Anthentification transparente dans un environnement numérique ubiquitaire

Takoua Guiga

Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité INFORMATIQUE

Préparée au sein de l'Université de Caen Normandie

Anthentification transparente dans un environnement numérique ubiquitaire

Présentée et soutenue par
TAKOUA GUIGA

Thèse soutenue le 05/07/2021 devant le jury composé de		
M. AUDUN JOSANG	Professeur, Université de Oslo - Norvège	Rapporteur du jury
M. AMINE NAIT-ALI	Professeur des universités, Université Paris-Est Créteil (UPEC)	Rapporteur du jury
M. YANNICK BENEZETH	Maître de conférences, Université de Bourgogne-Franche Comté	Membre du jury
M. JEAN-JACQUES SCHWARTZMANN	Ingénieur, Orange Labs Caen	Membre du jury
MME SAMIA SAAD-BOUZEFRANE	Professeur des universités, CNAM DE PARIS	Président du jury
M. CHRISTOPHE ROSENBERGER	Professeur des universités, ENSICAEN	Directeur de thèse

Thèse dirigée par CHRISTOPHE ROSENBERGER, Groupe de recherche en informatique, image, automatique et instrumentation

UNIVERSITÉ CAEN NORMANDIE

ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

MIIS

GREYC

"One must still have chaos in oneself to be able to give birth to a dancing star."

Friedrich Nietzsche

Remerciement

Tout au long de cette thèse, j'ai pris beaucoup de plaisir à travailler sur un sujet passionnant qui m'a beaucoup apporté, aussi bien sur un plan scientifique que personnel. Je tiens à remercier les personnes qui ont contribué et soutenu ce travail. Tout d'abord, je commence par un MERCI particulier à mon directeur de thèse Christophe Rosenberger, pour son accompagnement tout au long de cette thèse. Merci de m'avoir donné une grande autonomie dans mon travail de recherche, de croire en moi et d'être disponible surtout quand les échéances se rapprochaient. Je remercie également Jean-Jacques Schwartzmann pour l'encadrement et le suivi de cette thèse, tout en veillant à ce que je garde les objectifs finaux en tête.

Je suis reconnaissante envers Orange Labs et le Laboratoire GREYC pour l'opportunité de cette thèse. Je remercie particulièrement l'équipe Security-Privacy-Innovation d'Orange Labs Caen au sein de laquelle j'ai effectué mes travaux de recherche et l'équipe SAFE du GREYC. Je suis chanceuse de faire partie de ces deux grandes équipes, riches et expertes en Cybersécurité et en Biométrie. Mes remerciements vont également à Orange Labs Rennes, là où j'ai commencé mes premiers pas en sécurité informatique.

Je tiens enfin à remercier ma famille et mes amis, qui ont su m'accompagner avec patience et encouragements le long de ce parcours. Merci infiniment pour le soutien inconditionnel et toujours plein d'enthousiasme.

Résumé

L'authentification des individus est une tâche indispensable pour contribuer à la sécurité efficace des systèmes informatiques. Les solutions innovantes foisonnent et la recherche est très active, mais peu d'acteurs s'intéressent à l'expérience utilisateur pris dans la globalité de ses interactions numériques. Beaucoup promettent de remédier au « cauchemar des mots de passe » à l'aide de moyens matériels pour renforcer la sécurité, pourtant la promesse d'un moyen d'authentification universel, sécurisé et simple est rarement tenue. Dans ce contexte de multiplicité des facteurs d'authentification, la biométrie physiologique est souvent évoquée comme alternative. Ces technologies sont toutefois très controversées, en raison notamment du risque d'atteinte à la vie privée et de la non-révocabilité de ces données. La biométrie comportementale, moins intrusive et plus simple d'emploi, peut constituer une alternative intéressante, à même de concilier les exigences apparemment contradictoires de sécurité, d'usabilité et de respect de la vie privée.

Cependant, les systèmes d'authentification comportementale actuels s'appuient principalement sur un seul objet connecté, notamment le smartphone, ce qui semble naturel puisque d'une part celui-ci est devenu le terminal de référence des utilisateurs, et que d'autre part, la multitude de capteurs dont il dispose couplée à ses possibilités de calcul, de stockage et de connectivité en font un outil de choix pour récupérer et traiter les données nécessaires à l'authentification comportementale en continu. Or, le parcours numérique d'un individu ne se limite pas aux interactions avec son smartphone. De nombreux utilisateurs disposent d'autres terminaux tels que les ordinateurs personnels et tablettes dans un cadre professionnel ou privé. De plus, grâce à l'essor des objets connectés, dont beaucoup disposent de capteurs susceptibles de récupérer des informations comportementales fortement authentifiantes, l'utilisateur se trouve immergé dans un environnement numérique ubiquitaire dans lequel la fonction d'authentification devrait venir s'intégrer naturellement.

Dans le cadre de cette thèse, nous proposons l'utilisation de la biométrie pour lier l'utilisateur avec ses objets connectés implicitement avec une solution multidevice, basée sur un cercle de confiance partagé entre les différents objets connectés permettant une authentification sécurisée de l'utilisateur et ses devices, qu'on appelle

Aura d'authentification.

Nous avons réalisé une étude de l'état de l'art sur les systèmes d'authentification et leurs exigences sécuritaires, les algorithmes de protection des données biométriques et sur les Objets connectés et l'Internet des objets IoT. Nous avons défini une méthode d'authentification transparente via un unique objet connecté, limitant les actions de l'utilisateur tout en protégeant sa vie privée. Nous avons validé cette approche sur des bases de données conséquentes en prenant en particulier le smartphone et l'ordinateur portable comme exemple d'objets intelligents. Nous proposons une approche originale d'authentification transparente via plusieurs objets connectés dans un environnement numérique ubiquitaire, qu'on appelle dans la suite Aura d'authentification. Cette approche, basée sur le transfert de confiance entre les objets connectés, s'appuie sur de la biométrie et assure la facilité d'usage et la sécurisation de l'accès de l'utilisateur à ses terminaux et services dans le respect de sa vie privée.

Abstract

User authentication is an essential task to contribute to efficient security of IT systems. Innovative solutions abound and research is very active, but few actors are interested in the user experience considering all his/her digital interactions. Many promise to remedy the "password nightmare" by using hardware to strengthen security. Yet the promise of a universal, secure and simple authentication method is rarely fulfilled. In this context of multiple authentication factors, physiological biometrics are often mentioned as an alternative. However, these technologies are very controversial, particularly because of the non-respect of privacy risks and the non-revocability of this data. Behavioral biometrics, which are less intrusive and easier to use, may be an interesting alternative, capable of reconciling the contradictory requirements of security, usability and privacy.

Meanwhile, current behavioral authentication systems rely mainly on a single connected object, particularly the smartphone, which has become the reference terminal for users, thanks to the multitude of its sensors and its storage and connectivity capabilities. However, a user's digital journey is not limited to interactions with his/her smartphone. Many users have multiple devices such as personal computers and tablets for business or private use. Moreover, thanks to the rise of connected objects, many of which have sensors that can retrieve highly authenticating behavioral information, the user is surrounded by a ubiquitous digital environment in which the authentication task should be naturally integrated.

In this thesis, we propose to use biometrics in order to connect the user with his/her connected objects implicitly with a multidevice solution, based on a circle of confidence shared between the different connected objects allowing a secure authentication of the user and his devices, which we call Authentication Aura.

We have presented a state of the art on authentication systems and their security requirements, on biometric data protection algorithms and on connected objects and the Internet of Things IoT. We have defined a transparent authentication method via a single connected object, limiting the user's actions while protecting his privacy. We have validated this approach on large databases, taking in particular the

smartphone and the laptop as examples of connected objects. We have proposed an original approach for transparent authentication via multiple connected objects in a ubiquitous digital environment, which we call Authentication Aura. This approach, based on the transfer of confidence between connected objects, relies on biometrics and ensures the ease of use and the security of the user's access to his/her connected objects and services while respecting his privacy.

Table des matières

Chapitre 1

Introduction

Nous vivons à une époque où chacun d'entre nous est occupé par beaucoup de tâches à accomplir en temps limité, et reçoit une masse considérable d'informations à chaque minute, ce qui induit un stress important à gérer. La technologie, en particulier, évolue rapidement, se complexifie et se multiplie jour après jour. Elle a connu une immense intégration dans notre vie quotidienne, créant une forte dépendance vis-à-vis des systèmes informatiques, omniprésents dans l'environnement de tout individu. Certes, le rôle important de la technologie dans notre vie moderne est évident. Elle enrichit nos vies pour communiquer et partager de façon instantanée sur de grandes distances. Cependant, sa propagation massive et continuelle engendre un excès d'informations que nous devons assumer et gérer tous les jours. Entourés par divers objets connectés (smartphones, ordinateurs, téléviseurs, montres connectées, etc.) , nous sommes inondés d'informations et nous nous trouvons dans l'obligation d'une part, et l'incapacité d'autre part, de les traiter, ce qui perturbe la concentration et crée un sentiment de frustation.

Il devient donc nécessaire de limiter nos interactions superflues avec la technologie, en commençant par réduire au moins nos habitudes quotidiennes à l'égard des équipements informatiques. Le concept du minimalisme a été adopté comme étant un nouveau mode de vie qui représente une recherche de solutions requérant le minimum d'efforts et de bouleversement. Le minimalisme est d'abord né dans la peinture aux États-Unis, au début des années 60. Ensuite, il s'est rapidement diffusé à la grande majorité des domaines artistiques : Musique, Sculpture, Architecture, Design, Mode, etc. Récemment, ce courant a intégré nos quotidiens, nos pensées, ce qu'on possède et ce qu'on fait. Qu'il s'agisse de la vie matérielle ou spirituelle, le minimalisme nous encourage à vivre légers, de nous débarrasser de tout ce qui n'est pas essentiel ou n'a pas d'utilité et de savoir nous protéger contre l'artifice. Nous proposons dans cette thèse d'appliquer cette philosophie sur le domaine de la sécurité informatique, en particulier sur le processus d'authentification auprès de nos services numériques.

Quotidiennement, la majorité des personnes sont connectées à divers appareils

électroniques. Par exemple, on commence chaque matin par consulter ses différentes messageries et réseaux sociaux sur son smartphone, lire son livre préféré sur le chemin du travail sur une liseuse. Dès son arrivée au bureau, on se trouve face à son ordinateur. À chaque connexion, on doit s'authentifier auprès des différents terminaux pour confirmer son identité en fournissant plusieurs informations (parfois les mêmes) qu'on se trouve obligé de mémoriser (mot de passe) ou de posséder (un certificat, un badge ou une clé). Ces interventions répétitives deviennent pénibles, créent du stress, une perte de temps et encombrent notre quotidien avec des tâches à faible valeur ajoutée pour l'utilisateur. L'idée de cette thèse est d'assurer une interaction suffisante entre les objets connectés d'un individu tout en créant un niveau de confiance suffisant qui puisse être transféré entre eux sans l'intervention directe de l'utilisateur. Toutefois, la présence de ce dernier reste indispensable.

