DONNÉES BIOMÉTRIQUES ET LEUR SÉCURISATION

1 Introduction

La biométrie est l'utilisation de la physiologie et / ou du comportement pour déterminer ou vérifier l'identité des individus. Malgré les avantages de ces systèmes biométriques par rapport aux systèmes d'authentification traditionnels qui utilisent des mots de passe et des cartes d'identité, ils sont toujours vulnérables à des limitations spécifiques qui peuvent dégrader considérablement leurs fonctionnalités.

Dans ce chapitre, nous décrierons, la notion de la biométrie, en parlant des modalités biométriques, d'architecture des systèmes biométriques, des mesures de performance d'un système biométrique et en terminant par une description de la sécurité des systèmes biométriques.

2 Définition

La biométrie désigne une technique d'identification et d'authentification qui consiste à transformer une caractéristique biologique, morphologique ou comportementale en une clé d'identificateur unique. Son objectif est d'attester l'unicité d'une personne à partir de la mesure d'une partie inchangeable ou immaitrisable de son corps [1] . Autrement dit, c'est une reconnaissance automatisée des individus en fonction de leurs caractéristiques biologiques et comportementales.

Pour que la reconnaissance soit envisageable, fiable et de qualité, les caractéristiques doivent au moins garantir les conditions suivantes [2, 3] :

- **Universelles :** exister chez tous les individus ou la population.

- **Uniques :** permettre de différencier un individu par rapport à un autre.
- **Permanentes ou persistantes :** autoriser l'évolution dans le temps.
- **Enregistrables :** collecter les caractéristiques d'un individu (avec l'accord de celui-ci).
- **Mesurables :** autoriser une comparaison future.
- **Non-reproductibles :** la facilité ou non à falsifier une modalité biométrique.

3 Modalités biométriques

On peut compter un grand nombre de modalités biométriques, qui peuvent être regroupées en trois grandes catégories : physiologique (ou morphologique), comportementale et biologique.

3.1 Physiologique (ou morphologique)

Les biométries morphologiques sont les biométries qui utilisent une partie du corps humain[1]. Cette catégorie regroupe la reconnaissance de :

- **L'empreinte digitale** : ce procédé est le plus répandu et le plus ancien[4]. La donnée de base est le dessin représenté par les crêtes (les lignes dessinés sur la peau) et les vallées (espaces entre les crêtes), et de l'épiderme (jonctions, terminaisons aveugles, croisements). Une empreinte est caractérisée par une centaine de points particuliers portés par les crêtes (appelés minuties), dont un nombre de minuties entre (15 et 20) correctement localisées suffisent pour une identification. Certains modules de reconnaissance d'empreintes vérifient la température du doigt, sa conductivité, les battements de cœur, ainsi que d'autres paramètres biologiques.

FIGURE 1.1 – Empreinte digitale avec crête et vallée marquées

- **L'oreille** : les oreilles, comme d'autre parties du corps, présentent aussi une empreinte unique lorsqu'on les met sous contrainte contre une surface. Ce système se base sur l'identification de la forme et des dimensions de l'oreille externe (pavillon, hélix, tragus, etc..). En effet, même si celles-ci sont uniques, elles grandissent de 0 à 20 ans et au-delà de 50 ans tout en se déformant légèrement.

• **L'iris** : c'est la région annulaire située entre la pupille et le blanc de l'œil. La biométrie par ce trait est la plus récente et efficace. Les motifs de l'iris se forment au cours des deux premières années de la vie et sont stables.

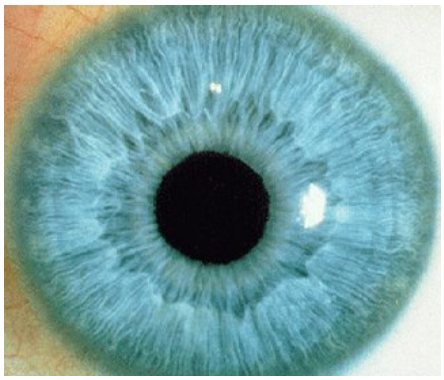

FIGURE 1.2 – Iris

• **La rétine** : elle est très peu utilisée et elle a été moins bien acceptée par le public à cause de la mesure qui doit s'effectuer à très faible distance du capteur [4] (Figure 1.3). Cette technique se base sur le fait que le schéma et le dessin formés par les vaisseaux sanguins de la rétine sont uniques pour chaque personne et assez stables toute la vie[4].

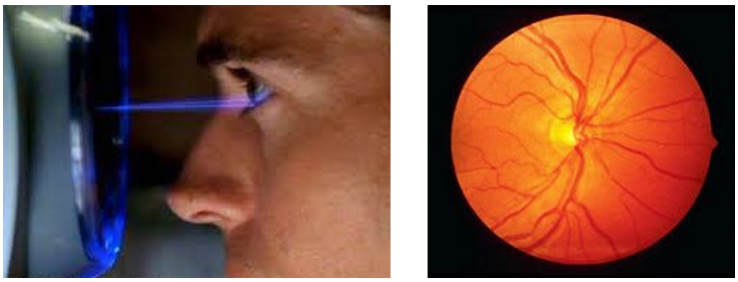

FIGURE 1.3 – Reconnaissance de la rétine

• **Géométrie de la main (hand-scan)** : c'est l'un des mesures biométriques les plus répandus, cela consiste à mesurer plusieurs caractéristiques de la main (jusqu'à 90) tel que la forme de la main, longueur et largeur des doigts etc. La technologie associée à cela est principalement de l'imagerie infrarouge[4].

• **Visage** : rien n'est plus naturel qu'utiliser le visage pour identifier une personne. C'est la biométrie la plus commune et la plus populaire. Elle implique la métrique des et entre caractéristiques distinctes dans le visage, se fondant moins sur des facteurs d'une nature changeante tels que la coupe des cheveux ou l'utilisation des produits de beauté. Néanmoins, le visage humain est sujet au changement avec le temps et cette réalité demeura un défi pour les systèmes d'identification de visage, comme le changement d'expression, la maladie, la vieillesse et d'autres facteurs normaux. En outre, les

FIGURE 1.4 – Géométrie de la main

facteurs humains et environnement joueront un très grand rôle dans l'efficacité d'un système de reconnaissance faciale.

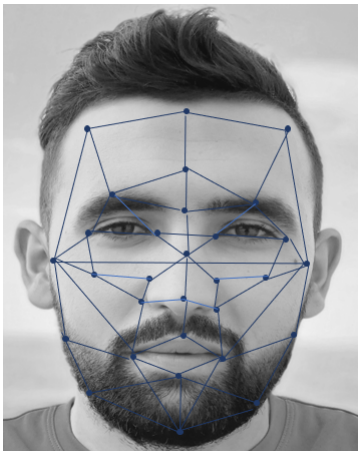

FIGURE 1.5 – Reconnaissance de visage

• **Reconnaissance vocale** : cette technique est très facilement falsifiable, en utilisant un enregistrement. La measure biométrique de la voix traite des données qui proviennent à la fois de facteurs physiologiques dépendants de l'âge, du sexe, de la tonalité, de l'accent, et de facteurs comportementaux comme la vitesse et le rythme. Ces éléments ont l'avantage d'être stables dans la vie d'un individu [5].

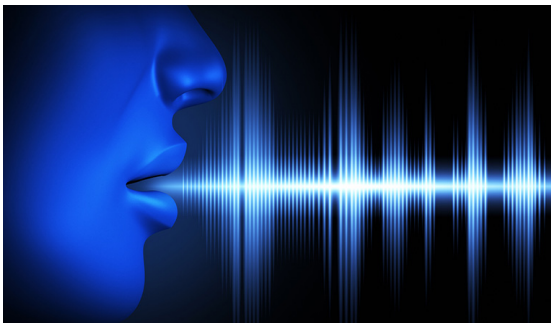

FIGURE 1.6 – Reconnaissance vocale

• **Reconnaissance Palmaire « Palmprints »** : Palmprint est l'une des nouvelles modalités biométriques les plus efficaces et qui s'appuie sur la texture de la paume de la main. Récemment, il a été montré que les lignes principales et les rides dans une image palmprint sont uniques. En général, la plupart des gens ont trois lignes principales : la ligne du cœur, la ligne de tête et la ligne de vie. Les rides sont considérées comme les modèles de ligne les plus fins et les plus irréguliers. Les rides prononcées autour des lignes principales, peuvent également contribuer à la discrimination de palmprint.

