation 5.4 as-

$$g(s, Y, Y*) = a_{01} \log P(genuine|s_1) + a_{02} \log P(genuine|s_2) +$$
$$a_1 \log P(W*|W) + a_2 \log P(A*|A) + a_3 \log P(G*|G) + a_4 \log P(H*|H) \quad (5.7)$$

The various values of the trait wise associated weights (i.e. a_i's) in the above equation are shown in Table 5.5. These variable values were altered for studying the performance of the biometric system in the resulting distinct scenarios. The additional constraint which has been imposed here (along with the previous one) is $a_{01}, a_{02} \gg a_1, a_2, a_3, a_4$.

The advantages of incorporating soft biometric information in the system are clearly observed in the resulting ROC curves for fingerprint, face and (fingerprint + face), shown in Figure 5.7, Figure 5.8 and Figure 5.9 respectively. Evidently, the inclusion of any soft biometric trait produces a significant boost in the performance of the system. More importantly, this general observation holds true for any biometric system irrespective of the choice of primary modality. For instance, including all the soft traits results in GAR's of 99% and 98% corresponding to a FAR of 3% for the fingerprint based system and the multimodal system respectively, whereas the

Traits		**Variables**					
Primary	**Soft**	a_{01}	a_{02}	a_1	a_2	a_3	a_4
Fingerprint	Height	0.8	0	0.2	0	0	0
	Weight	0.8	0	0	0.2	0	0
	Age	0.8	0	0	0	0.2	0
	Gender	0.8	0	0	0	0	0.2
	All	0.8	0	0.05	0.05	0.05	0.05
Face	Height	0	0.8	0.2	0	0	0
	Weight	0	0.8	0	0.2	0	0
	Age	0	0.8	0	0	0.2	0
	Gender	0	0.8	0	0	0	0.2
	All	0	0.8	0.05	0.05	0.05	0.05
Fingerprint + Face	Height	0.4	0.4	0.2	0	0	0
	Weight	0.4	0.4	0	0.2	0	0
	Age	0.4	0.4	0	0	0.2	0
	Gender	0.4	0.4	0	0	0	0.2
	All	0.4	0.4	0.05	0.05	0.05	0.05

TABLE 5.5: Variation of a values for all traits.

GAR's drop down to about 45% and 81% corresponding to the same FAR when the soft information was not included. The best result in all the experiments was obtained for the face based frameworks, wherein the GAR greatly increases from 50% to 98% for a FAR of 1% upon the inclusion of all the soft information.

One important noticeable observation in these results relates to the characteristics of the soft biometric traits which have been utilized. The impact of these various traits on the system performance distinctively varies. The usefulness of the soft features is most prominent for the 'weight' and 'age' attributes, whereas the 'height' and 'gender' features have very little effect on the system performance. Interestingly, this property is apparent for all the three primary modalities. The dominance of 'weight' and 'age' could be attributed to the fact that these measurements provided a wide range of possible values. On the other hand, 'height' and 'gender' varied very

little over the entire population. In fact the 'height' values ranged approximately from 1.473 m to 1.854 m and 'gender' was limited to male and female categories (i.e. 0 and 1), thus garnering very little variance. This generic observation emphasizes the importance of a proper ordering scheme for the soft traits during the fusion process.