L'authentification des individus, fonction critique d'une transaction électronique, est devenue une sorte de bulle technologique déconnectée d'un parcours client par ailleurs de plus en plus fluide : les solutions innovantes foisonnent [1] et la recherche est très active [2], mais peu d'acteurs s'intéressent à l'expérience utilisateur pris dans la globalité de ses interactions numériques. Beaucoup promettent de remédier au « cauchemar des mots de passe » à l'aide de moyens matériels pour renforcer la sécurité, pourtant la promesse d'un moyen d'authentification universel, sécurisé et simple est rarement tenue : la plupart du temps, un objet (badge, bracelet, calculette) vient encombrer un peu plus l'utilisateur final, assorti d'un nouveau facteur mémoriel, pour un usage finalement limité de services. Dans ce contexte de multiplicité des facteurs d'authentification, la biométrie morphologique est souvent évoquée comme une alternative. Apple a démontré son exploitabilité industrielle pour le grand public en intégrant en 2013 un capteur d'empreinte digitale sur son iPhone, supplanté en 2017 par une nouvelle technologie de reconnaissance faciale sur ses modèles haut de gamme. Par ailleurs, certains États [3] financent des projets d'infrastructure très ambitieux basés sur la biométrie pour répondre aux enjeux de sécurité régalienne. Ces technologies sont toutefois très controversées, en raison notamment du risque d'atteinte à la vie privée et de la non-révocabilité de ces données. De plus, contrairement aux mots de passe, ces systèmes fournissent des résultats probabilistes [4], et par ailleurs les possibilités d'attaques sont réelles [5]. Par rapport à ces modalités morphologiques, la biométrie comportementale, moins intrusive et plus simple d'emploi, peut constituer une alternative intéressante, à même de concilier les exigences apparemment contradictoires de sécurité, d'usabilité et de respect de la vie privée.

Les caractères comportementaux commencent à être pris en compte, et c'est ainsi que Google a initié en 2015 son projet Abacus [6] qui devait aboutir à une mise en production sous le nom de Google Trust API [7] : il s'agit d'établir un score de confiance en cumulant différents facteurs comportementaux récupérés à partir du smartphone de l'utilisateur. Orange a choisi une approche similaire et a lancé la thèse CIFRE de Julien Hatin sur le sujet en 2014 [8], dont l'un des résultats marquants est un démonstrateur d'authentification comportementale présenté fin 2016 au Salon de la Recherche d'Orange. D'autres initiatives semblables existent, comme

FIGURE 1.1: Environement ubiquitaire (© Adobe Stock).

celle de la startup SecuredTouch, partenaire de la banque israélienne Leumi Card, ou bien encore UnifyID, basé sur la reconnaissance de la démarche de l'utilisateur [9]. Toutes ces innovations annoncent clairement un changement de paradigme dans le contrôle d'accès des utilisateurs à leurs ressources. Un score de confiance global calculé à partir de leurs habitudes comportementales devant supplanter à terme les preuves d'authentification mémorielles et matérielles que les utilisateurs sont actuellement contraints de fournir avec comme corollaire une rupture dans le parcours client.

Cependant, les systèmes d'authentification comportementale actuels s'appuient principalement sur le smartphone, ce qui semble naturel puisque d'une part celui-ci est devenu le terminal de référence des utilisateurs, et que d'autre part, la multitude de capteurs dont il dispose couplée à ses possibilités de calcul, de stockage et de connectivité en font un outil de choix pour récupérer et traiter les données nécessaires à l'authentification comportementale en continu. Or, le parcours numérique d'un individu ne se limite pas aux interactions avec son smartphone. De nombreux utilisateurs disposent d'autres terminaux tels que les ordinateurs personnels et tablettes dans un cadre professionnel ou privé. De plus, grâce à l'essor des objets connectés, dont beaucoup disposent de capteurs susceptibles de récupérer des informations comportementales fortement authentifiantes, l'utilisateur se trouve immergé dans un environnement numérique ubiquitaire dans lequel la fonction d'authentification devrait venir s'intégrer naturellement.

Dans le cadre de cette thèse CIFRE proposée par Orange Labs en collaboration avec l'équipe SAFE (Security, Architecture, Forensics, biomEtrics) du laboratoire GREYC (Groupe de Recherche en Informatique, Image, Automatique et Instrumentation), nous proposons l'utilisation de la biométrie comportementale pour lier l'utilisateur avec ses objets connectés implicitement avec une solution multidevice, basée sur un cercle de confiance partagé entre les différents objets connectés per-

mettant une authentification sécurisée de l'utilisateur et ses devices, qu'on appelle par la suite Aura d'authentification. Nous présentons deux contributions principales visant à permettre l'authentification transparente de l'utilisateur dans un contexte mono-device et multi-devices.

— Nous avons défini une méthode d'authentification transparente mono-device limitant les actions de l'utilisateur tout en protégeant sa vie privée. Nous avons validé cette approche sur des bases de données conséquentes en prenant en particulier le smartphone et le PC comme exemple d'objets intelligents.

— Nous proposons une approche originale d'authentification transparente multi-devices dans un environnement numérique ubiquitaire, qu'on appelle dans la suite "Aura d'authentification". Cette approche s'appuie sur de la biométrie comportementale et assure la facilité d'usage et la sécurisation de l'accès de l'utilisateur à ses terminaux et services dans le respect de sa vie privée. La proximité des objets intelligents mesurée via leur aura est à la base de l'approche proposée.

Le manuscrit est organisé en 4 chapitres en vue d'introduire le contexte et les motivations de cette thèse et de présenter nos contributions sur ce sujet. Ensuite, un état de l'art sur les systèmes d'authentifications, les objets connectés et la protection des données est décrit, pour formuler les objectifs de la thèse par la suite. Nous présentons les travaux effectués sur l'authentification transparente mono-device et l'authentification transparente multi-devices, ainsi que les résultats et les validations, pour finir par une conclusion et des perspectives. Le manuscrit est articulé de la manière suivante :

— Le **Chapitre 2** positionne la problématique de la thèse, donne un état de l'art sur : les systèmes d'authentification et leurs exigences sécuritaires, les algorithmes de protection des données biométriques, les Objets connectés et l'Internet des objets IoT. Nous détaillons enfin les objectifs de la thèse.

— Le **Chapitre 3** étudie le concept d'authentification transparente mono-device , montre les validations sur des bases de données conséquentes, en particulier sur smartphone et sur PC. Nous présentons les méthodes de classification utilisées pour évaluer les performances du système en prenant compte la protection des données.

— Le **Chapitre 4** présente le concept d'Aura d'authentification permettant une authentification transparente multi-devices, le procédé proposé pour l'estimation de la confiance d'Aura, et la validation sur des données réelles.

— Le **Chapitre 5** conclut ce manuscrit et donne plusieurs perspectives à cette thèse.

Chapitre 2

Positionnement de la problématique

Resume: *Ce chapitre expose le rôle et la définition du principe de l'authentification. Un aperçu technique des approches d'authentification existantes est présenté. Un état de l'art sur les algorithmes de protection des données biométriques et sur l'Internet des Objets IoT est également abordé avant de définir les objectifs de cette thèse.*

2.1 Introduction

L'authentification est une tâche indispensable pour contribuer à la sécurité efficace des systèmes informatiques. Pour comprendre la nécessité de l'authentification, il est important d'établir le contexte global dans lequel elle s'inscrit. En analysant le domaine informatique, la technologie disponible et les problèmes qui se posent, on comprend pourquoi les approches d'authentification jouent un rôle si important dans la sécurisation des systèmes. Il est également utile de comprendre le fonctionnement de base des technologies d'authentification existantes, l'état actuel de leur implémentation, ainsi que leurs forces et leurs faiblesses.

L'objectif de ce chapitre est donc de présenter le principe de l'authentification et de donner une définition précise que nous allons suivre dans ce manuscrit. Nous détaillons les différents facteurs et critères d'authentification et les exigences d'un système d'authentification transparente en terme de sécurité, vie privée et usabilité. Nous présentons un état de l'art des algorithmes de protection des données biométriques et sur les objets connectés dans le contexte de l'Internet des Objets IoT. Ceci nous permettra par la suite de définir les objectifs de la thèse.

2.2 Système d'Authentification

Nous commençons dans cette partie par définir l'authentification.

2.2.1 Définition et principe

L'authentification est une procédure, par laquelle un système informatique vérifie l'identité d'une personne ou d'une machine. L'authentification machine est utilisée pour garantir la sécurité des interactions entre machines (terminaux, serveurs, routeurs, noeuds de réseaux) par leur vérification d'identité mutuelle. L'authentification utilisateur est le processus par lequel le système s'assure de l'identité revendiquée par un individu. Il s'agit d'un mécanisme qui permet d'associer une requête entrante à un ensemble de preuves d'identité. Lawrence O'Gorman [10] explique la différence entre l'authentification utilisateur et l'authentification machine sur la figure 2.1 : Alice, étant une personne physique, doit prouver son identité au site A. Dans l'authentification machine, les sites A et B peuvent s'authentifier mutuellement. Dans l'authentification utilisateur, le site A demande à Alice de lui prouver qu'elle est bien Alice et non un usurpateur.

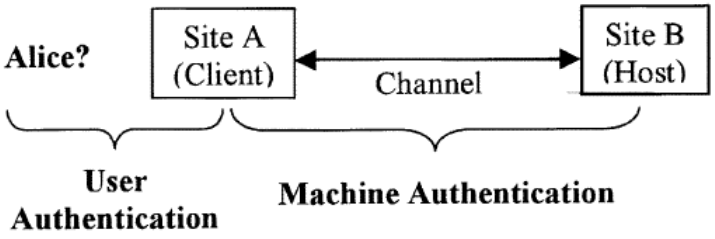

FIGURE 2.1: Authentification Utilisateur, Authentification Machine

Nous nous focalisons dans la suite de ce manuscrit sur l'authentification utilisateur. Chuang et al. [11] présentent l'authentification comme suit : *Compte tenu d'une paire (identité, échantillon), le système d'authentification doit déterminer si l'échantillon correspond d'une manière légitime à l'identité de l'utilisateur.* Hatin et al. [8] définissent l'authentification comme : *Le processus consistant a fournir des éléments en vue d'établir un certain niveau de confiance dans l'identité d'une entité. Une entité peut être une personne physique, morale ou une infrastructure ayant un rôle dans ce processus.*

Lorsqu'un utilisateur veut accéder à un système d'information, il doit dans un premier temps effectuer une procédure d'identification et d'authentification (voir Figure 2.2).

— **L'identification** consiste à établir l'identité de l'utilisateur. Elle permet de répondre à la question « qui êtes-vous ? ». L'utilisateur utilise un identifiant qui l'identifie d'une manière unique.

— **L'authentification** permet à l'utilisateur d'apporter la preuve de son identité. Elle permet de répondre à la question « Êtes-vous vraiment cette personne ? » en utilisant un élément de preuve de son identité.