FIGURE 1.7 – Palm print

• **Veines** : le motif des veines du doigt ou de la paume de la main sert de critère d'authentification des personnes. Grâce à un scanner infrarouge et une caméra grand angle intégrée, le système capte, en quelquesmillisecondes, la structure veineuse et donc l'identité univoque d'une personne.

FIGURE 1.8 – Reconnaissance des veines

3.2 Comportementale

Cette catégorie utilise un trait personnel du comportement [6]. Se base sur l'analyse de certains comportements d'une personne. Elle concerne l'étude des actions répétitives et usuelles des personnes. On peut compter les suivants :

• **Reconnaissance de la dynamique de la frappe au clavier** : dans cette technique les durées entre frappes, la fréquence des erreurs et la durée de la frappe elle-même sont étudiées de façon statistique. En revanche, cette technologie est tributaire de l'état physique et psychique de la personne qui utilise le clavier. La fatigue, le stress sont autant de facteurs qui feront varier la qualité de la frappe.

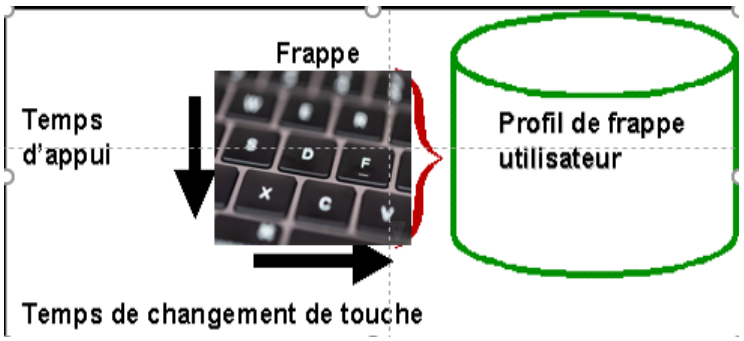

FIGURE 1.9 – Reconnaissance de la dynamique de la frappe au clavier

• **Reconnaissance de la dynamique de signature** : dans ce système d'identification, l'utilisateur doit signer avec un stylo électronique sur une tablette graphique, le système analyse ensuite les variations de vitesse du stylo, ses accélérations et ses pressions sur la tablette[4]. Le point faible de cette technique est qu'un individu qui ne signe pas toujours de la même façon se verra souvent refuser l'accès au système.

FIGURE 1.10 – Reconnaissance de la dynamique de signature

• **Reconnaissance de la démarche** : chaque être humain à une façon très personnelle de démarche qui peut être modélisée en se basant sur plusieurs éléments tels que la vitesse, l'accélération, les mouvements du corps, etc. La marche peut être aussi affectée par plusieurs facteurs comme le choix des chaussures, la surface de marche et les vêtements. Les systèmes de reconnaissance de la démarche, qui sont encore au stade de développement, utilisent le traitement d'image afin de détecter la silhouette humaine et les attributs spatiotemporels associés.

3.3 Biologique

Une biométrie de cette catégorie est basée sur l'identification de traits biologiques particuliers qui, pour toutes personnes, sont uniques et permanents [7]. Ce type de biométrie est très complexe à mettre en œuvre dans un système usuel de reconnaissance et n'est utilisé que dans un cas d'extrême nécessité (ex : Enquête criminelle, test de paternité... etc.) [8]. Cette catégorie regroupe :

• **Reconnaissance de l'ADN** : présent dans les cellules du corps, il est spécifique d'un individu à un autre et permet de l'identifier de manière certaine à partir d'un simple fragment de peau, d'une trace de sang ou d'une goutte de salive. Actuellement, le temps requis pour une analyse et le coût associé à celle-ci restreignent son utilisation dans des domaines autres que celui de l'identification judiciaire. Cependant, ce procédé biométrique fait l'objet de recherche intensive puisqu'il représente la technologie d'identification par excellence avec une marge d'erreur bien en dessous des autres moyens biométriques.

FIGURE 1.11 – Reconnaissance de l'ADN

• **Reconnaissance de l'odeur** : chaque personne dégage une odeur particulière définie par des composantes chimiques. Les systèmes biométriques basés sur cette modalité analysent ces composants pour extraire des données comparatives.

• **Reconnaissance de la thermographie faciale** : une caméra thermique est utilisée pour réaliser un cliché infrarouge du visage, ce qui permet de faire apparaître une répartition de la chaleur unique à chaque individu , voire de cartographier le réseau veineux du visage invisible à l'œil nu. Cette technique permet de distinguer même les vrais jumeaux.

FIGURE 1.12 – Reconnaissance de la thermographie faciale

Remarque
Les modalités biométriques peuvent être regroupées aussi selon la coopération ou non de l'individu [9], on peut trouver :

— **Techniques intrusives** : ces techniques requièrent un contact physique avec l'individu pour l'identifier, tels que les empreintes digitales, la rétine, l'iris ou la forme de la main. Leur usage est généralement mal accepté.

— **Techniques non intrusives** : ces techniques ne requièrent pas la coopération de l'individu en question, leur application peut se faire à distance en utilisant des capteurs qui ne nécessitent pas de contact direct avec l'utilisateur (visage, démarche, ...).

4 Architecture des systèmes biométriques et modes de fonctionnements

Un système biométrique est essentiellement un système de reconnaissance de formes. Ce système fonctionne en acquérant des traits biométriques, construisant des modèles et ensuite en comparant ces modèles par les caractéristiques stockées au préalable dans une base de données pour pouvoir enfin exécuter une action ou prendre une décision à partir du résultat de cette comparaison [1].

4.1 Architecture des systèmes biométriques

Un système biométrique est composé de 5 modules suivants (voir figure 1.13) :

a. **Module d'acquisition ou capture** : Le module d'acquisition peut mesurer les caractéristiques biométriques d'origine à laide de caméras, lecteurs d'empreintes digitales, caméras de sécurité,...etc. Pour des raisons d'efficacité et de rapidité, des traitements préliminaires ont été effectués à ce niveau [1].

b. **Module de pré-traitement** : Il consiste en un prétraitement et une atténuation du bruit, et par l'application d'une série d'opérations continues (comme le filtrage, la normalisation, etc.) pour faire apparaître des paramètres pertinents et des paramètres utiles [1].

c. **Module d'extraction des caractéristiques** : Sert à représenter les données biométriques prétraitées dans l'étape précédente par de nouvelles représentations ou ce qu'on appelle les modèles. Ces modèles sont obtenus par l'extraction des caractéristiques les plus pertinents. Idéalement, ces modèles devrait être unique à chacun et relativement constante pour les changements intra-classe [1].

d. **Module du stockage** : Qui contient l'ensemble des modèles biométriques des utilisateurs enrôlés du système. En principe, les informations stockées ne sont

jamais les images d'origine, mais un modèle mathématique des éléments qui distinguent l'échantillon biométrique d'un autre [1].

e. **Module de Matching et de décision** : Il s'agit de la dernière étape, ou nous pouvons prendre les décisions appropriées en fonction des exigences de l'application, aprés le calcul de la similitude entre le et la base de référence [1].