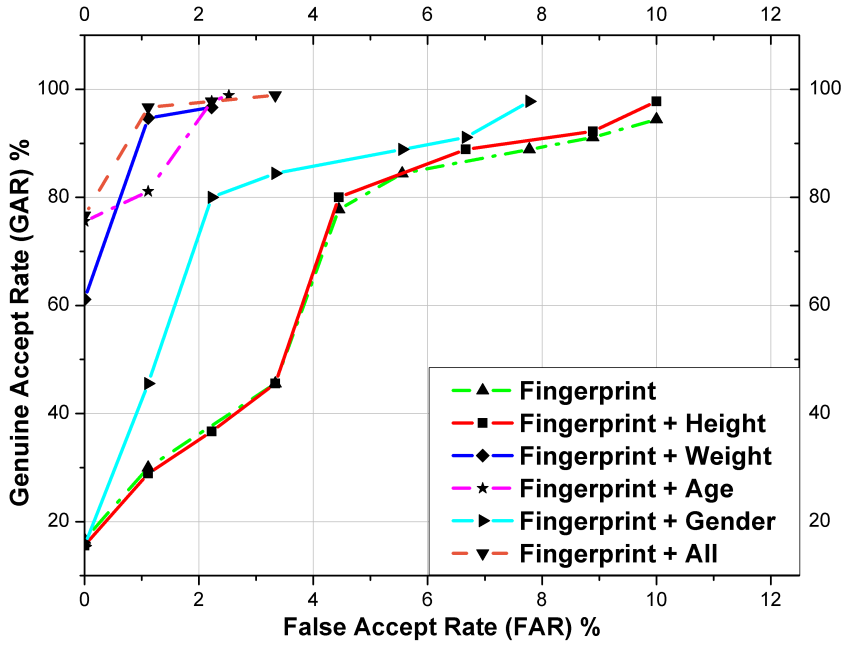


FIGURE 5.7: Performance comparison of biometric systems based on fingerprint

5.9.2.3 Noise Analysis

The security paradigm of the proposed QBBS directly depends upon the notion of differential privacy, which is based on the Laplacian noise addition mechanism. In this section, the effects of adding this random noise on both the individual traits and the whole framework are analyzed. For achieving this purpose two new parameters have been introduced, namely 'Deviation (Dev)' and 'Distortion ($Dist$)'.