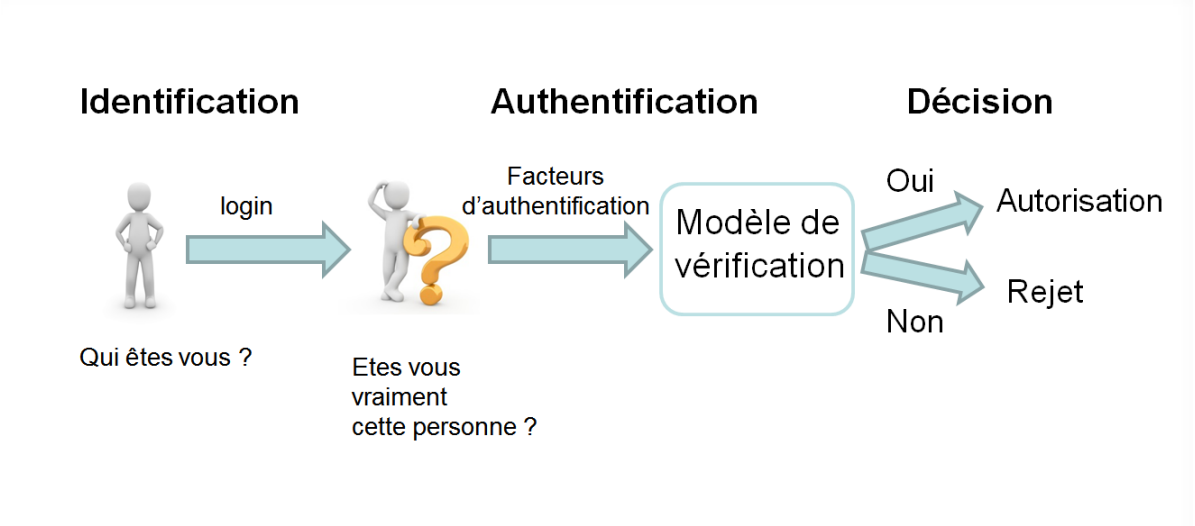

FIGURE 2.2: Procédure d'identification et d'authentification

2.2.2 Facteurs d'authentification

Un facteur d'authentification est la preuve d'identité qu'un utilisateur peut présenter à un système informatique afin d'être authentifié. Généralement, les facteurs d'authentifications sont classés en trois catégories :

— Facteur mémoriel (Ce que l'utilisateur sait)

— Facteur matériel (Ce que l'utilisateur possède)

— Facteur biométrique (morphologique (ce que l'utilisateur montre) et comportementale (ce que l'utilisateur sait faire))

Nous présentons dans la section suivante, un aperçu technique des approches d'authentification utilisées pour chaque catégorie présentées ci-dessus, afin de comprendre l'état de l'art actuel de l'authentification, de présenter l'évolution des travaux réalisés et d'établir des comparaisons sur le fonctionnement et la performance des techniques d'authentification utilisées.

Authentification basée sur un secret

Les mots de passe textuels

Les mots de passe représentent la technique la plus courante et la plus utilisée dans une procédure d'authentification. Cependant, les vulnérabilités de cette

technique traditionnelle sont bien connues. L'un des principaux problèmes est la difficulté de s'en souvenir. Des études ont montré que les utilisateurs ont tendance à choisir des mots de passe courts ou d'autres qui sont facilement mémorisables, par exemple le nom de l'animal de compagnie ou de l'artiste préféré [12].

Malheureusement, ces mots de passe peuvent être facilement devinés ou cassés. Selon une étude faite par Avast [13], 42% des français utilisent des mots de passe faibles, laissant ainsi leurs données vulnérables face aux attaques. Un mot de passe fort est en effet l'un des principes de base à respecter pour sécuriser ses données, avec au moins quatorze caractères. Par contre, ces mots de passe sont difficiles à deviner ou à casser mais sont souvent difficiles à retenir. Des études de recherche antérieures réalisées par l'Institut national de la santé et de la recherche médicale Inserm [14] ont montré que la mémoire humaine ne peut se souvenir que d'un nombre limité de mots de passe textuels. A cause de cette limitation, les utilisateurs sont susceptibles d'écrire leur mot de passe en clair sur différents supports et ont également tendance à utiliser un même mot de passe pour différents types d'applications.

Les mots de passe graphiques

Les mots de passe graphiques ont été proposés comme alternative plus sécurisée et plus facile à utiliser aux mots de passe textuels, motivés notamment par le fait que l'homme se souvient mieux des images que du texte [15]. Les images sont généralement plus faciles à retenir ou à reconnaître que le texte et sont plus difficiles à deviner ou à casser, surtout si le nombre des images proposées est suffisamment large alors les possibilités de combinaisons d'un mot de passe graphique peut dépasser celui des textuels et offrir ainsi une meilleure résistance aux attaques. En raison de ces avantages, il y a un intérêt croissant pour le mot de passe graphique avec des implémentations sur des postes de travail, sites Web, applications de connexion, distributeurs automatiques de billets et appareils mobiles. Les mots de passe graphiques sont classifiés en deux catégories [16] :

— Techniques basées sur la reconnaissance : l'utilisateur s'authentifie en identifiant et reconnaissant un ensemble d'images qui ont déjà été vues, par exemple l'approche basée sur le choix d'une série de visages "Passfaces" décrite dans la figure 2.3 où le mot de passe est composé d'une série d'images pré-sélectionnées par l'utilisateur.

— Techniques basées sur le rappel : l'utilisateur doit reproduire une actions réalisée sur une image . Par exemple, l'approche "Click-based" demande à l'utilisateur de se rappeler et sélectionner une série de points sur une image afin d'assurer son authentification, comme illustré sur la Figure 2.4.

Authentification basée sur la possession

L'authentification basée sur une possession est fondée sur un objet qu'une personne porte sur elle et qu'on l'appelle généralement Token. Il existe une grande variété de Tokens, conçus pour servir à diverses fins comme l'accès physique aux bâtiments, aux voitures et aux bureaux ou l'accès digital à des systèmes tels que les paiements électroniques, les téléphones portables et les ordinateurs. Ils peuvent être

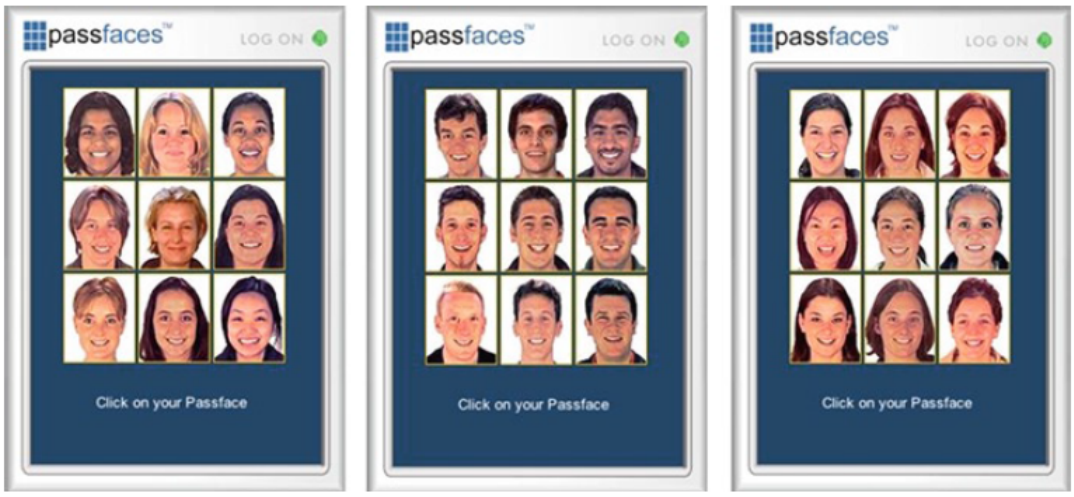

FIGURE 2.3: Authentification graphique : Passfaces [17]

FIGURE 2.4: Authentification graphique : approche "click-based".

sous différentes formes, un certificat, un badge ou une clé, et ils doivent respecter les caractéristiques suivantes :

— L'unicité : chaque token doit être unique,

— La non duplication : un token doit être difficile à dupliquer,

— La portabilité : avoir une forme convenable et facile à porter,

— Le coût : être peu coûteux et facile à remplacer en cas de perte ou de vol.

N.Clarke [18] regroupe les tokens en deux catégories, les passifs et les actifs :

— **Les Tokens passifs** : sont ceux qui enregistrent un secret qui sera présenté à un système extérieur pour le traitement et la validation afin de s'authentifier, comme les badges et les cartes de crédit magnétiques par exemple.

— **Les Tokens actifs** : ils enregistrent aussi un secret mais qui ne sera pas communiqué à un système externe. Le traitement est généré plutôt par le token lui même et le résultat de vérification sera transmis à un système externe. La clé PKI présentée dans la figure 2.5 est un exemple d'un Token actif, où l'information est combinée avec un mot de passe unique donc techniquement il s'agit plutôt d'une authentification à deux facteurs, que nous détaillerons dans la suite.

FIGURE 2.5: le token actif : clé PKI

2.2.3 Authentification biométrique

L'authentification biométrique est un processus qui valide l'identité d'un utilisateur qui souhaite se connecter à un système en mesurant certaines caractéristiques intrinsèques de cet utilisateur. Elle représente la solution préventive pour l'accès non autorisé rencontré lors d'une authentification traditionnelle par mot de passe ou par les tokens. Ainsi, plusieurs études sur les différents types de systèmes d'authentification biométrique ont été réalisées et permettent, selon les caractéristiques utilisées de la personne, de les classer en deux catégories : la biométrie morphologique et la biométrie comportementale. Nous donnons ici un rapide aperçu des modalités biométriques les plus courantes [19]. Pour les caractéristiques morphologiques, nous décrivons la reconnaissance faciale, d'empreintes digitales, de géométrie de la main, d'empreintes palmaires et l'iris.

Pour les caractéristiques comportementales [20], nous décrivons les systèmes basés sur la voix, la façon d'écrire et d'interactions avec un écran tactile, la dynamique de frappe au clavier, la démarche, le rythme cardiaque ECG et les signaux cérébraux EEG. Il existe d'autres méthodes biométriques basées sur, l′ADN, l'odeur corporelle, les veines de la main, les empreintes palmaires, la forme de l'oreille et des lèvres etc., que nous ne détaillons pas ici. Il est important de noter qu'il n'existe aucune caractéristique biométrique idéale. A chaque application correspond une ou plusieurs modalités biométriques appropriées.

Biométrie physiologique

Il s'agit d'exploiter des données relatives à différentes mesures pouvant s'effectuer sur partie du corps de la personne à authentifier/identifier. Nous traitons dans la suite des caractéristiques physiologiques utilisées dans les principaux systèmes biométriques existants.

1. **Empreinte digitale**

   L'analyse d'empreintes digitales est basée sur une étude des minuties associées [21] permettant, à partir des points caractéristiques 2.6, d'identifier ou authentifier un individu. Elles possèdent l'intéressante propriété de ne pas dépendre du patrimoine génétique et d'être invariantes dans le temps. Il existe de nombreuses méthodes d'acquisition des empreintes digitales. La plus ancienne consiste à couvrir le bout du doigt d'une fine couche d'encre et à l'imprimer sur une feuille de papier. L'empreinte ainsi imprimée peut ensuite être numérisée. Les appareils d'acquisition numériques des empreintes digitales sont basés sur la capture optique, thermique, électromagnétique ou sur les ultrasons.

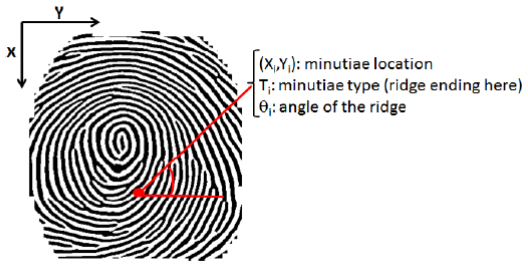

FIGURE 2.6: Schéma montrant comment sont extraites les minuties à partir d'une image [21]

2. **Reconnaissance faciale**

   Il s'agit d'extraire et d'analyser les caractéristiques faciales, depuis quelques années à partir de méthodes d'apprentissage profond [22], pour l'identification/authentification d'individus. La reconnaissance faciale est devenue une approche de plus en plus populaire utilisée dans différentes applications. C'est l'une des rares approches morphologiques qui peut être appliquée directement de manière transparente sans que l'utilisateur n'ait besoin de fournir explicitement un échantillon. Dans le cas des empreintes digitales par exemple, l'utilisateur est obligé de toucher le capteur. Néanmoins, l'authentification faciale pose des problèmes que différents travaux de recherche essayent à les résoudre. Les enjeux concernent principalement leur robustesse face à des conditions d'acquisition tels que l'éclairage, le point de vue du visage, les expressions, vieillissement, maquillage et lunettes, etc.