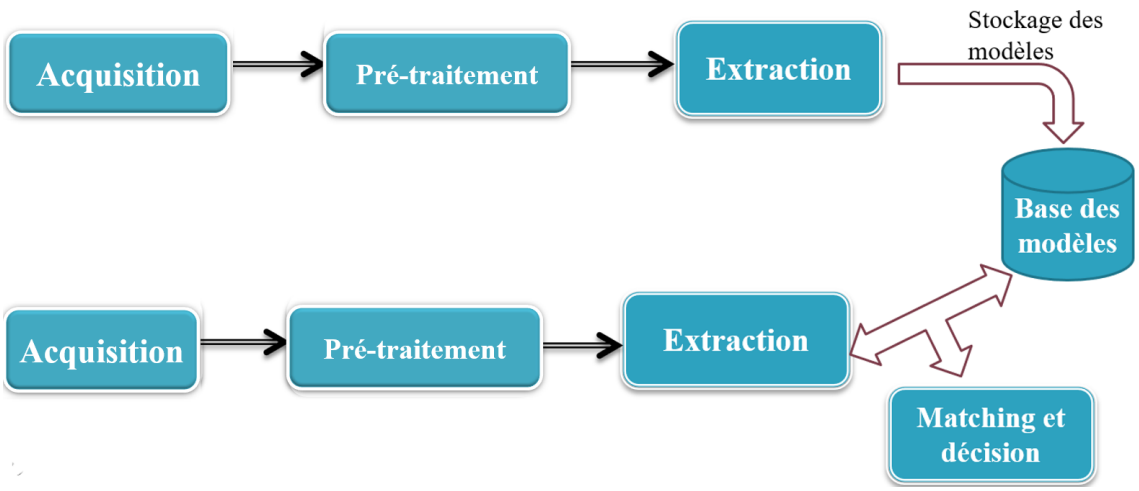

FIGURE 1.13 – Architecture d'un système biométrique

4.2 Modes de fonctionnement

Les systèmes biométriques peuvent fonctionner en deux modes principaux : l'authentification (vérification) et l'identification. Il existe une étape avant les deux modes précédents qui s'appelle "l'enrôlement".

a . **Enrolement**

C'est la première phase de tout système biométrique, il s'agit de l'étape pendant laquelle un utilisateur est enregistré dans le système pour la première fois et où une ou plusieurs modalités biométriques sont capturées et enregistrées dans une base de données [4] (voir la figure 1.14), cet enregistrement peut s'accompagner par l'ajout d'information biographique dans la base de données.

b . Authentification

Permet de prouver l'identité revendiquée par un utilisateur(voir la figure 1.15). Le système doit répondre à une question de type : "Suis-je bien la personne que je prétends être ? ". Techniquement, le dispositif vérifie par rapport à un code (identifiant) saisi sur un clavier, ou lu par le passage d'un badge (carte à puce, magnétique, proximité, etc.) que l'échantillon biométrique fourni correspond bien au gabarit désigné par l'identifiant[4].

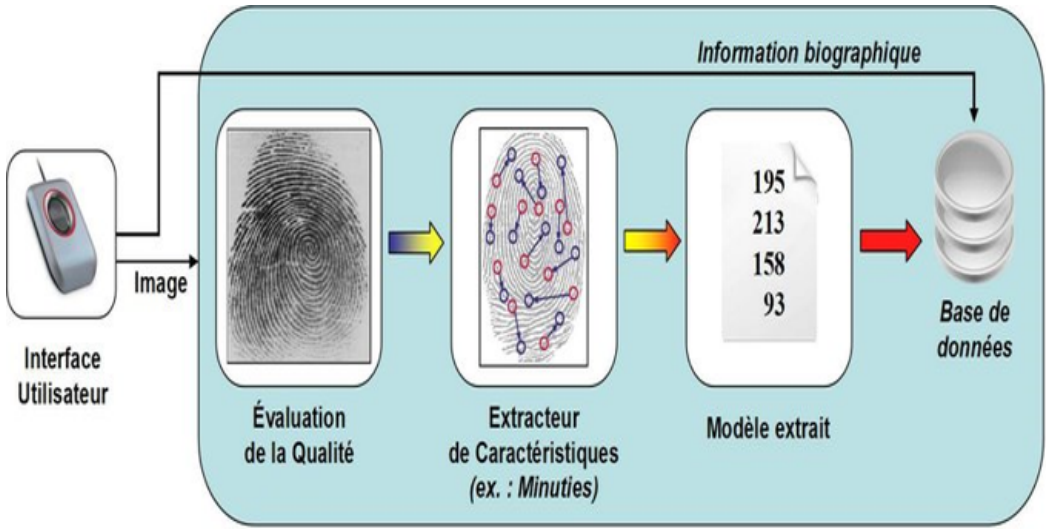

FIGURE 1.14 – Enrolement d'une personne dans un système biométrique [10]

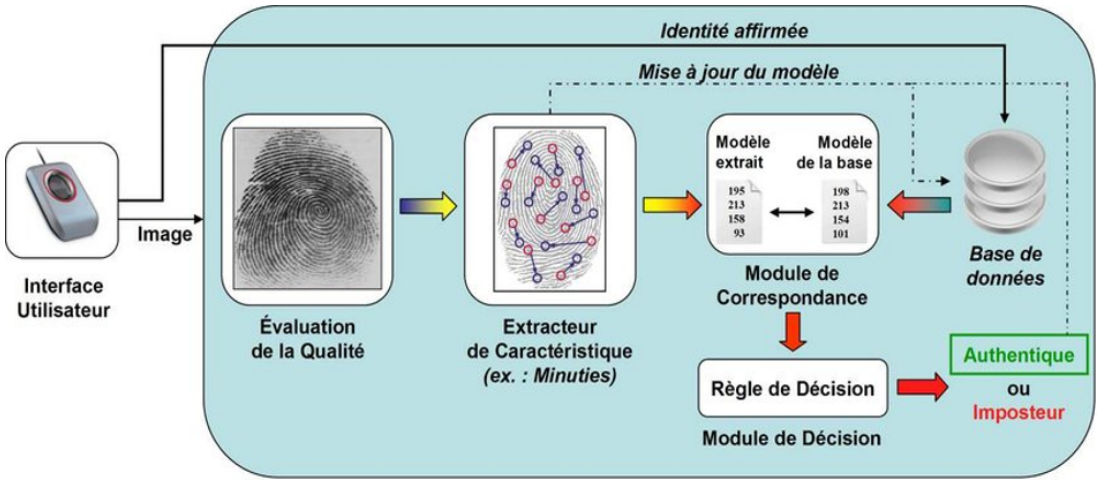

FIGURE 1.15 – Authentification d'un individu dans un système biométrique [10]

c . Identification

Permet de vérifier que l'identité d'un individu qui se présente existe bien dans la base de référence [9]. Le système doit deviner l'identité de la personne. Il répond donc à une question de type "Qui suis-je ? ". À partir de l'échantillon biométrique fourni, le dispositif cherche le gabarit correspondant dans sa base de données[4].

L'identification et l'authentification sont donc deux problèmes différents. L'identification peut être une tâche redoutable lorsque la base de données contient des millions d'identités, tout particulièrement lorsqu'il existe des contraintes de type temps réel sur le système. Ces difficultés sont analogues à celles que tend à résoudre les systèmes d'indexation de documents multimédias [11].

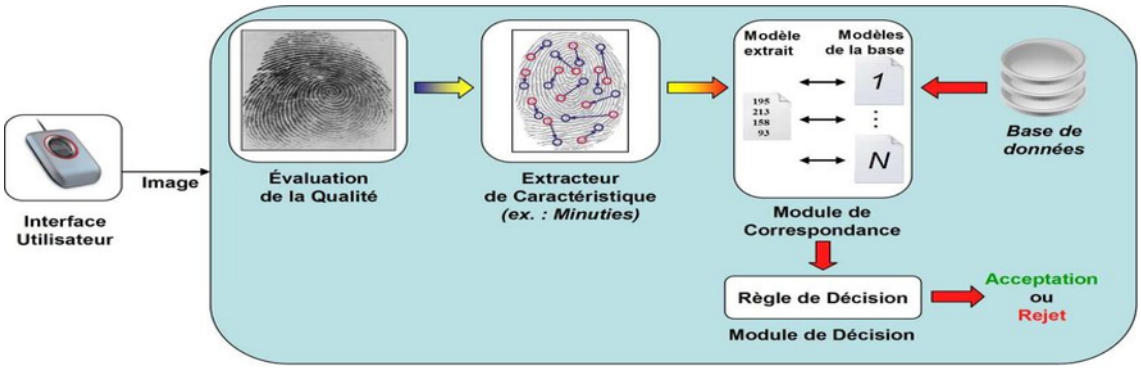

FIGURE 1.16 – Identification d'un individu dans un système biométrique [10]

5 Mesure de performance d'un système biométrique

En biométrie, chaque système est en face de deux populations :

1) Les clients appartenant au système, ceux qui sont autorisés à pénétrer dans la zone protégée.
2) Les imposteurs n'appartenant pas au système, mais généralement qui essayent de rentrer [7].