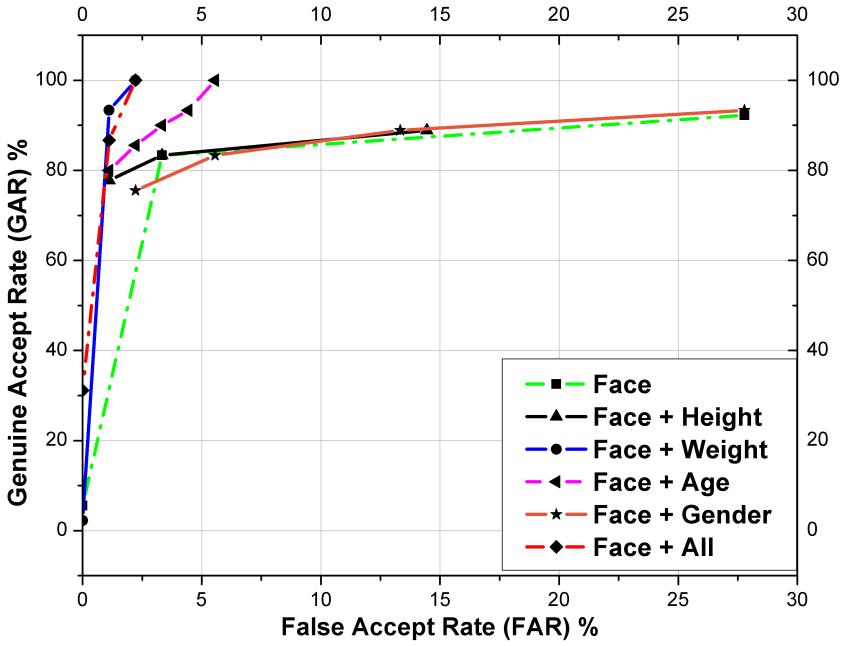


FIGURE 5.8: Performance comparison of biometric systems based on face

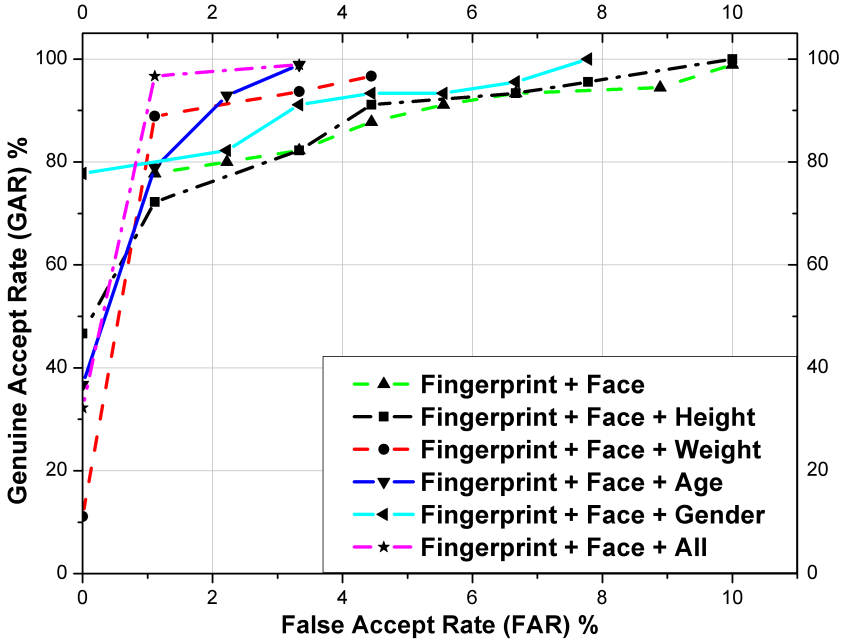


FIGURE 5.9: Performance comparison of biometric systems based on fingerprint and face

Deviation (*Dev*) in a QBBS is described as the difference between the true value and the computed value of a particular soft biometric trait. More specifically, the

true value corresponds to the original value of a soft trait which gets stored in the database during enrollment whereas the computed value is the final output of the synthesis function executed by the matcher module. In its basic form, this parameter captures the amount of noise associated with every trait. Recollecting the notations of the proposed secure QBBS,

$$Dev_i = |Y_i' - Y_i|$$

where, $|.|$ is the absolute value function. To recall, Y_i' are the trait wise estimates of the original soft trait values Y_i.

The term ϵ essentially controls the net amount of external noise which gets incorporated in the framework, thus ultimately representing the privacy level of the framework. A small ϵ value induces the addition of a large amount of noise, which consequently results in better levels of privacy. Conversely, a large ϵ value diminished the privacy levels by generating smaller quanta of noise. The trait-wise variation of Dev with ϵ is depicted in Figure 5.10. As expected, the Dev values for all the four soft traits exponentially decreases with increase in ϵ. This decreasing nature directly follows from facts that the Laplacian probability density function has an exponential distribution and ϵ is inversely related to the net amount of generated noise. Figure 5.10 displays another pattern that the generated noise for 'weight' and 'age' is comparatively much more than that for 'height' and 'gender'. This characteristic of the curves is caused due to the higher global sensitivity values for the first

two traits (i.e. weight and age), which in turn results due to the wide range of their domains.