3. **Reconnaissance de l'iris**

   L'iris est la région annulaire située entre la pupille et le blanc de l'œil. Les motifs de l'iris se forment au cours des deux premières années de la vie et ils sont stables et uniques. La reconnaissance de l'iris est aussi considérée comme une des modalités biométriques les plus fiables. Un état de l'art des méthodes d'analyse d'iris est donné dans ce papier [23].

4. **Géométrie de la main**

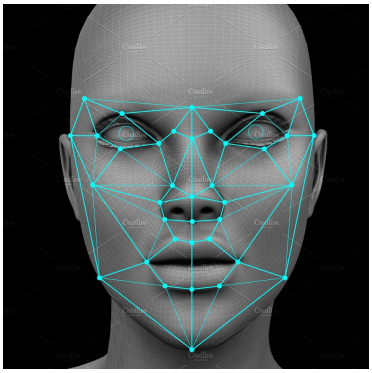

FIGURE 2.7: Reconnaissance faciale

FIGURE 2.8: Reconnaissance de l'Iris

Cette méthode consiste à déterminer les caractéristiques de la main d'un individu : sa forme, la longueur, la largeur, la courbure des doigts, etc. Les systèmes de reconnaissance de la géométrie de la main sont simples d'usage. L'utilisateur doit poser la paume de sa main sur une plaque. Une photo de la face de la main est ensuite prise par un appareil photo numérique. Une photo de profil peut aussi être prise pour obtenir de l'information sur l'épaisseur de la main. La géométrie de la main a un faible pouvoir discriminant et les systèmes peuvent être facilement trompés par de vrais jumeaux ou même par des personnes de la même famille [24].

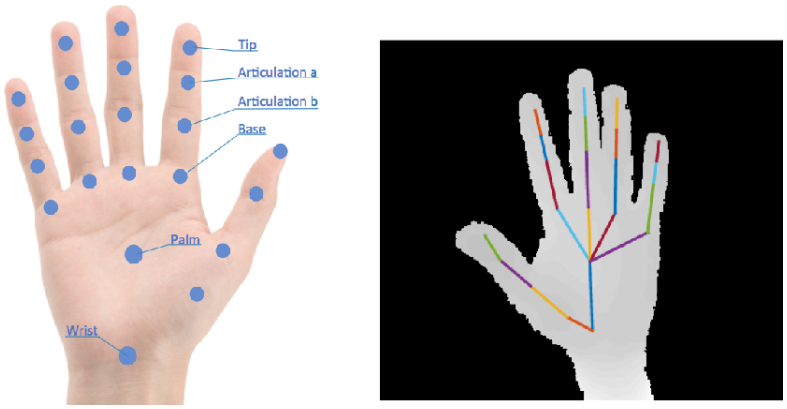

FIGURE 2.9: Reconnaissance 3D de la géométrie de la main [25]

Biométrie comportementale

La biométrie comportementale est basée sur la reconnaissance d'une personne par l'analyse de sa façon de réaliser une certaine action, plutôt que sur celle d'une caractéristique morphologique. Nous présentons dans la suite les caractéristiques comportementales utilisées dans les principaux systèmes biométriques existants.

1. **Dynamique de frappe sur clavier**

   Le processus d'authentification d'une personne à partir de la façon dont elle tape sur un clavier, connu aussi sous le nom de dynamique de frappe au clavier, est basé sur des caractéristiques uniques tels que la vitesse de frappe, le temps entre deux frappes, la fréquence des erreurs de frappe, le temps d'appui sur une touche, etc. D. Migdal [26] a utilisé cette caractéristique biométrique comportementale afin de créer un code binaire générant une signature unique basée sur l'identité de l'individu.

2. **Dynamique d'interaction tactile (Swipe)**

   Les smartphones sont maintenant équipés d'écrans tactiles. L'utilisateur interagit au travers de ces écrans tactiles avec les applications de son appareil. Il est alors possible de récupérer sans avoir à demander d'action spécifique à l'utilisateur les mouvements effectués sur les applications du téléphone mobile. Ce type d'authentification présente l'avantage d'être disponible sur l'ensemble des smartphones. Ceci permet d'avoir une modalité qui est utilisée en quasi continuellement. Différents travaux de recherche comme ceux de T.Feng et al. [27] par exemple, ont montré que les données d'interaction tactile sont discriminantes et chaque personne a sa propre façon de "swiper" sur son smartphone. La figure 2.10 montre différentes interactions tactiles de différents utilisateurs pendant la lecture d'un document sur leurs téléphones. Il est intéressant de voir que même pour une même tâche, les données tactiles de différents utilisateurs montrent des différences remarquables.

3. **La démarche**

   La démarche d'un individu peut être utilisée afin de vérifier son identité à un service numérique. L'authentification basée sur la démarche peut être réalisée par trois techniques utilisant chacune un type de capteur différent :

   — La vision par ordinateur : s'appuie sur l'utilisation de caméra pour capter et reconnaître la démarche de la personne.

   — Les "wearable" capteurs : un ensemble de capteurs portables non intrusifs peut être attaché à la personne pour identifier sa démarche comme les smartphones, les montres connectées ou les chaussures avec capteurs intégrés.

   — Les capteurs de sol : la démarche d'une personne est enregistrée et identifiée lorsqu'elle marche sur une surface surveillée par des capteurs.

   Dans leur étude, M.Derawi et P.Bours [29], ont utilisé les données issues de l'accéléromètre afin d'identifier les cycles de marches d'un utilisateur et de l'authentifier en utilisant l'algorithme de déformation temporelle dynamique.

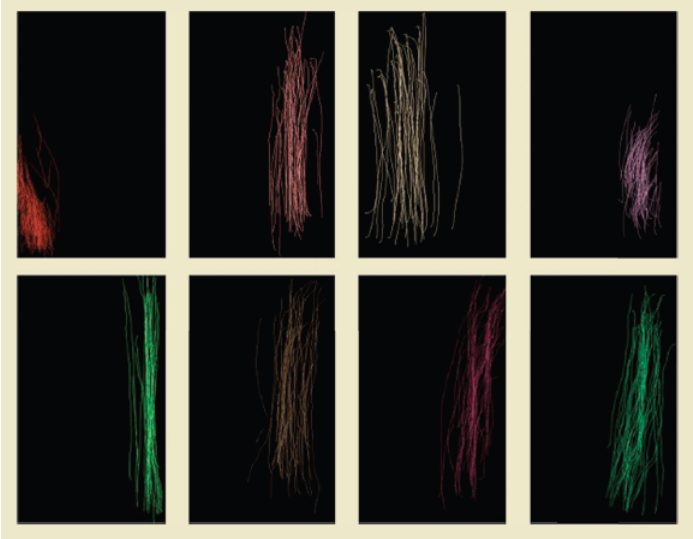

FIGURE 2.10: Interactions tactiles de différents utilisateurs pendant la lecture d'un texte sur smartphone [28]

Ils obtiennent un FRR (taux de faux rejet) de 10.7 % pour un FAR (taux de fausse acceptation) de 1.4% .

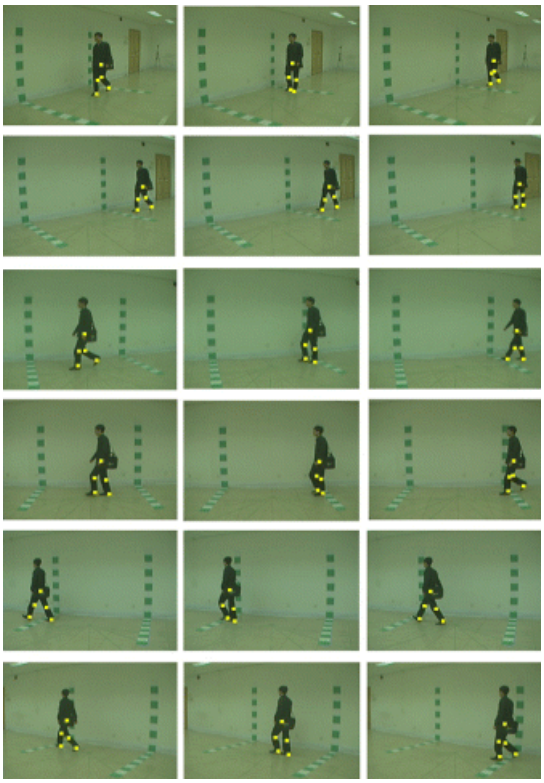

FIGURE 2.11: Authentification par reconnaissance de la démarche [30]

4. **Reconnaissance vocale**

La reconnaissance du locuteur est une modalité biométrique considérée à la fois comportementale et morphologique [31]. Lors de l'usage d'un smartphone dont la première fonction est les communications téléphoniques, il est possible de réaliser une reconnaissance du locuteur de façon transparente.

Après ce bref tour d'horizon de modalités biométriques pouvant être utilisées dans le contexte de la thèse, nous nous focalisons sur les exigences d'un système d'authentification.

2.2.4 Exigences d'un système d'authentification

Les systèmes actuels d'authentification sont évalués selon Bonneau et al. [32] autour de trois axes majeurs, qui sont l'Usabilité, le Déploiement et la Sécurite de la solution d'authentification. Un axe récurrent est celui de la protection de la vie privée. Hatin et al [8] ont choisi de se confronter aux exigences suivantes :Vie privée, Usabilité et Sécurité. Compte tenu de l'usage des systèmes d'authentification, il est important de trouver un compromis entre ces trois exigences pour assurer leur complémentarité en toute cohérence. Nous adaptons le triangle proposé par Hatin et al [8] donné dans la figure 2.12 pour représenter l'importance des ces exigences dans un système d'authentification.

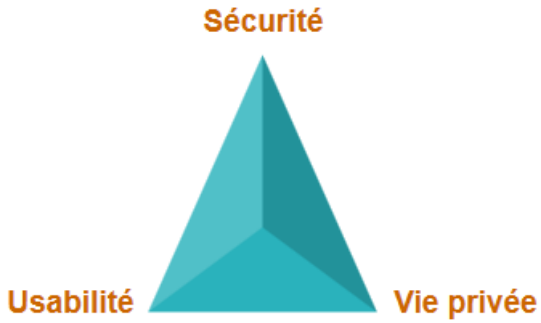

FIGURE 2.12: Représentation graphique des exigences d'un système d'authentification

Vie privée

La protection de la vie privée des utilisateurs est une notion qu'il est désormais nécessaire de prendre en compte dès la conception d'un système ("privacy by design"). La CNIL dans son guide « Gestion des risques vie privée » propose des méthodes et des recommandations pour minimiser les risques liés aux données à caractère personnel. Au niveau européen, le Règlement (UE) 2016/679 du parlement européen et du Conseil du 27 Avril 2016 [33], régit le traitement des données à

caractère personnel et la protection de la vie privée dans le secteur des télécommunications. Le concept de « privacy by design » est activement défendu sur Internet. Lors de la gestion de ces risques, les méthodes à mettre en place (chiffrement, hachage, anonymisation) sont dépendantes des données traitées. L'objectif est d'éviter la divulgation d'informations sensibles. Afin de définir les données à protéger, il est nécessaire d'évoquer la notion d'éléments d'intérêts.

Définition 1. Un élément d'intérêt est un objet qui ne doit pas être divulgué à un attaquant potentiel pour le propriétaire de cet élément. On distingue plusieurs types de données dont il faut prêter attention particulièrement au moment de l'anonymisation.

— **Les identifiants** : ce sont des attributs qui identifient directement un utilisateur : un nom, un numéro de téléphone. Ceux-ci doivent être retirés de la base.

— **Les attributs sensibles** : Ils doivent être impérativement protégés d'un attaquant. Si celui-parvient à relier un attribut sensible à l'identité réelle d'une personne, cela représenterait une atteinte à la vie privée de cette personne. Un historique d'achat sur internet ou des requêtes sur un moteur de recherche sont des exemples des données sensibles.