Pour évaluer les performances d'un système biométrique, plusieurs mesures sont employées . Les exigences des applications sont diverses, et par conséquent un tel système de reconnaissance assurant certains critères est recommandé pour telle ou telle application. Les critères principaux utilisés pour évaluer la performance des systèmes de reconnaissance biométriques sont [12] [13] :

- La fiabilité qui est mesurée par des taux d'erreurs et des courbes de performances.
- L'efficacité (rapidité), qui est mesurée par le temps CPU et l'espace mémoire.
- L'exigence en termes de quantité et de qualité d'exemples d'apprentissage et de test.

Dans la section qui suit, un aperçu sur les mesures d'évaluation des systèmes biométriques est présenté.

5.1 Taux d'erreur

Les systèmes d'authentifications sont généralement évalués par le taux de faux rejets et le taux de fausses acceptations. Tandis que les systèmes d'identifications peuvent être évalués par le taux d'identification, taux de faux-négatif d'identification, taux de faux-positif d'identification, et erreur de l'algorithme de présélection.

5.1.1 Taux d'erreur de systèmes d'authentification

On peut trouver plusieurs métriques pour mesurer la performance d'un système biométrique donné lors d'authentification. Les plus importants sont : le taux de fausse acceptation (TFA), le taux de faux rejet (TFR) et le taux d'erreur égal (TEE) :

a. **TFA (FAR)** : Taux de Fausses Acceptations, ("False Accept Rate" ou FAR) ; ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système.

$$TFA = \frac{Nombre\,imposteurs\,acceptes\,(FA)}{Nombre\,total\,d'accesimposteur}$$

b. **TFR (FRR)** : Taux de Faux Rejets, ("False Reject Rate" ou FRR). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système.

$$TFR = \frac{Nombre\,de\,client\,rejetées}{Nombre\,total\,d'acces\,clients}$$

c. **TEE (EER)** : Taux d'Erreur Egale, ("Equal Error Rate" ou EER). Donne un point sur lequel : TFA = TFR.

$$TEE = \frac{(Nombre\,de\,fausses\,acceptations\,(FA) + Nombre\,de\,faux\,rejets(FR))}{Nombre\,total\,d'acces}$$

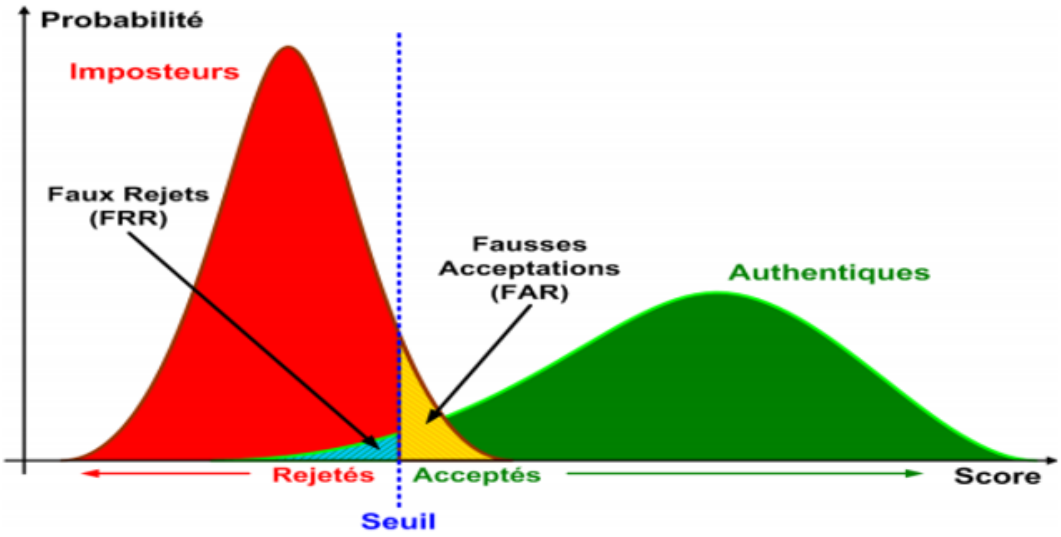

FIGURE 1.17 – Illustration du FRR et FAR
[4]

5.1.2 Taux d'erreur de systèmes d'identification

Dans le cas des systèmes d'identification, on peut trouver les taux suivants :

a. **Taux d'identification (identification rate, IR)** : Appelé aussi « taux de reconnaissance ». Il est donné sous forme de taux d'identification de rang-1(Rank-1). Il présente la proportion de tentatives d'identification authentiques pour lesquelles l'inscription correcte est indiquée dans la liste des identifiants [14] [12].

$$Rang-1 = \frac{N_i}{N}.100\%$$

Où Ni représente le nombre d'images attribuées avec succès à l'identité correcte (bien classées) et N représente le nombre total d'images essayant d'assigner une identité [12].

b. **Taux de faux-négatif d'identification (false-negative identification-error rate, FNIR)** : Proportion de transactions d'identification, par des utilisateurs enrôlés dans le système, pour lesquels l'identifiant de l'utilisateur ne figure pas dans la liste des identifiants retournée [12].

c. **Taux de faux-positif d'identification (false-positive identification-error rate, FPIR)** : Proportion de transactions d'identification, par des utilisateurs non enrôlés dans le système, pour lesquels la liste des identifiants retournée est non vide [12].

d. **Erreur de l'algorithme de présélection (pre-selection error)** : L'algorithme de présélection permet de réduire le nombre de modèles biométriques à comparer avec l'image acquise pendant la phase d'identification. L'erreur de l'algorithme de présélection est l'erreur qui se produit quand le modèle correspondant à la donnée biométrique acquise ne figure pas dans la liste retournée des modèles [12].

5.2 Courbes de performance

Pour évaluer la performance d'un système biométrique par courbes, on peut trouver les courbes suivantes selon le mode de fonctionnement :

a. **Pour l'authentification**

⇒ **Courbe ROC (« Receiver Operating Characteristic » en anglais)** : La courbe ROC (voir figure 1.18) trace le taux de faux rejet en fonction du taux de fausse acceptation. Plus cette courbe a tendance à suivre la forme de l'indice de référence, plus le système est efficace, c'est-à-dire avec un taux de reconnaissance global élevé.

b. **Pour l'identification**

⇒ **Courbe CMC ("Cumulative Match Characteristic" en anglais)** : La courbe CMC donne le pourcentage de personnes reconnues en fonction d'une

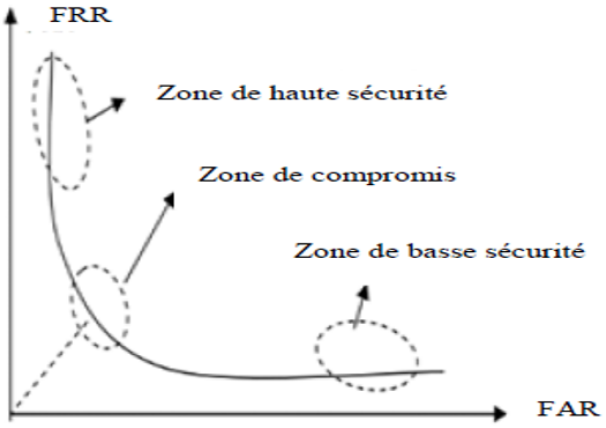

FIGURE 1.18 – Courbe ROC
[4]

variable que l'on appelle le rang. On dit qu'un système reconnaît au rang 1 lorsqu'il choisit la plus proche image comme résultat de la reconnaissance. On dit qu'un système reconnaît au rang 2, lorsqu'il choisit, parmi deux images, celle qui correspond le mieux à l'image d'entrée, etc. On peut donc dire que plus le rang augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible. Le taux d'identification de rang-n pour différentes valeurs de n peut être résumé en utilisant la courbe CMC (voir figure 1.19). Où n varie de 1 à N. N est le nombre d'utilisateurs dans la base de données [10].