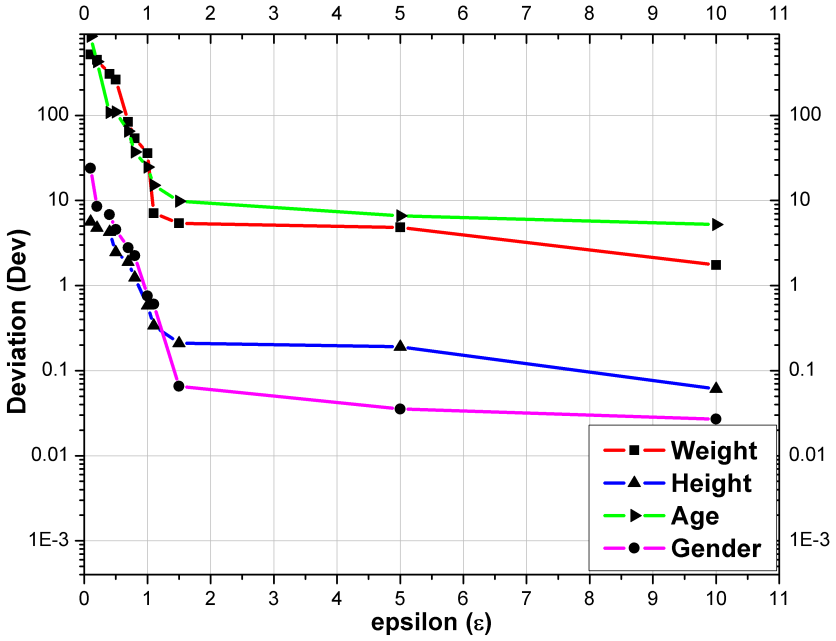


FIGURE 5.10: Effects of noise on Deviation

For the purpose of analysis on a holistic scale, the 'Distortion ($Dist$)' parameter is introduced and consequently analyzed. Distortion measures the noise introduced in the overall framework by computing the differences between true responses and perturbed responses for all the traits. This parameter is analogous to the Mean Squared Error (MSE) which is used in statistics to calculate the average of the squares of the errors of an estimator. In this scenario, the errors correspond to the noise which has been introduced in the system. Distortion is defined as -

$$Dist = \frac{1}{m \times n} \sum_{k=1}^{m \times n} (R_k - R'_k)$$

The change of *Dist* with the privacy parameter ϵ is shown in Figure 5.11. Similar to *Dev*, *Dist* decreases exponentially with increasing ϵ. This property also directly follows from the nature of Laplacian distribution, which is exponential. A low value of ϵ results in the generation of high noise content and consequently a bigger *Dist* value, whereas conversely a higher ϵ value generates low levels of noise and a smaller *Dist* value. It is up to the choice of the designer to achieve a desired level of privacy for the framework by selecting a fixed ϵ value.

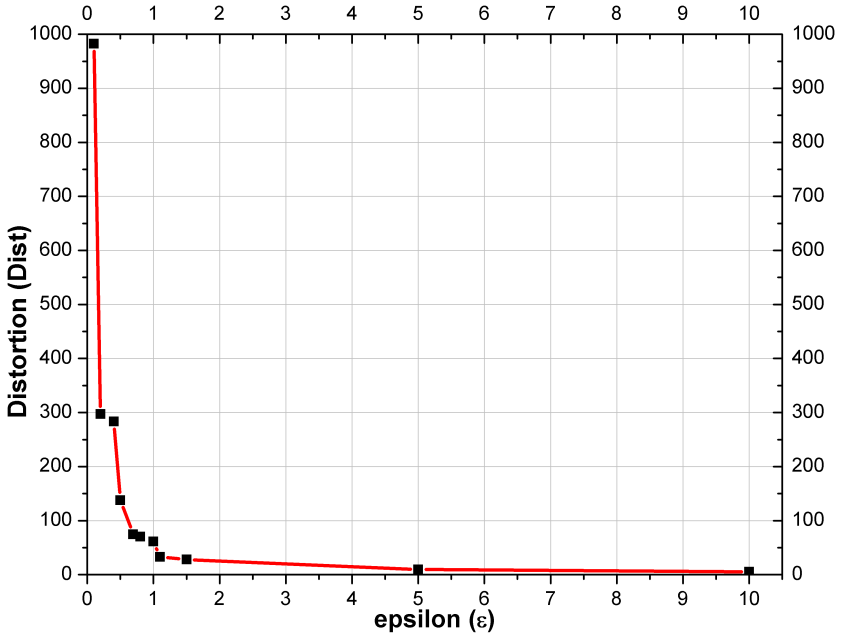


FIGURE 5.11: Effects of noise on Distortion

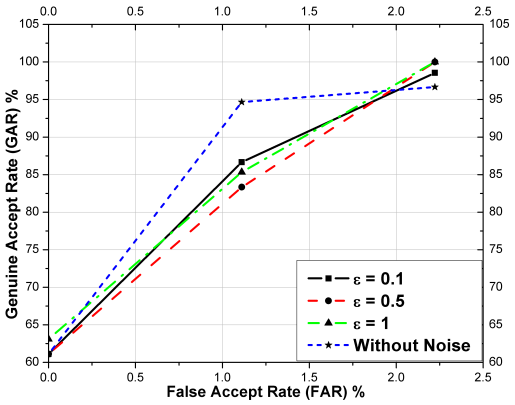
5.9.2.4 Privacy Preserving QBBS

Now the results obtained from analyzing the proposed secure QBBS are presented. Similar to the previous cases, the soft biometric information is fused here with three primary modalities (viz. fingerprint, face and multimodal combination of fingerprint