— **Les quasi-identifiants** : ce sont des attributs qui n'identifient pas une personne directement. A première vue, un quasi-identifiant ne parait pas être sensible, cependant, en utilisant des informations externes, ils peuvent être utilisés pour identifier un utilisateur. Un quasi-identifiant peut-être une date de naissance ou une adresse IP par exemple.

Le rapport technique de Pfitzmann et Hansen [34] établit une technologie pour décrire la protection de la vie privée par minimisation des données. Les termes décrits sont :

— **Anonymat** : Un sujet ne peut pas être identifié au sein d'un groupe par un attaquant.

— **Impossibilité d'établir un lien** : un ou plusieurs objets d'intérêt (sujet, message, actions,...) ne doivent pas permettre de relier le même individu par un attaquant.

— **Indétectabilité** : L'indétectablité d'un objet d'intérêt signifie qu'il est impossible pour un attaquant de savoir si cet objet existe.

— **Inobservabilité** : Un objet d'intérêt est inobservable s'il est à la fois anonyme et indétectable.

Les contraintes de protection de la vie privée sont définies vis-à-vis d'un attaquant, c'est-à-dire d'une personne visant à utiliser des données à caractère personnel de manière illégitime. Ces contraintes visent plus généralement les systèmes de gestion d'identités. Cependant, lors de la conception d'un système d'authentification, le stockage et l'utilisation des attributs privés peuvent amener des faiblesses et des attaques sur la vie privée des utilisateurs. Par exemple, le lien entre photographie

utilisée pour la reconnaissance faciale et la couleur de peau peut être établi. Ceci nuit à la protection des données à caractère personnel puisqu'une information à caractère racial peut être déduit de la méthode de stockage ou de transmissions de l'attribut privé. L'objectif de la protection de la vie privée et des données à caractère personnel est de garantir les libertés fondamentales de l'individu. Les données et en particulier les données biométriques sont des données qui lui appartiennent et sur lesquels il doit garder les pleins pouvoir.

Sécurité

L'authentification est la première barrière contre l'usurpation d'identité. Une fois une identité numérique usurpée, l'attaquant peut accéder à des données à caractère personnel ou professionnel. Il est aussi possible de réaliser des actions au nom de l'utilisateur. La sécurité est une question globale qui inclue aussi la gestion de ressources non numériques. Nous ne nous intéressons ici qu'aux aspects sécuritaires de l'authentification. Menezes et al. [35] définissent les attaques possibles sur un protocole d'authentification en plus de la recherche exhaustive :

— **Attaque par imitation** : Une tromperie pour laquelle une entité prétend être une autre.

— **Attaque par rejeu** : Une imitation ou une autre tromperie impliquant l'utilisation d'une information provenant d'une exécution précédente unique du protocole auprès d'un même ou d'un autre vérificateur.

— **Attaque par entrelacement** : Une imitation ou une autre tromperie impliquant une combinaison sélective d'informations, d'une ou plusieurs, précédentes ou simultanées, exécutions du protocole (sessions parallèles), incluant la possibilité qu'une ou plusieurs exécution du protocole provenant de l'adversaire lui-même.

— **Attaque par réflexion** : Une attaque par entrelacement impliquant l'envoi d'informations depuis un protocole en cours d'exécution en retour vers l'auteur de l'information.

— **Attaque par retard forcé** : Un retard forcé se produit lorsqu'un attaquant intercepte un message (qui contient un numéro de séquence par exemple) et le retransmet plus tard dans le temps. Il ne s'agit pas d'une attaque par rejeu.

— **Attaque par texte choisi** : Une attaque dans laquelle l'attaquant choisit le challenge afin d'obtenir de l'information sur le secret partagé.

En addition des attaques réalisables sur le protocole, il est aussi nécessaire de vérifier l'implémentation d'un protocole d'authentification. Ainsi, Kainda et al. [36], proposent d'évaluer un système en supposant que l'utilisateur suivra toujours le chemin de plus faible résistance. C'est à dire le chemin qui l'emmènera le plus rapidement possible à l'accomplissement de sa tâche. Ceci peut engendrer de failles dans le système d'authentification. Un exemple criant de ce type de comportement est que lorsque les mots de passe deviennent de plus en plus compliqués, les utilisateurs ont tendance à les écrire à proximité plutôt que d'essayer de les mémoriser.

Usabilité

Selon l'ISO 9241-11, l'utilisabilité ou (usabilité) est définie par le "degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié". Nielsen [37] explique l'utilisabilité comme une sous-partie de l'acceptabilité des systèmes. L'acceptabilité d'un système est la composante qui décrit si un système est suffisamment bon pour répondre à tous les besoins d'un utilisateur et des autres parties prenantes. On peut encore décomposer l'acceptabilité en deux parties qui sont :

— L'acceptabilité sociale

— L'acceptabilité pratique

FIGURE 2.13: Modèle des attributs de l'acceptabilité d'un système [37]

L'acceptabilité pratique peut encore être séparée entre l'utilité et l'utilisabilité. La figure 2.13 reprend le positionnement de l'utilisabilité. L'usabilité est traditionnellement décrite avec les cinq aspects suivants :

— **Facilité d'apprentissage** : Le système doit être facile à apprendre de manière à ce que l'utilisateur puisse rapidement réaliser des tâches avec le système.

— **Efficacité d'utilisation** : Le système doit être efficace à utiliser, après la phase d'apprentissage pour garantir un haut niveau de productivité.

— **Facilité de mémorisation** : Le système doit être facile à mémoriser de telle façon qu'un utilisateur occasionnel puisse revenir au système après une période de non utilisation sans avoir à tout réapprendre.

— **Erreurs** : Le système doit avoir un faible taux d'erreur et il doit être facile de corriger les erreurs éventuelles.

— **Satisfaction** : Le système doit être plaisant à utiliser, et les utilisateurs doivent être subjectivement satisfaits en l'utilisant.

2.2.5 Les modes d'authentification

Dans cette section, nous détaillons les différents modes de réalisation d'une authentification d'un utilisateur.

Authentification Ponctuelle

L'authentification permet d'établir un certain niveau de confiance dans l'identité d'une entité. Dans le cadre d'une authentification classique, ce niveau de confiance est établi de manière ponctuelle en début de session. Si nous représentons le niveau de confiance en fonction du temps, il est impossible de garantir qu'après l'authentification, l'utilisateur soit toujours l'utilisateur légitime une fois la preuve d'authentification fournie. Ce cas arrive lorsqu'une session est ouverte par le détenteur pour un tiers. Par exemple, il est possible pour Bob de déverrouiller son téléphone pour qu'Alice puisse appeler avec. Nous avons ainsi représenté l'apport de confiance fourni par l'authentification ponctuelle comme un trait sur la figure 2.14 . Les authentifications ponctuelles sont aujourd'hui la norme. Ce système est plus simple à implémenter et s'attend à recevoir un attribut apportant une preuve d'identité (comme un mot de passe, une donnée biométrique, ...) qui est prédéterminé. Cependant, les authentifications réalisées de manière ponctuelle ne permettent pas de s'assurer que l'utilisateur légitime est toujours le même après l'ouverture de la session. En plus, dans ce genre d'authentification, une action explicite est toujours demandée de l'utilisateur pour apporter une preuve de son identité ce qui occasionne une certaine gêne.

Authentification Continue

A l'opposé de la vérification ponctuelle, l'authentification continue vérifie l'identité de l'utilisateur sans interruption. Il n'est ainsi pas possible, avec une authentification continue, pour Bob d'ouvrir une session à Alice.

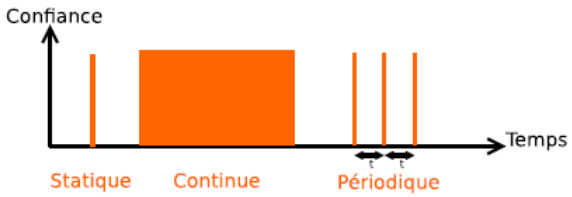

FIGURE 2.14: Apport de confiance dans l'authentification ponctuelle et continue

Les techniques d'authentification continue sont généralement utilisées en complément d'une authentification explicite. Dans l'étude comparative de Khan et al. [38], l'authentification continue repose sur le principe d'authentifier une personne en fonction de ses agissements. Il s'agit donc de l'authentifier en utilisant de la biométrie comportementale. On peut citer la probabilité d'être à un lieu donné à une heure donnée, les appels vers un numéro connu/inconnu (les habitudes d'appels), l'accès à des sites web connus / inconnus, la reconnaissance faciale et vocale...

Authentification Transparente

Afin d'assurer une authentification continue, réduire la charge sur l'utilisateur (mémoriser différents mots de passe, entrer une preuve d'identité pour une ré-authentification) et renforcer la sécurité, il faut pouvoir passer d'une authentification intrusive (qui demande une tâche explicite de l'utilisateur) à une authentification non intrusive, implicite ou ce qu'on appelle authentification transparente.

L'authentification transparente peut être réalisée par toute approche d'authentification capable d'obtenir l'échantillon requis pour la vérification de l'identité de l'utilisateur de manière non intrusive [18]. Les techniques de la biométrie comportementale sont utilisées pour assurer une authentification transparente comme par exemple la reconnaissance faciale de l'utilisateur lors d'une visioconférence ou la reconnaissance du locuteur lors d'une conversation téléphonique .

Authentification forte ou Multifacteurs

L'objectif de l'authentification est d'établir un lien de confiance entre un système numérique et l'utilisateur qui demande l'accès au système. Plus spécifiquement, l'authentification vérifie et garantit un niveau de confiance en ce qui concerne "qui l'utilisateur revendique être". Il s'ensuit que plus il y a de facteurs d'authentification présentées par l'utilisateur pour prouver son identité, plus fort sera le lien de confiance. De même, plus les facteurs sont solides, plus le lien de confiance est renforcé.

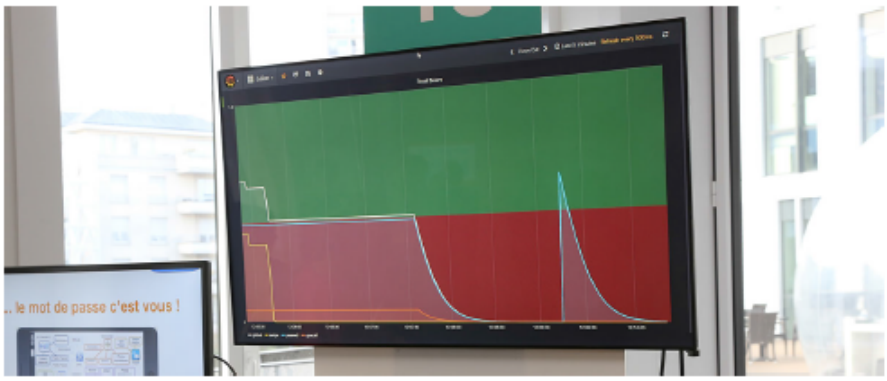

FIGURE 2.15: Photographie de la démonstration mise en place au Salon de la recherche d'Orange Labs sur l'évolution de l'indice de confiance dans une authentification transparente.