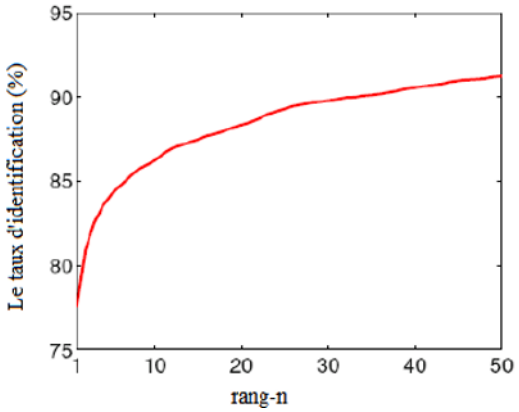

FIGURE 1.19 – Courbe CMC

Malgré les bienfaits d'un système biométrique, il nécessite d'être sécurisé. La section suivante permet d'expliquer ce point, dont on parle de différents cotés telle que les menaces et les vulnérabilités, les problèmes de sécurité de ces systèmes et leurs sécurisation.

6 Sécurité des systèmes biométriques

Au cours de temps et pour plus de sécurité, les systèmes biométriques sont de plus en plus utilisées dans de nombreuses applications. Malgré les avantages des systèmes biométriques par rapport aux systèmes d'authentification traditionnels (pin, mot de passe...), ils restent vulnérables (n'est pas sécurisés à 100%) à des attaques spécifiques. Ces derniers peuvent dégrader considérablement leur fonctionnalité. Dans cette section, nous présentons premièrement des vulnérabilités et menaces liées à l'utilisation des modèles biométriques, puis nous passons à exposer les principaux points d'attaques de ces modèles, et nous terminons par une présentation des méthodes de sécurisation d'un système biométrique.

6.1 Modèle biométrique : vulnérabilités et menaces

Les systèmes biométriques ont plusieurs faiblesses où un modèle biométrique est enregistré dans la base de données sans aucune protection. Parmi les menaces et vulnérabilités qui touchent les modèles biométriques on peut citer :

A. **Risques de violation de la vie privée**

L'analyse de la conformité de la confidentialité d'un système de reconnaissance automatique basé sur la biométrie est un problème principal à la fois pendant le processus de conception du système et pour son déploiement dans des applications réelles. On peut citer les préoccupations principales suivants liées à l'utilisation de la biométrie :

- Les données biométriques peuvent être collectées ou partagées sans l'autorisation spécifique d'un utilisateur, des connaissances adéquates ou sans objectif spécifique[15].
- Les données biométriques, qui ont été collectées à des fins spécifiques, peuvent être utilisées ultérieurement à une autre fin non voulue ou non autorisée.
- L'utilisation de la biométrie peut violer le « principe de proportionnalité » [16], qui stipule que les données biométriques ne peuvent être utilisées que si elles sont adéquates, pertinentes et non excessives par rapport à l'objectif du système.
- Les données biométriques peuvent être mal stockées et / ou transmises. Cela exposerait les données biométriques à des attaques externes.

B. **Risques d'usurpation d'identité**

Le principe est qu'un individu collecte les informations biométriques d'un autre et se fabrique une « fausse identité », parce qu'il est parfois possible de contrefaire des mesures biométriques de manière artisanale par différentes techniques[17].

Une autre technique d'usurpation d'identité « attaque par rejeu », qui consiste

à contourner la capture de l'image biométrique, avant sa conversion en gabarit, par l'accès au système par une image préalablement prélevée.
« Substitution attack » est une autre façon ou technique d'usurpation d'identité par l'insertion des caractéristiques biométriques d'un pirate ayant réussi à accéder à une banque de données aux renseignements personnels d'une autre personne.
Alors qu'un mot de passe est facilement renouvelable, la donnée biométrique deviendra caduque et ne pourra être réutilisée une fois subtilisée. En effet, les données biométriques ont la particularité d'être irrévocables et tout se complique si l'utilisateur légitime se fait pirater ses données [18].

Un système biométrique peut soumettre à d'autres types d'attaques. dont les lignes suivantes présentent les différents points d'attaques d'un système biométrique.

6.2 Modèle biométrique et problèmes de sécurité

Ratha et al. [19] ont classé les attaques sur un système biométrique générique en 8 niveaux ou classes. La figure 1.20 définit les emplacements possibles de ces attaques dans un système biométrique générique :

A. **Données biométriques falsifiées** : une reproduction de la donnée biométrique utilisée sera présentée au capteur biométrique (comme la présentation d'une copie d'une signature).

B. **Transmission de données biométriques interceptées** : une ancienne donnée biométrique enregistrée est rejouée dans le système sans passer par le capteur biométrique (comme la présentation d'une ancienne copie de l'image de l'empreinte).

C. **Attaque sur le module d'extraction des caractéristiques** : ce module pourrait être remplacé par un cheval de Troie de manière à produire des informations choisies par l'attaquant.

D. **Altération de caractéristiques extraits** : Après l'obtention de données par le module d'extraction de caractéristiques, ceux-ci sont altérés voire remplacés par d'autres données définies par l'attaquant.

E. **Module de calcul de similarité est remplacé par un module malveillant** : ce module pourrait être remplacé par un cheval de Troie afin de produire artificiellement de hauts ou bas scores.

F. **Altération de la base de données** : la base de modèles biométriques est disponible localement, à distance ou distribuée sur plusieurs serveurs. Dans ce type d'attaque, l'attaquant modifie un ou plusieurs modèles afin d'autoriser un imposteur voire d'empêcher un utilisateur légitime d'y accéder.

G. **Attaque sur le canal entre la base de données et le module de calcul de similarité** : dans ce type d'attaque, les modèles sont altérés sur le lien de transmission reliant la base de modèles et le module de calcul de similarité.

H. **Altération des décisions (acceptées ou rejetées)** : ce type d'attaque altère la décision booléenne (oui ou non) pris par le module de calcul de similarité. La dangerosité de cette attaque est élevée puisque même si le système est robuste en termes de performance, il a été rendu inutile par ce type d'attaque.

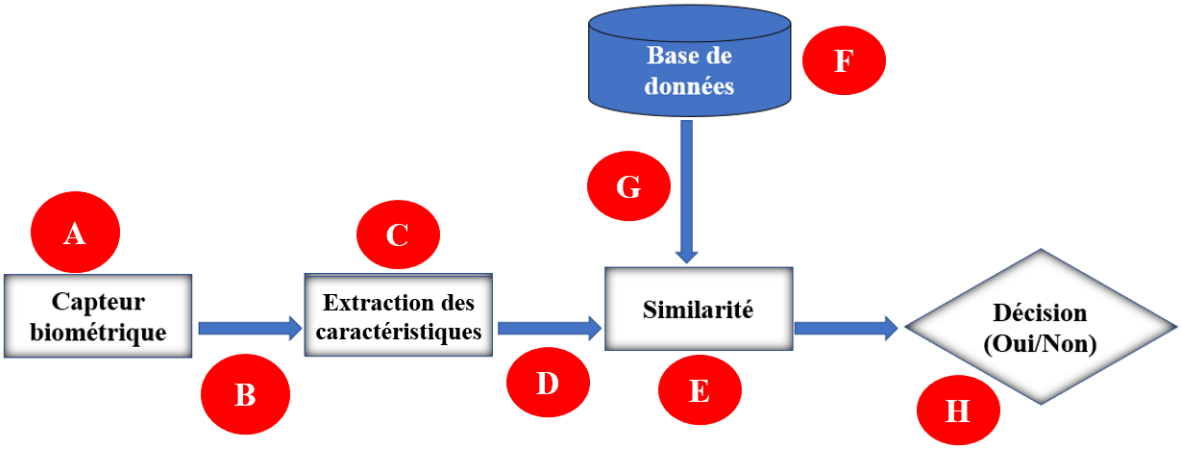

FIGURE 1.20 – Emplacements des points de compromission d'un système biométrique [19]

Les menaces relatives de ces attaques reposent généralement sur plusieurs facteurs que sont la modalité biométrique (il est plus difficile de reproduire la rétine que de forger une signature), le type du capteur (2D ou 3D, les capteurs 3D permettent de mieux détecter les tentatives de fraudes) et les paramètres de sécurité (illustrés par le FAR) du système.