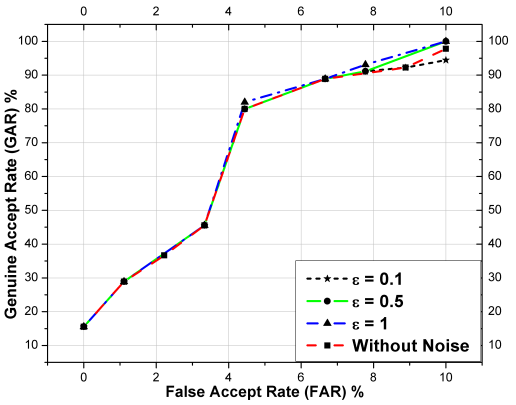
+ face). A total of 15 trait combinations were obtained for the experiments by varying the associated variables according to Table 5.5. The effects of random noise addition (by varying ϵ) on the utility of the recognition system are analyzed for each one of these separate combinations. For this case, this utility factor is defined by the system performance, which is in turn represented by the corresponding ROC curve.

The ROC curves associated with Fingerprints are shown in Figure 5.12(a), Figure 5.12(b), Figure 5.12(c), Figure 5.12(d) and Figure 5.12(e). From all these observations it is vividly evident that for fingerprint based biometric systems, the overall performance is not affected by the variation of net amount of added noise. Although there is a tiny discrepancy between the performances attributed to 'weight', 'age' and 'all soft traits', the same for 'height' and 'gender' is negligible. However, this discrepancy diminishes on reaching higher recognition rates. This is evident from Figure 5.12(a) where, between the noised and noiseless frameworks, there is a GAR difference of about 7.5% at a FAR of 1.5%, but the same disparity reduces to about 2% at a FAR of 2%. Conversely, there is a maximum difference of only 4% in the GAR across the entire range of FAR related to 'height' and 'gender' traits.

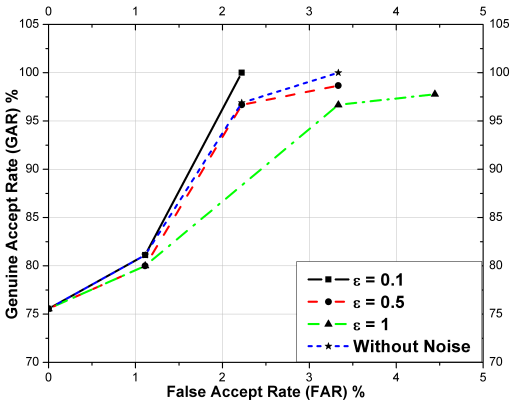
The performance for face based systems are compared in Figure 5.13(a), Figure 5.13(b), Figure 5.13(c), Figure 5.13(d) and Figure 5.13(e). All the results portray observations similar to that of the aforementioned fingerprint based system. In this scenario, the maximum discrepancy between the noised and the noiseless frameworks was observed when face was merged with the 'weight' attribute. Herein a difference of 18% in GAR is observed for low recognition rate regions (FAR = 2%), but it gets