Nous pouvons donc définir l'authentification multifacteurs comme étant la procédure d'authentification qui associe au moins deux types de facteurs différents. Généralement, c'est l'association d'un mot de passe avec un token ou des caractéristiques biométriques ou les trois facteurs ensemble. Les applications basées sur l'authentification à deux facteurs sont diverses. Nous citons quelques exemples d'applications les

plus courantes pendant ces dernières années, comme par exemple, Google Authenticator, LastPass Authenticator, Authy, Yubico et Microsoft Authenticator. Nous détaillons aussi l'API d'authentification comportementale, solution d'authentification multifacteurs continue et transparente, présentée au salon de la recherche d'Orange Labs, basée sur les travaux de thèse de Julien Hatin [8] intitulée " Évaluation du niveau de confiance dans un processus d'authentification ". Cette solution offre une authentification comportementale continue en utilisant les modalités de d'authentification Swipe et Gait qui sont respectivement le mode de déplacement du doigt sur l'écran dans une application et la démarche d'une personne. Une plateforme en libre service a été proposée afin d'enrôler les utilisateurs, calculer et fournir les scores de confiance et mesurer les performances des modalités. La figure 2.15 montre une photographie de la démonstration mise en place au Salon de la recherche d'Orange Labs sur l'évolution de l'indice de confiance dans une authentification transparente.

Nous présentons aussi le principe l'OTP biométrique dans une transaction 3DSecure basée sur l'algorithme du BioHashing pour protéger les données personnelles des clients [39], comme le montre la figure 2.16. Il s'agit d'un démonstrateur (collaboration entre le GREYC et Orange Labs) d'un générateur de code biométrique à usage unique pour un cas d'usage bancaire. C'est une application Android et un service web qui simule une interaction entre un site d'e-commerce, une banque et un client. La banque, appelée OrangeBank, intègre un nouveau protocole de paiement sur internet qui permet aux clients de la banque de confirmer un paiement en renseignant leurs données biométriques. Les modalités retenues pour le démonstrateur est le chemin secret et la façon de swipe. Les coordonnées du chemin tracé ainsi que les données de temps sont récupérées pour calculer un template puis un Biocode.

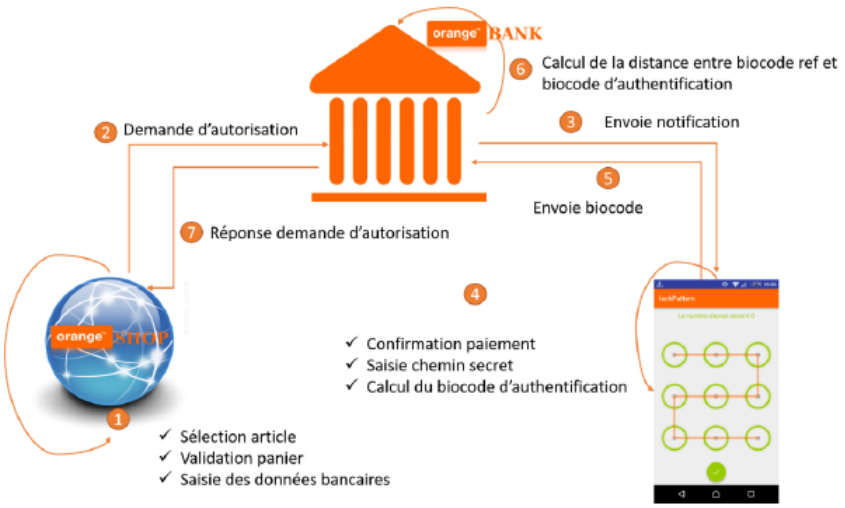

FIGURE 2.16: OTP biométrique dans une transaction 3DSecure

2.2.6 Protection des données personnelles sensibles

L'utilisation d'un grand nombre de données personnelles pose la problématique de leur protection. Différentes menaces sont possibles. La première concerne l'usurpation d'identité consistant pour un imposteur de produire une preuve illégitime d'authentification d'un individu. La combinaison d'informations à caractère personnel doit permettre de lutter contre cette menace. Or, cette collection de données personnelles ne doit pas être une atteinte à la vie privée de l'utilisateur notamment sur la conscience du contenu. A cette fin, la solution d'authentification doit permettre à la fois de prendre une décision sur l'identité de l'utilisateur sans révéler des informations personnelles utilisées. La confidentialité des informations peut être remplie par du chiffrement (classique ou plus avancé) ou par des transformations non inversibles conservant la similarité comme proposé par Ratha et al. en 2001 [40]. L'avantage de cette dernière technique est de masquer au vérifieur le contenu sémantique des informations utilisées pour générer la preuve d'authentification. La preuve d'authentification ne doit pas permettre de faire le lien entre plusieurs identités numériques du même utilisateur même si les données collectées sont les mêmes. Il est alors nécessaire de diversifier la preuve d'authentification en fonction du service, de la transaction ou du temps.

Évolution des algorithmes de protection biométriques

Nous commençons tout d'abord par présenter l'évolution des différents schémas de protection biométriques donnée par la figure 2.17.

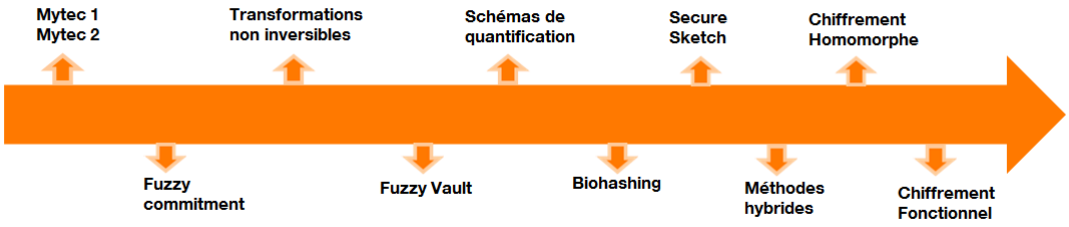

FIGURE 2.17: Évolution des algorithmes de protection de données biométriques.

En 1996, Soutar *et al.* [41] ont construit un système cryptographique appelé Mytec 1. Deux ans plus tard, ils ont présenté Mytec 2, une version développée de de Mytec 1. Juels et Wattenberg [42] ont initialisé une nouvelle méthode nommée "Fuzzy Commitment Scheme" qui associe la cryptographie aux codes de correction d'erreur pour assurer une meilleure protection des templates biométriques. Dans le début des années 2000, Ratha *et al.* [43] ont présenté les algorithmes de transformation non inversibles où les données initiales ne peuvent plus être récupérées. En 2002, le "Fuzzy Vault Scheme" a été présenté par Juels et Sudan [44]. Au cours de la même année, Feng et al. [45] ont proposé les cryptosystèmes à génération de clé en utilisant les schémas de quantification "Quantization Scheme". En 2003, Goh *et al.* [46],

ont présenté le Biohashing comme étant une nouvelle technique de transformation de caractéristiques biométriques. Cet algorithme a été proposé à l'origine pour les empreintes faciales et digitales par Teoh *et al.* in [47]. En 2004, Dodis *et al.* [48], ont développé les "Fuzzy extractors" qui ont été nommés plus tard par "Secure Sketch" par Li *et al.* [49]. Cette approche est similaire aux fonctions de hachage robustes et les transformations qui gardent la similarité. Scheirer *et al.* [50] ont développé des méthodes hybrides présentant des Biotokens révocables associant des cryptosystèmes avec des algorithmes de transformation de caractéristiques. En 2009, l'algorithme de chiffrement homomorphe "Homomorphic Encryption" a été introduit par Leo *et al.* [51]. En 2010, le chiffrement fonctionnel "Functional Encryption" a été formalisé par Boneh *et al.* [52] comme une généralisation des chiffrements à clés publiques.

Les algorithmes de protection biométriques peuvent être classés en 3 grandes catégories : les algorithmes de biométrie révocable, les cryptosytèmes biométriques et les systèmes hybrides comme le montre la figure 2.18.

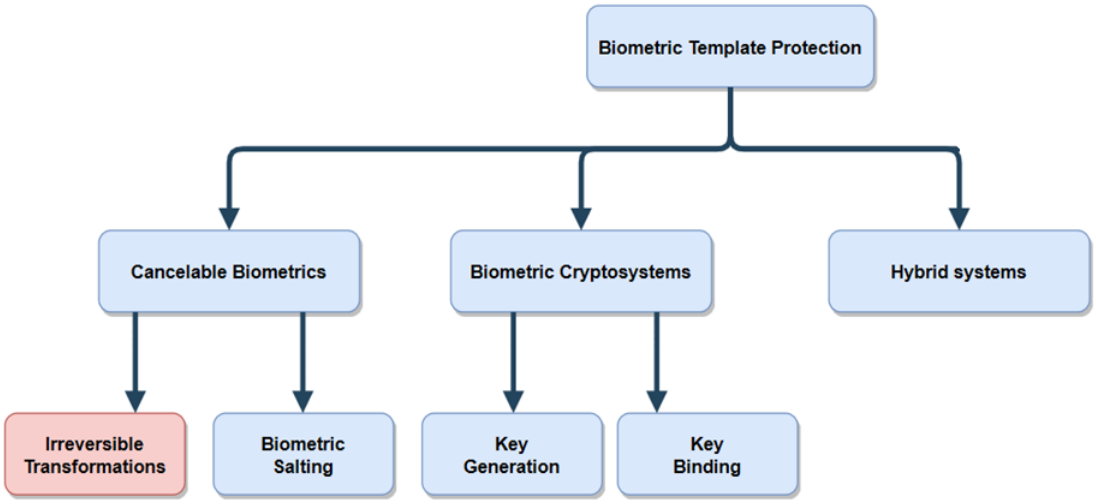

FIGURE 2.18: Les algorithmes de protection biométriques

Nous intéressons par la suite à la biométrie révocable, pour détailler notamment l'algorithme du BioHashing, basé sur des transformations irréversibles, ses différentes versions et le BioPhasor.

Biométrie révocable

La biométrie révocable est une approche consistant à transformer une donnée biométrique de façon non inversible tout en conservant la similarité. Cette transformation est généralement paramétrée par une clé permettant de révoquer le résultat de la transformée en changeant de clé. Le concept de protection de la vie privée des données biométriques a été défini en 2001 dans un document de référence [43]. Depuis lors, de nombreuses méthodes ont été proposées parmi les approches de pro-

jections aléatoires [53], les méthodes de BioHashing [47], les filtres Bloom [54], pour n'en citer que quelques-unes. Par la suite, nous nous intéressons particulièrement, à étudier et appliquer l'algorithme BioHashing sur nos données.

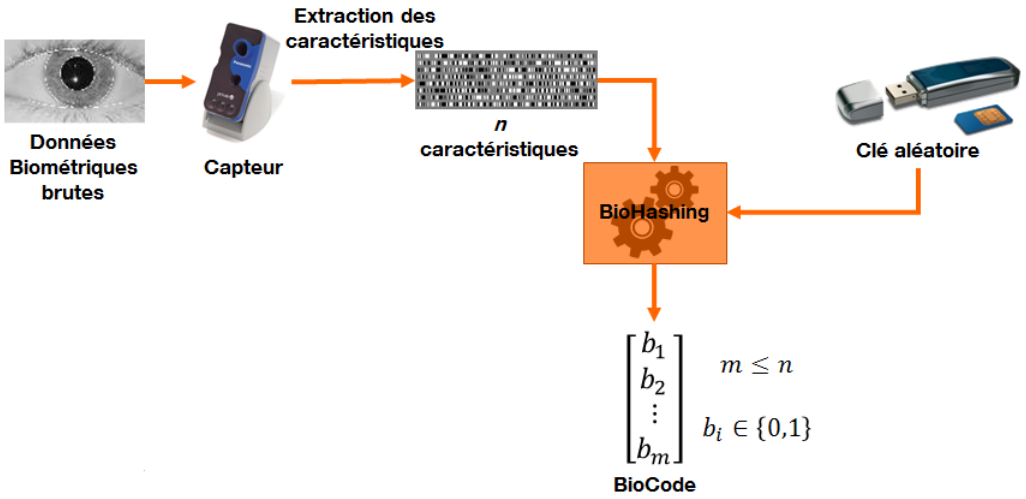

FIGURE 2.19: Schéma de biométrie révocable

Algorithme de BioHashing

Le BioHashing est un algorithme permettant de protéger des données biométriques.Il est donc appliqué sur des modèles biométriques qui sont représentés par des vecteurs de longueur fixe à valeur réelle (la métrique utilisée pour évaluer la similarité entre deux caractéristiques biométriques est donc la distance euclidienne). Il génère des modèles binaires de longueur inférieure ou égale à la longueur originale (ici, la métrique D_T utilisée pour évaluer la similarité entre deux modèles transformés est la distance de Hamming). Cet algorithme a été proposé à l'origine pour les empreintes faciales et digitales par Teoh *et al.* in [47]. Ensuite, l'algorithme BioHashing transforme le modèle biométrique $T = (T_1, \ldots T_n)$ en un modèle binaire $B = (B_1, \ldots B_m)$, avec $m \leq n$ dans l'algorithme 1. L'algorithme BioHashing est détaillé [47], l'un des schémas de protection des modèles les plus populaires.