Pour surmonter ces différentes attaques, il faut penser à sécuriser le système biométrique dont la section suivante présente un perçu sur les différentes méthodes de sécurisation d'un tel système.

6.3 Sécurisation du modèle biométrique

La sécurité du modèle biométrique est toujours une tâche très importante lors de la conception d'un système biométrique sécurisé. Avant de présenter les techniques utilisées pour la sécurisation des modèles biométriques, on présente d'abord des notions de base de la cryptographie qui sont utilisés pour atteindre ce but.

6.3.1 Notions générales de cryptage

La cryptographie désignant l'ensemble des techniques permettant de chiffrer des messages, c'est-à-dire permettant de les rendre inintelligibles sans une action spécifique.

A. **Cryptographie symétrique(cryptographie à clé privée)**

Le chiffrement symétrique est basé sur des fonctions mathématiques réversibles. Le chiffrement symétrique repose sur un principe de clé unique pour chiffrer et déchiffrer(comme le montre dans la figure 1.21). Le chiffrement symétrique se

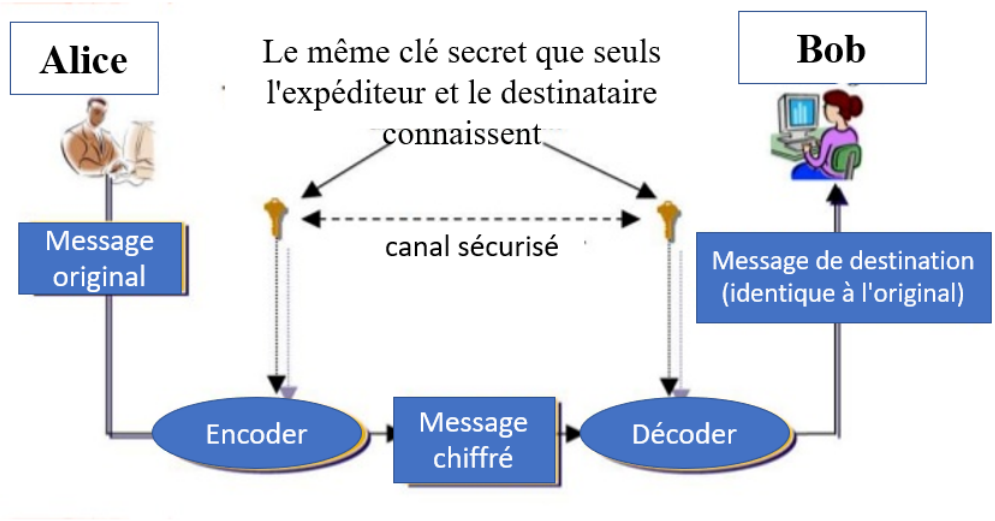

FIGURE 1.21 – La gryptographie symétrique

déroule en 3 étapes, de la manière suivante :

1ère étape

- Génération de la clé secrète par Alice
- Envoi de cette clé secrète à Bob, de manière sécurisée

2 ème étape

- Chiffrement du message original par Alice, avec la clé secrète générée
- Envoi de ce message chiffré à Bob

3 ème étape

- Réception du message chiffrée par Alice
- Déchiffrement du message avec la clé secrète reçue auparavant

B. **Cryptographie asymétrique(cryptographie à clé public)**

La cryptographie asymétrique est l'une des plus grandes fondations de la cybersécurité. Par exemple, chaque interaction sécurisée sur le Web public repose sur la cryptographie à clé publique (connexion cryptée SSL). Contrairement à la cryptographie symétrique, il utilise deux clés différentes, une clé pour chiffrer, et une autre pour le déchiffrer (ressemblent mathématiquement mais qui ne sont pas identiques).

La cryptographie asymétrique se déroule selon les étapes suivantes :

- Anis écrit un message, et souhaite l'envoyer à Mohammed, les deux possèdent une paire de clés, et chacun connait la clé publique de l'autre.
- Afin de chiffrer un message pour le destinataire, Anis va alors utiliser la clé publique du Mohammed.

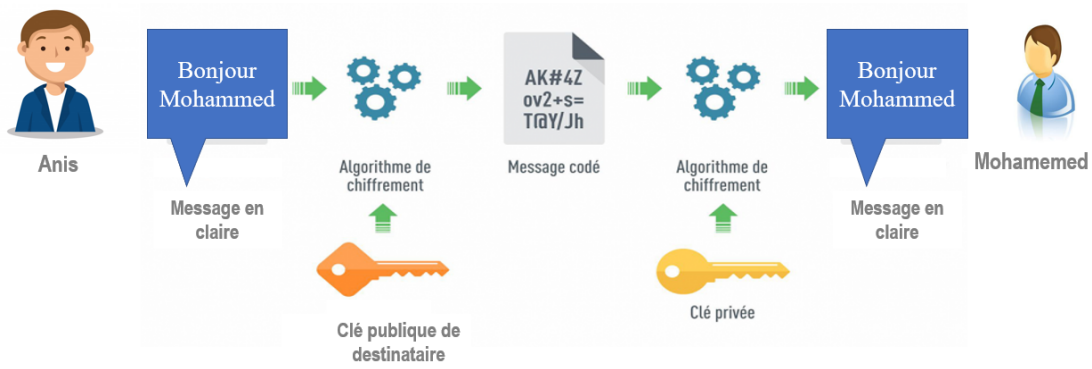

FIGURE 1.22 – La cryptographie asymétrique

• Cette clé active un algorithme, et le message écrit est alors transformé en texte incompréhensible, qui peut alors être envoyé au Mohammed.

• Lorsque Mohammed reçoit le message chiffré, il devra utiliser sa propre clé privée, celle que lui seul détient, afin d'activer l'algorithme pour le déchiffrer.

C. **Fonction de hashage cryptographique**

La fonction de hachage cryptographique est une fonction mathématique qui prend n'importe quelle chaîne d'entrée (données) de n'importe quelle longueur et génère une chaîne alphanumérique de taille fixe [20]. La chaîne de sortie est appelée valeur de hachage ou empreinte numérique ou somme de contrôle. De plus, la sortie est de longueur fixe et unique. La fonction produit toujours le même hachage à partir des mêmes données malgré le nombre de recalculs. Le hachage ne peut pas être inversé pour obtenir l'entrée données (très difficile) et, par conséquent, il peut être utilisé pour vérifier l'intégrité des données. Ainsi, il est également appelé fonction de hachage unidirectionnelle. La fonction de hachage a trois propriétés principales :

— **Résistance à la collision** : cette propriété rend très improbable (probabilité très faible) que deux entrées aléatoires génèrent le même résultat de hachage et qu'il est impossible (par calcul) de trouver un ensemble de données différent qui génère le même résultat de hachage donné d'un autre ensemble de données malgré la recalcul plusieurs fois. Plus formellement, la résistance à la collision d'une fonction de hachage peut être définie comme suit :

Il est très difficile de trouver deux entrées différents X, Y : Hash (x) = Hash (y).