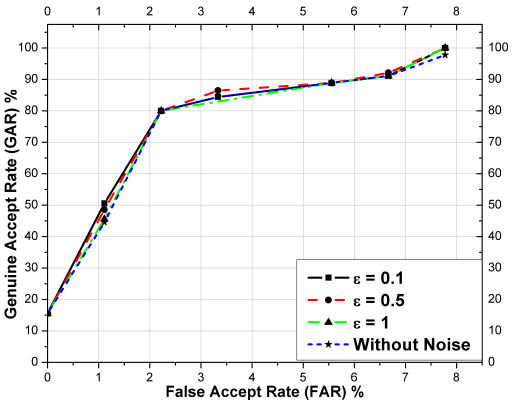
(a) Effect for Weight



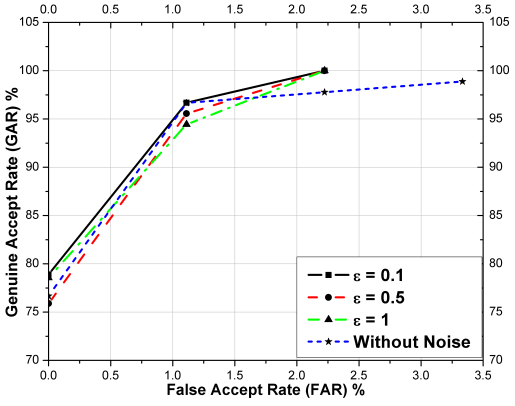
(b) Effect for Height



(c) Effect for Age



(d) Effect for Gender



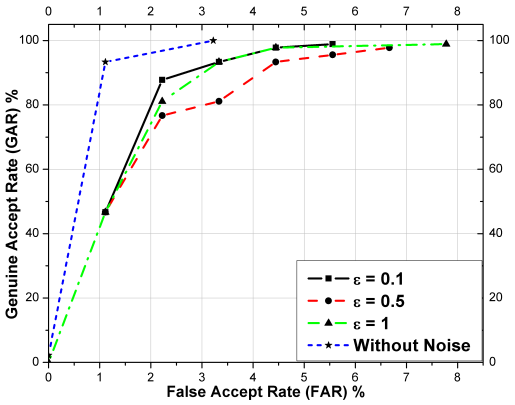
(e) Effect for All Traits

FIGURE 5.12: Effects of noise on utility for fingerprint based QBBS.

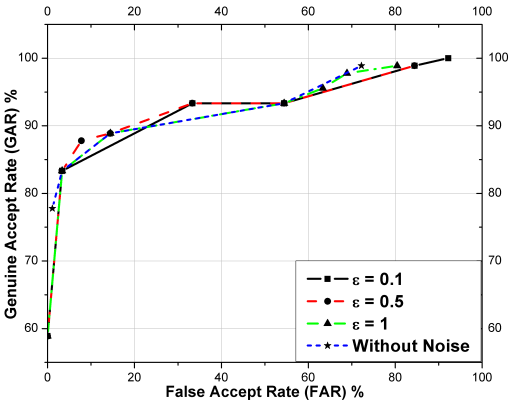
comprehensively reduced to almost 2% for higher recognition rate regions (FAR = 5%). A like observation is seen in Figure 5.13(e) wherein the difference is approximately 10% for a low FAR of 1.2%, but reduces to less than 1% for higher FAR of 2.2%. For all the other soft biometric trait based combinations, the maximum difference incurred at any point is less than 5%. In some cases (e.g. 'gender') this difference is almost zero.

The performance graphs for soft traits fused with the multimodal combination of face and fingerprints are shown in Figure 5.14(a), Figure 5.14(b), Figure 5.14(c), Figure 5.14(d) and Figure 5.14(e). The obtained ROC curves also strongly vindicate the idea of noise having no (or little) effect on the performance of the system. The maximum difference in the GAR's (at the same FAR) for noise based and noise free frameworks is less than 8% for all the combinations. This worst case occurs at a FAR of 1.1% during the fusion of all the soft biometric traits. The maximum GAR differences occurring for other individual soft traits are approximately 6%, 1%, 2% and 7% corresponding to 'age', 'height', 'gender' and 'weight' respectively. However similar to the previous two cases, these difference minimize for high recognition regions. For instance, the same combinations result in the GAR differences of 4%, 1%, 1.5% and 2% for higher FAR values.