La particularité de l'algorithme BioHashing est qu'il utilise une fonction unidirectionnelle et un germe aléatoire de m bits. Il est important de noter que chaque caractéristique biométrique enregistrée utilise un germe différent afin de créer un BioCode spécifique. La performance de cet algorithme est assurée par les produits scalaires avec les vecteurs orthonormaux. Le processus de quantification de la dernière étape assure la non invertibilité des données (même si $n = m$, car chaque coordonnée de l'entrée T est une valeur réelle, alors que les coordonnées de la sortie B sont un bit unique). Enfin, le germe aléatoire garantit à la fois les propriétés de diversité et de révocabilité.

Algorithm 1 BioHashing

1: Inputs
2: $T = (T_1, \ldots, T_n)$: biometric template,
3: K_z : secret seed
4: Output $B = (B_1, \ldots, B_m)$: BioCode
5: Generation with the seed K_z of m pseudorandom vectors $V_1, \ldots, V_m$ of length n,
6: Orthogonalize vectors with the Gram-Schmidt algorithm,
7: **for** $i = 1, \ldots, m$ **do** compute $x_i =< T, V_i >$.
8: **end for**
9: Compute BioCode :

$$B_i = \begin{cases} 0 & \text{if} \quad x_i < \tau \\ 1 & \text{if} \quad x_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

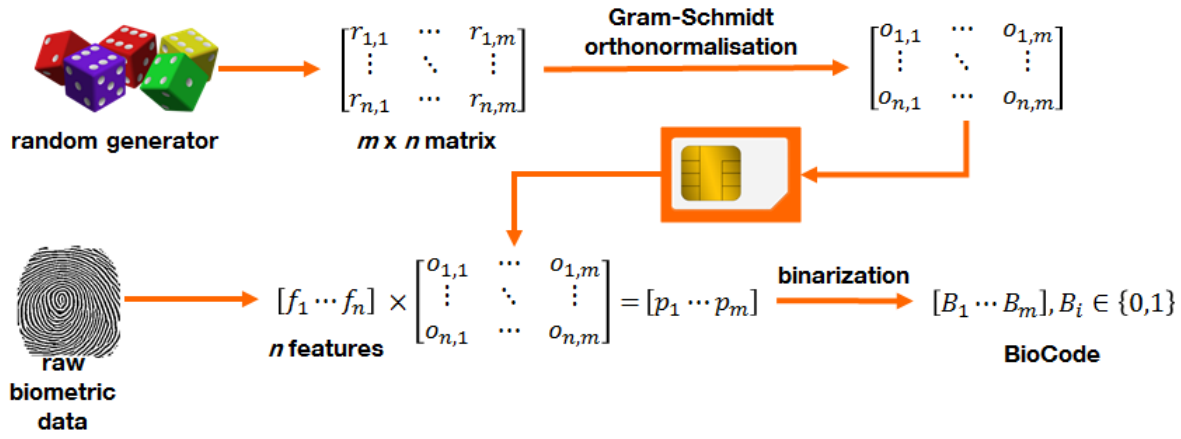

FIGURE 2.20: Schéma de BioHashing

L'algorithme BioPhasor

L'algorithme BioPhasor a été proposé par Teoh et al. dans [55] et a été présenté comme une forme de biométrie révocable basée sur le mélange itératif entre le nombre pseudo-aléatoire spécifique à l'utilisateur et la caractéristique biométrique. L'algorithme BioPhasor est considérée comme une amélioration de l'algorithme BioHashing et il est détaillé dans l'algorithme 2 ci-dessous. Le point intéressant des algorithmes de BioHashing et BioPhasor est qu'aucune phase d'apprentissage n'est requise. Néanmoins, ils sont tous les 2 dépendants de la clé secrète, ce qui crée certaines limites, évoquées dans [56] [57] [58]. Principalement, la matrice de projection est uniquement liée à la clé secrète. Si un imposteur obtient la clé secrète (pouvant être vu comme l'interception de l'attaquant d'un des deux facteurs d'authentification), l'attaque devient assez facile, notamment en la combinant avec une vraie donnée biométrique, la sienne ou celle d'un autre utilisateur. En effet, avec la

connaissance de la clé, l'imposteur utilisera la même base de projection que l'utilisateur légitime. Ainsi, il augmente la probabilité de succès de son attaque.

Algorithm 2 BioPhasor

Inputs
$T = (T_1, \ldots, T_n)$: biometric template,
K_z : secret seed
Output $B = (B_1, \ldots, B_m)$: BioCode
Generation with the seed K_z of m pseudorandom vectors $V_1, \ldots, V_m$ of length n,
Orthogonalize vectors with the Gram-Schmidt algorithm,
for $i = 1, \ldots, m$ **do** compute $h_i = 1/n \sum_{j=1}^{n} arctan(T_j^2/V_i^2)$.
end for
Compute BioCode :

$$B_i = \begin{cases} 0 & \text{if} \quad h_i < \tau \\ 1 & \text{if} \quad h_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

GREYCHashing

Dans l'intention d'améliorer la sécurité de l'algorithme BioHashing, les auteurs de [59], ont proposé l'algorithme GREYCHashing qui vise à limiter l'impact des attaques basées sur le vol du secret. Concrètement, l'algorithme de BioHashing calcule un changement de base qui est déterminé de manière unique par le germe. Si on regarde du coté sécurité, nous pouvons voir que le germe a une plus grande importance que les données biométriques. Ainsi, dans l'algorithme du GREYCHashing, il a été proposé de générer différemment la matrice de projection : elle est dérivée à la fois du germe et des données biométriques, comme décrit dans la figure 2.21. Par conséquent, la connaissance du secret est moins importante qu'avec les méthodes précédentes. De plus, une première transformation du modèle biométrique est proposée, il s'agit d'une projection sur une hypersphère de rayon R. Une description détaillée du GREYCHashing est donnée dans l'algorithme 3.

2.3 Objets connectés et IOT

L'IoT (Internet des Objets) est une conséquence de la numérisation croissante et des évolutions technologiques facilitant la mise en réseau de divers objets générant et/ou traitant des masses croissantes d'informations. Plus précisément, nous pouvons considérer l'Internet des Objets comme un environnement ultra-connecté permettant une infinité d'interactions entre des objets physiques diffus et leurs représentations virtuelles. Selon [60], l'IoT décrit "un réseau de réseaux permettant, via

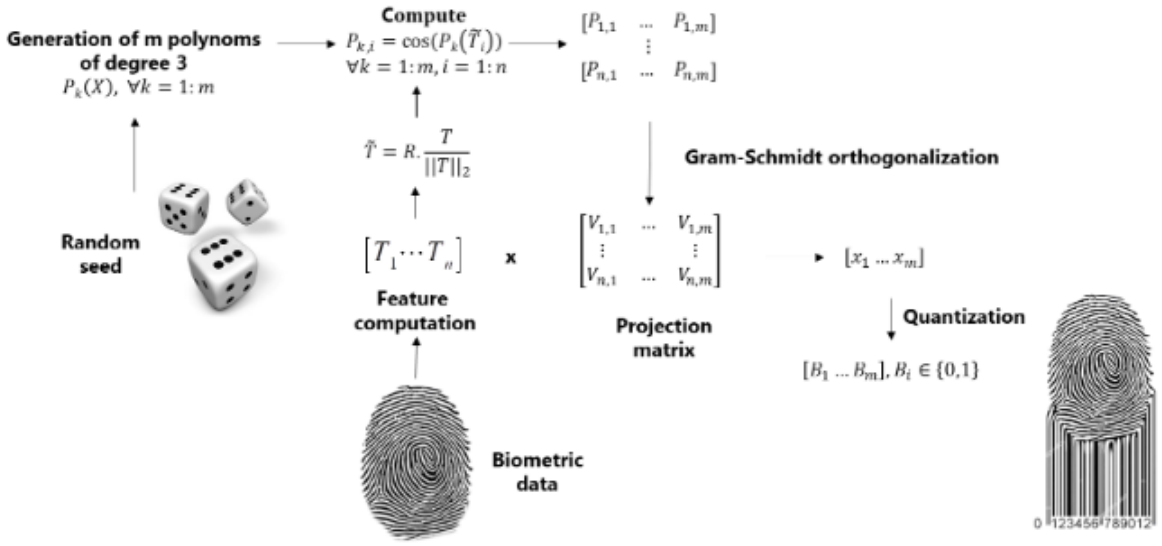

FIGURE 2.21: Principe général de l'algorithme GREYCHashing

Algorithm 3 GREYCHashing (transformation)

Inputs
$T = (T_1, \dots, T_n)$: biometric template
C : constant
K_z : secret seed
Output $B = (B_1, \dots, B_m)$: BioCode
Compute $\tilde{T} = R \times \frac{T}{||T||_2}$
Generation with the seed K_z of m polynomials P_k with $k = 1 : m$
Evaluate $P_{k,i} = cos(P_k(\tilde{T}_i)), \forall i = 1 : n, \forall k = 1 : m,$
Orthogonalize the matrix P with the Gram-Schmidt algorithm,
for $i = 1, \dots, m$ **do** compute $x_i =< T, P_i >$.
end for
Compute BioCode :

$$B_i = \begin{cases} 0 & \text{if} \quad X_i < \tau \\ 1 & \text{if} \quad X_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

des systèmes d'identification électronique et des dispositifs mobiles sans fil, d'identifier des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant".

2.3.1 Interactions entre Objets connectés

Les objets connectés existent depuis déjà plusieurs années, mais la conjonction de l'intégration technologique, de l'évolution des technologies de communication, des nouvelles capacités à traiter les données captées et l'apparition des nouveaux services

qui en découlent, a accéléré leur développement pour les rendre indispensables. Les interactions entre objets sont rendues possibles grâce à l'intégration facilitée de capteurs et contrôleurs (c'est le cas dans la plupart des systèmes complexes actuels), et du développement d'un Internet de qualité et bon marché, disponible partout. Le concept d'IoT a donc été créé pour décrire le phénomène émergent selon lequel chaque objet a la possibilité de fournir des informations sur son état présent, son passif et son devenir. Ordinateurs, téléphones, tablettes viennent évidemment à l'esprit, mais les objets considérés peuvent être de toute taille : les bus, trains et voitures, maisons, sont par exemple équipés de capteurs permettant de renvoyer des informations aux utilisateurs ou à des salles de commande plus éloignées.