— **Résistance à la pré-image** : la deuxième propriété stipule que la fonction de hachage doit être une fonction unidirectionnelle. Cette propriété implique qu'étant donné la sortie d'une fonction de hachage, il ne devrait y avoir aucun moyen de récupérer l'entrée d'origine.

— **Distribution uniforme** : La troisième propriété indique que les résultats de

hachage sont uniformément distribués dans l'espace de sortie. Étant donné une entrée aléatoire, la probabilité d'obtenir un résultat choisi est la même pour toutes les valeurs dans l'espace de sortie. Cela signifie que toutes les sorties possibles ont la même chance d'être "touchées". (plus de détails sur ce point dans le chapitre 2)

D. **Signature numérique**

La signature numérique est la méthode de cryptographie la plus sécurisée pour assurer la sécurité des informations. Pour prouver l'origine(authentification), l'intégrité des données et la non-répudiation du message, il est courant d'envoyer une signature numérique avec le message lui-même. Le processus de signature illustrer dans les étapes suivantes et la figure 1.23.

1. Calcul de l'empreinte de hachage(hash) des données à signer.
2. Chiffrement de l'empreinte à l'aide de la clé privée. On obtient alors la signature qui sera liée avec un certificat pour authentifie l'identité du signataire.
3. Déchiffrement de la signature avec la clé publique. Cela permet de retrouver l'empreinte associée aux données signées.
4. Calcul de l'empreinte des données signées. On vérifie que cette empreinte correspond à la précédente, auquel cas la signature est valide : les données sont donc intègres et l'identité de l'expéditeur est vérifiée[21].

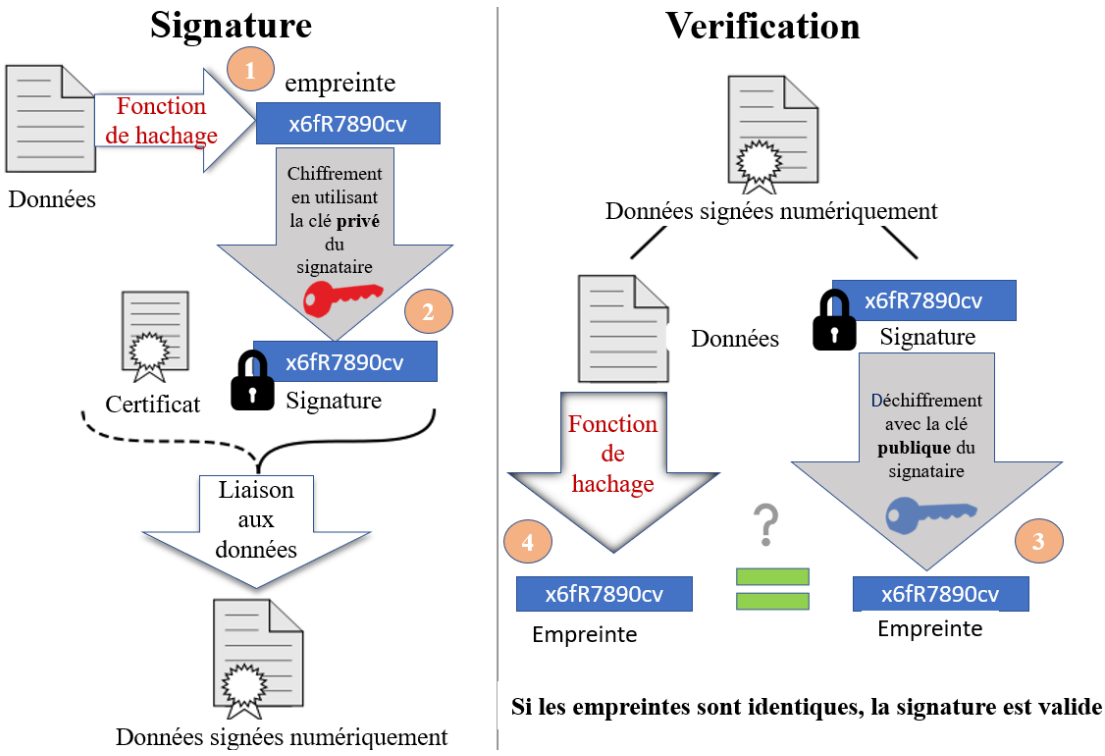

FIGURE 1.23 – Illustration de signature et vérification d'un message [22]

E. **Arbre de Merkle**

Un arbre de Merkle où arbre de hachage est une structure de données binaires arborescente qui permet de condenser un ensemble de blocs de données en un seul code de hachage au moyen d'une fonction de hachage cryptographique. Les feuilles contiennent les valeurs à stocker et les autres noeuds internes sont le hachage de ses deux fils. l'arbre de Merkle tire son nom de Ralph Merkle, qui considérer comme l'inventeur de ce type de structure en 1979 [23].

La figure 1.24 explique le fonctionnement de l'arbre de Merkle, il s'agit de hacher les blocs de données L1,..,L4 (les « feuilles »), puis de concaténer les empreintes (hashes)résultantes (Hash(L1),..,hash(L4)) deux à deux et de les hacher, et ainsi de suite jusqu'à l'obtient d'un seul hash qui s'appelle **Racine de Merkle** (Merkle root), et la modification de n'import quelle données d'un neoud entraînera la modification complète de la valeur de la racine. De cette façon, l'intégrité d'une quantité arbitraire de données peut être efficacement assurée.

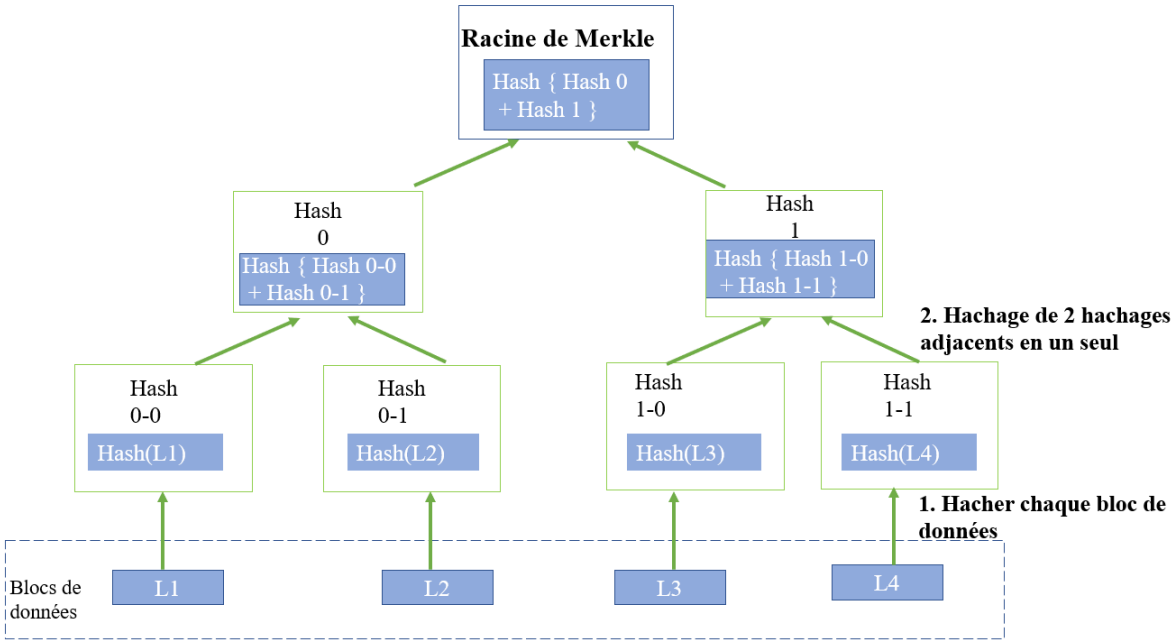

FIGURE 1.24 – Representation de l'arbre de merkle

Ces techniques de cryptage même si elles sont inventées pour sécuriser d'autres types de données, elles sont adaptées pour sécuriser les données biométriques. La section suivante présente les différentes méthodes de sécurisation des données biométriques tels que les approches matériels et les approches logiciels.