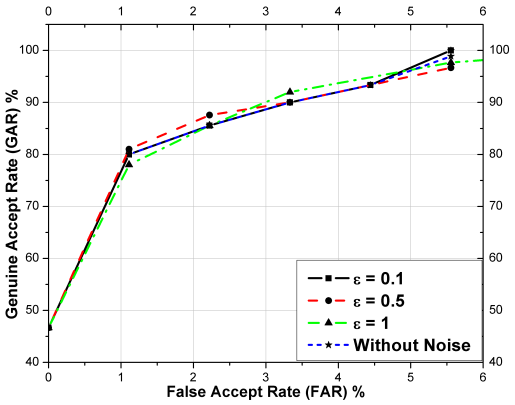
Thus it can be generally stated that for lower error rates, the addition of Laplace noise to a QBBS has some minor effects on the overall performance. However if shifted towards a higher recognition region (which is more common in commercial



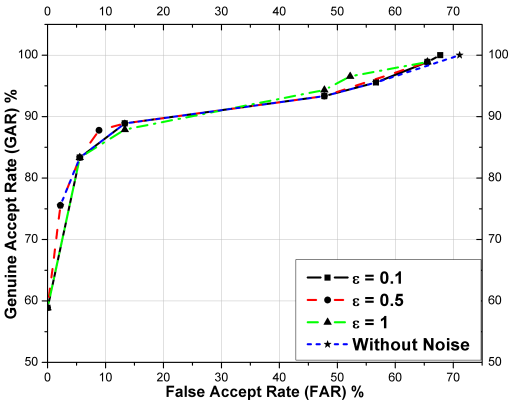
(a) Effect for Weight



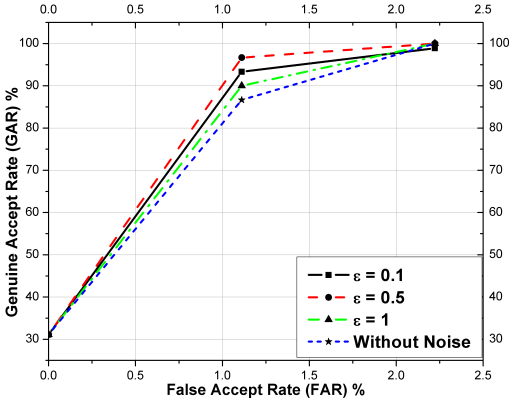
(b) Effect for Height



(c) Effect for Age

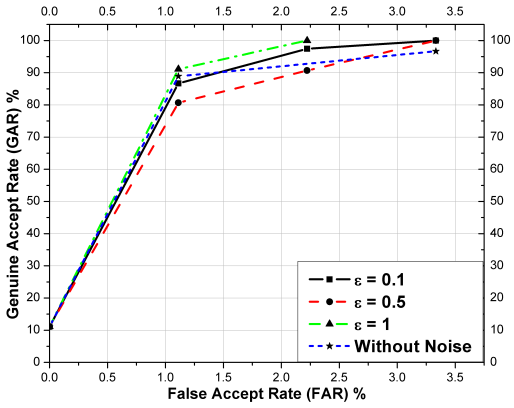


(d) Effect for Gender

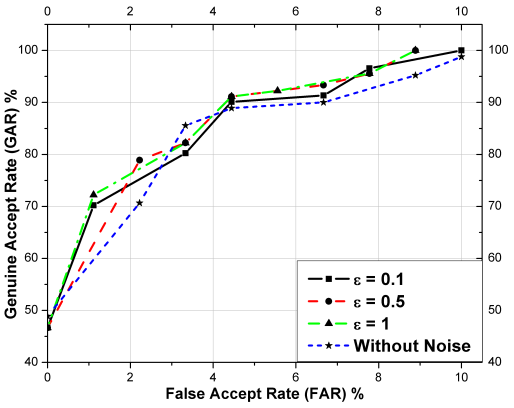


(e) Effect for All Traits

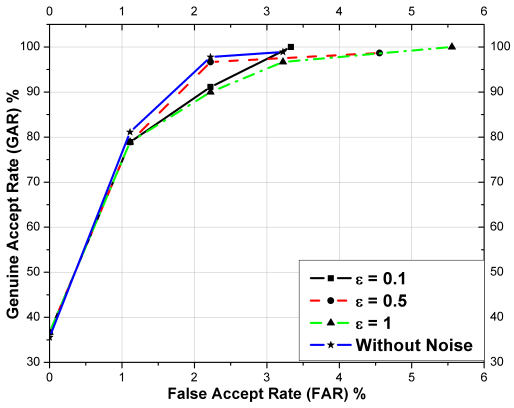
FIGURE 5.13: Effects of noise on utility for face based QBBS.



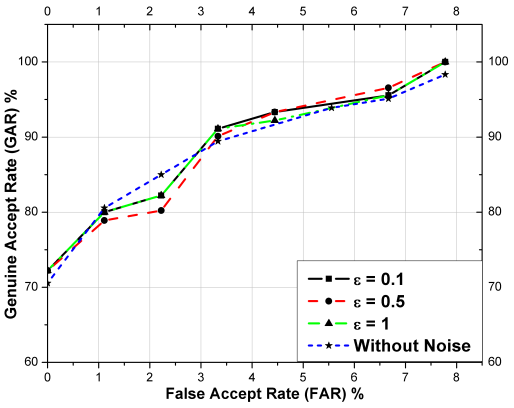
(a) Effect for Weight



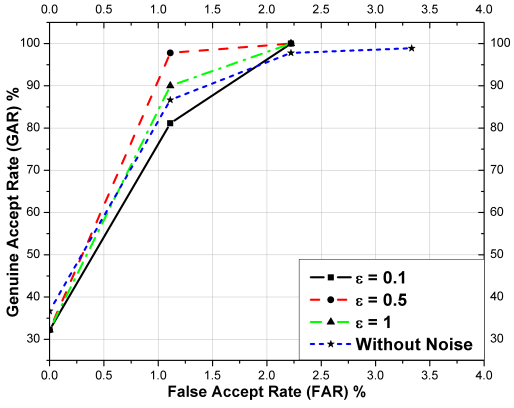
(b) Effect for Height



(c) Effect for Age



(d) Effect for Gender



(e) Effect for All Traits

FIGURE 5.14: Effects of noise on utility for fingerprint and face based QBBS

applications), the noise has practically no significant effect on the system performance. This demonstrates that the proposed approach does not mitigate the utility of the biometric system, while providing adequate (and variable) levels of user privacy.

5.10 Conclusion

In the modern technological world, protecting or safeguarding the privacy of individuals is an important factor. This chapter addresses this issue in a soft biometric based system by a two fold approach. The first part is dedicated to developing a formal model which quantifies privacy loss in such scenarios, whereas the second part deals with the construction of a privacy preserving framework for the same.

Explicitly speaking, the privacy levels of individual users have been quantified if a soft biometric database gets leaked. The privacy threats conceivable in such situations is very practical and this study successfully characterizes them. In this work, not only a theoretical framework for modeling such scenarios has been developed but also the achievable privacy levels have been defined. While constructing such a framework, the various possibilities in which an adversary may try to learn sensitive information about an individual have been considered and the appropriate levels of privacy in each case have been estimated. A potential solution to this privacy problem has also been proposed in the second part of this chapter. For doing so, an interactive biometric framework termed as a Query Based Biometric System (QBBS)

was initially developed. The final privacy preserving design is basically an implementation of ϵ- differential privacy in this QBBS. The correctness of this scheme is validated since the proposed framework compares the user's data (obtained in both the enrollment and verification phases) in a transformed domain. The analysis of the obtained results indicates that the suggested framework achieves its aim of privacy preservation while retaining the original recognition accuracy rates.

The basic cause for the majority of privacy issues points to the multiple attribute based links which exist between a micro database and a soft biometric database. As long as people would continue providing their personal information in various events, these linking and consequently the privacy risks would increase proportionally. This phenomenon is more apparent in social network based scenarios where a plethora of personalized data is available for mining. To put it briefly, privacy preservation of individuals is a complex but necessary requirement and this study tries to model one such practical circumstance along with providing a viable solution.