6.3.2 Approche matériel

Il s'agit d'assurer le stockage sécurisé du modèle biométrique sur un dispositif dédié (secure element) comme une carte à puce. Différentes solutions peuvent être proposées :

— Store-on-Card (SoC) : il s'agit d'éliminer la base de données centrale et de la remplacer par un dispositif sécurisé, dont en stockant le modèle biométrique sur le dispositif sécurisé[24].

— Match-on-Card (MoC) : Se réfère aux solutions où le module de comparaison est sur l'élément sécurisé. Le capteur et le module d'extraction sont sur une plateforme hôte[24].

— System-on-Device (SoD) : Le capteur, les modules d'extraction et de comparaison sont embarqués sur le même dispositif [24].

6.3.3 Approches logicielles

On peut trouver plusieurs solutions logiques, tels que :

A. **Chiffrement du modèle biométrique**

Le chiffrement des données traduit les données sous une autre forme, ou code, de sorte que seules les personnes ayant accès à une clé secrète (clé de déchiffrement) ou à un mot de passe peuvent les lire.

Le chiffrement du modèle biométrique se base sur des mécanismes de cryptographie. Le cryptage biométrique consiste à créer une clé à partir de la donnée biométrique qui servira à chiffrer et à déchiffrer un identifiant. Cette clé sera générée de manière aléatoire et différente à chaque demande d'authentification, ni la clé, ni la donnée biométrique ne sont conservées, seule la version « hachée » de la clé est conservée. Il est également impossible de relier les clés entre elles, ni de les tracer. Le cryptage biométrique permet ainsi d'utiliser la biométrie de manière anonyme et sans trace.

Il permet de réduire trois risques liés à la protection des données personnelles :

— Assurer la minimisation de la collecte de données car aucune donnée biométrique, ni gabarit ne sont conservés, cela permet de réduire les risques de perte ou de détournement de finalité.

— La personne garde le contrôle sur ses données.

— La sécurité est augmentée.

B. **Bases de données anonymes**

L'idée dans les données anonymes est de vérifier le statut d'adhésion d'un utilisateur sans connaître sa véritable identité. Une question clé dans une base de données anonyme est la nécessité d'une collaboration sécurisée entre deux parties le serveur biométrique et l'utilisateur [24]. Dans [25] où les base de données anonymes l'accès se basant sur la biométrie, l'objectif est de permettre au serveur de connaitre l'appartenance ou non du client à la base de données sans d'autres informations supplémentaires que cela soit son identité ou sa biométrie en claire (non chiffrée). Les auteurs utilisent la modalité d'iris combinée au système homomorphique de Paillier [24]. Car dans les bases de données anonymes les modèles

stockés dans la base restent en clair, alors l'information reste vulnérable aux attaques.

6.3.4 BioHachage

Basé sur la transformation du modèle biométrique à l'aide de projections pseudo-aléatoires générées à l'aide d'une clé ou d'un jeton spécifié par l'utilisateur. Cette solution a attirée beaucoup d'attention car il améliore la précision de la vérification par rapport à l'utilisation uniquement des données biométriques. Elle permet la révocation du modèle et préserve la confidentialité [26].

Toutes les méthodes de BioHashing partagent le principe commun de générer un BioCode unitaire (la donnée biométrique, après transformation) à partir de deux données : la biométrique (par exemple la texture ou les minuties pour la modalité d'empreinte digitale) et un nombre aléatoire qui doit être stocké (par exemple sur une clé USB, ou plus généralement sur une token), appelé nombre aléatoire tokenisé[24]. Le même schéma (détaillé ci-dessous) est appliqué à la fois :

— A l'étape de l'enrôlement, où seul le BioCode est stocké, au lieu des données biométriques originales brutes.

— A l'étape de la vérification, où un nouveau BioCode est généré, à partir du nombre aléatoire stocké.

Ensuite, la vérification repose sur le calcul de la distance de Hamming entre le BioCode de référence et le nouvellement émis, ce principe permet l'annulation et la diversité du BioCode en utilisant différents nombres aléatoires pour différentes applications.
Plus précisément, le processus BioHashing est illustré par la figure 1.26. On peut voir qu'il s'agit d'un schéma de protection d'authentification à deux facteurs, en ce sens que la fonction de transformation combine un nombre aléatoire spécifique dont la graine est stockée dans un jeton avec la caractéristique biométrique exprimée comme vecteur de longueur fixe F = (f1, ..., fn), F $\in$ Rn.

Ces techniques, même si elles permettent d'offrir un certain niveau de sécurisation des données biométriques, elles restent incapable d'assurer la protection nécessite à la conception d'un système biométrique sécurisé. On remarque que la possibilité d'avoir le modèle originale si un adversaire peut accéder au modèle transformé est toujours possible, ainsi qu'une fonction de hachage doit être conçus soigneusement pour que les performances de reconnaissance ne se dégradent pas.
Pour mieux répondre aux besoins de sécurisation des données biométriques, une autre technique est récemment adoptée, qui est la blockchaine.

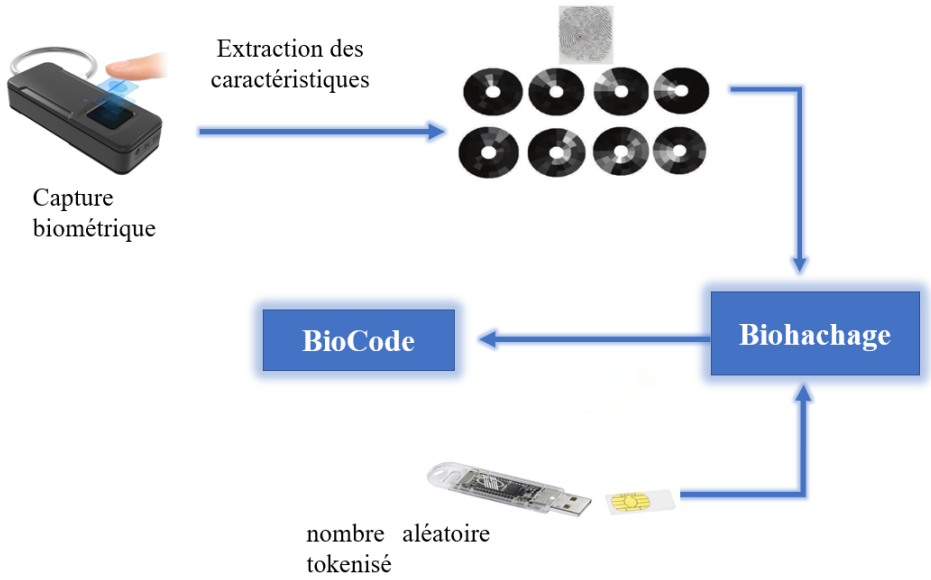

FIGURE 1.25 – BioHachage
[27]

6.3.5 Blockchain

La technologie Blockchain (chaine de blocs) est une nouvelle technologie qui intègre la décentralisation, le calcul distribué, le chiffrement asymétrique, le hachage (fonction de hachage et arbre de merkle) l'horodatage et l'algorithme de consensus. La Blockchain est une technologie qui permet de stocker et de transmettre les informations de manière sécurisée, fiable et transparente. Actuellement, son utilisation touche beaucoup de secteurs y compris la biométrie. Plus de détails sur cette technologies sont présentées dans le chapitre 2.

7 conclusion

Dans ce chapitre, on a expliqué la notion de la biométrie, les modalités biométriques, l'architecture des systèmes biométriques et modes de fonctionnements, mesure de performance d'un système biométrique, et on a terminé par la sécurité des systèmes biométriques. Cette sécurisation est reste toujours insuffisance et ne répond pas aux exigences de la protection des systèmes biométriques.

Pour répondre à ce défi, on va essayer de le combiner avec une autre technique qui s'appelle la chaine de blocs (blockchain) expliquée dans le prochain chapitre.