Nous nous trouvons donc face à une croissance importante des objets inter-connectés qui inondent notre quotidien. Gartner [61] estime le marché mondial des objets connectés en 2020 à 69 milliards de dollars, soit une augmentation de 70% par rapport au début 2020, et prévoit une nouvelle croissance de 18% pour 2021. Un objet IoT regroupe plusieurs fonctionnalités dont certaines sont essentielles au fonctionnement du service offert. Mais que définit-on exactement par objet IoT ou objets connectés ? Quelles fonctionnalités définissent mieux un objet IOT ? Selon le rapport de recherche de GigaOM [62], cinq concepts clés sont à associer à un objet IoT : Identification, Sensibilité, Interactivité, Représentation Virtuelle et Autonomie.

— **Identification :** L'identification par type ou par entité est une notion fondamentale de l'IoT. En général, les identifiants sont numériques. Par exemple, les produits de consommation ont généralement un code barre, les livres des ISBN, etc. Des objets isolés peuvent également avoir des numéros attribués : les puces RFID stockent des codes de produits électroniques grâce à des suites de 96-bits. Les adresses IP de nos ordinateurs sont un autre exemple d'identification.

— **Sensibilité à son environnement :** S'il peut reporter son état, un objet est également susceptible de communiquer des informations sur son environnement : température, humidité, niveau de vibrations, de bruit ou géolocalisation. Si la bande passante le permet, un objet peut également enregistrer ou jouer un flux audio et vidéo.

— **Interactivité :** Les dernières avancées technologiques ont rendu possible l'interconnexion d'une grande variété d'objets et d'équipements. La plupart du temps, il n'est pas nécessaire que les objets soient connectés en permanence aux réseaux auxquels ils sont rattachés. De nombreux objets dits "passifs" tels que les puces RFID n'ont besoin d'être activés que lorsqu'ils ont besoin d'échanger des informations. Les objets "actifs" peuvent eux être connectés en permanence ou lorsqu'une connexion est disponible.

— **Représentation virtuelle :** Elle caractérise la possibilité pour un programme présent sur le cloud d'agir au nom d'un objet physique auquel il est rattaché et dont il a parfaitement connaissance. Ainsi, même un objet ne portant aucune intelligence physique peut théoriquement avoir une représentation virtuelle

complexe. Cette représentation virtuelle est quelquefois nommée cyber-objet ou agent virtuel.

— **Autonomie :** Les objets sont traités de manière individuelle, en général d'un point isolé, et opérés indépendamment d'un contrôle à distance. La notion d'apatride est ici extrêmement importante : il ne doit pas y avoir d'intelligence centrale contrôlant l'ensemble des objets individuels de manière totalitaire. Au contraire, chaque objet est en quelque sorte autonome et indépendant, avec la capacité d'être interrogé et d'interagir avec d'autres objets du réseau lorsque nécessaire.

Dans une procédure d'authentification multidevices, les interactions entre les différents devices se multiplient avec l'augmentation de nombre de devices. Dans ce cas, les attaques contre un système d'IoT sont logiquement plus susceptibles d'augmenter. En raison de confiance crée entre devices,lorsqu'un attaquant pirate l'accès à un seul device, automatiquement,il sera capable de collecter des informations de tous les autres devices. Cependant, est ce vraiment sécurisé de multiplier le nombre de devices pour une procédure d'authentification ? Afin de répondre à cette question, Godquin et al. [63] ont élaboré dans leurs travaux de recherche une taxonomie des menaces liées aux différentes fonctionnalités d'un objet IoT tout en proposant une réflexion concernant les contraintes supplémentaires s'imposant en termes de sécurité, suite à l'ajout de fonctionnalités aux solutions IoT. Nous traiterons de ces menaces dans le section suivante.

2.3.2 Menaces liées à l'IOT

La popularité de l'IoT évolue et de nombreux projets se forment autour de cette thématique. La course à l'innovation entre ces acteurs encourage une sortie de solutions toujours plus rapide. La sécurisation est perçue comme la dernière étape avant la commercialisation d'un produit. Afin de réduire les temps de conception, cette dernière est souvent négligée. Dans ce contexte, HP a mené une étude en 2014 [64] au-cours de laquelle les 10 appareils connectés les plus populaires ont été analysés. Lors de cette étude, HP a révélé que 70% des appareils évalués possédaient des vulnérabilités avec en moyenne 25 failles différentes par appareil. Une taxonomie des menaces liées aux éléments de l'IoT a été effectuée par Babar et al. en 2010 [65]. Cette taxonomie se représentée sous la forme de 5 catégories :

— **Communication** : comprenant les différentes menaces associées aux attaques sur les canaux de communication. Cela comprend notamment les attaques par déni de service (DOS), les attaques dites man-in-the-middle (MITM) ou encore les attaques par injections réseau.

— **Identification** : représente les menaces liées aux mécanismes de gestion de l'identité dont l'authentification, le contrôle d'accès ou encore le provisionnement.

— **Physique** : catégorie spécifique au domaine de l'IoT. L'objet pouvant se retrouver dans les mains d'un client malintentionné ou exposé à des tiers (envi-

ronnement hostile), la menace de l'accès physique à l'objet n'est pas à négliger. Le reverse engineering ou l'injection de fautes sont des attaques représentatives de cette menace.

— **Sécurité embarquée** : regroupe l'ensemble des menaces au niveau des couches Physical et Media Access Control (du modèle OSI) mises en œuvre dans l'objet. Cela comprend la falsification des données à ce niveau, les menaces portant sur un environnement ou un élément de sécurité, ainsi que les attaques par canaux auxiliaires.

— **Stockage** : catégorie de menaces fortement liées à la présence de l'objet dans un environnement hostile. Elle comprend les menaces liées à la gestion de clés cryptographiques et celles concernant l'intégrité et la confidentialité.

La sécurité absolue n'existe pas. A ce titre, il est impossible d'affirmer qu'un système est exempt de menaces. Il est nécessaire d'adapter le niveau de sécurité du système selon les fonctionnalités mises en œuvre et les menaces considérées. Les solutions IoT offrent une multitude de fonctionnalités via un ou plusieurs objets connectés. Chaque fonctionnalité possède des menaces associées et le cumul des fonctionnalités est sujet à une multiplication des menaces. Ci-dessous sont énumérées les principales fonctionnalités des objets connectés avec leurs menaces associées : monitoring / gestion, interface web, cloud, wireless, firmware update et applications mobiles.

1. **Monitoring / Gestion** Le monitoring est utilisé dans les objets connectés afin de faire remonter des informations auprès de leur constructeur, distributeur ou d'un autre objet. Cela peut consister à prévenir un utilisateur que son appareil n'est plus à jour, analyser sa consommation ou encore aider à la construction de statistiques pour l'évolution du produit. Cette fonctionnalité est particulièrement sensible aux menaces sur la communication et le stockage. Effectuer une attaque sur le système est possible avec un faible coût et un gain potentiel important. L'attaque de Target démontre une vulnérabilité pouvant affecter un système mettant en œuvre ce type de fonctionnalité. L'attaque de cette fonction de monitoring permet à un attaquant de bénéficier d'un accès au système pour le détourner à son profit. Les réglementations pour cette catégorie de systèmes (besoin d'une authentification double facteur) n'ont pas été respectées ouvrant ainsi l'ensemble de l'architecture réseau à cette attaque.
2. **Interface web**
   Les objets connectés bénéficient rarement d'interface utilisateur physique (écran ou clavier). Afin de combler cette absence,les constructeurs proposent communément une interface web consultable depuis un navigateur internet. Elle est principalement utilisée afin d'interagir avec l'objet (configuration et envoi de commandes). Une interface utilisateur web introduit dans le système des menaces relatives à la gestion d'identité, aux communications et au stockage. Les interfaces de ce type rencontrent les mêmes problématiques que les pages web traditionnelles ainsi que celles associées à une communication avec un serveur. Lors de l'étude effectuée par HP en 2014 [6], sur les 10 appareils les plus

populaires, 6 d'entre eux bénéficiaient d'une interface utilisateur web vulnérable. Le système s'ouvre ainsi à des attaques faiblement coûteuses telle que les injections SQL ou encore les attaques par Cross-Site Scripting. Le gain de l'attaquant est important car il bénéficie alors d'un contrôle total sur les appareils.

3. **Cloud**

   Les faibles capacités de calcul et de stockage des objets connectés impliquent fréquemment l'inter-fonctionnement de ces derniers avec un cloud. L'utilisation du cloud ouvre le système à de nouvelles menaces concernant les communications, la gestion d'identité et le stockage. D'après [64], 50% des applications mobiles IoT ne chiffrent pas leurs communications vers le cloud,internet ou le réseau local. Il est alors facile d'effectuer sur cette fonctionnalité une attaque man-in-the-middle. De plus, un système ayant absolument besoin du cloud pour fonctionner est sujet à des attaques par déni de service comme l'attaque sur Amazon Key [66]. Le coût des attaques sur un tel système est faible et offre un gain important. Les risques présentés peuvent être réduits en utilisant de la cryptographie pour protéger les canaux de communication et en prévoyant une solution locale assurant la continuité du service en cas de perte de connexion.

4. **Wireless**

   La communication sans fil est omniprésente dans les objets connectés. L'ajout de cette fonctionnalité auprès d'un système introduit des menaces sur les canaux de communications, le stockage des clés de sécurité et la gestion d'identité. Les attaques sur ces communications sont variées mais l'attaque man-in-the-middle reste la plus fréquente et facile à mettre en œuvre. En 2013, Vidgren et al. [67] proposent une attaque de ce type offrant ainsi la possibilité de récupérer la clé utilisée par le protocole radio ZigBee. Plus récemment, en 2017, Vanhoef et Piessens [68] ont proposé une attaque sur le protocole WPA2 permettant 'a l'attaquant de déchiffrer les communications. Le gain de cette catégorie d'attaques reste limité : l'attaquant peut écouter toutes les communications, mais, sans l'exploitation de vulnérabilités supplémentaires, il ne dispose pas du contrôle de l'appareil. Il est possible de prévenir ces attaques en utilisant le paramètre "High Security level" du protocole pour l'attaque sur ZigBee et en mettant à jour le système pour l'attaque sur le protocole WPA2. L'utilisation de plusieurs protocoles de communications offre potentiellement à l'attaquant autant d'accès au système, il est ainsi recommandé de limiter leur nombre.

5. **Mise à jour du firmware**

   La mise à jour du firmware est une fonctionnalité primordiale dans les objets connectés. Une fois l'objet déployé, il est important de permettre une mise à jour de l'appareil à distance. Malgré la nécessité de cette fonctionnalité, cette dernière introduit de nouvelles menaces sur les communications, la sécurité embarquée, le stockage et des menaces d'ordre physique. Il est important de toutes les prendre en compte. D'après HP [6], 60% des appareils analysés ne chiffrent pas leurs données lors du téléchargement d'une mise à jour. Il

est alors facile de mettre à jour le firmware avec une version modifiée de ce dernier en utilisant une attaque man-in-the middle. C'est le cas de l'attaque présentée par Barcena et Wueest [69]. Le gain de l'attaquant est alors fort puisqu'il bénéficie ensuite d'un contrôle total sur l'appareil. Le chiffrement des communications ainsi que la vérification des signatures complexifie la mise en place d'une telle attaque. La présence d'un secure boot peut également participer à la prévention de ce type d'attaque en empêchant le système de démarrer si l'intégrité du firmware n'est pas vérifiée.

6. **Applications mobiles**
   Dans leur papier Barcena et Wueest [69] montrent que 84% des objets connectés étudiés peuvent interagir avec une application mobile. Ces applications servent principalement à communiquer avec l'objet, le paramétrer ou le commander. L'utilisation d'une application mobile introduit des menaces de communication, de stockage, d'identité et d'autres relatives aux attaques physiques (reverse engineering ou injection de fautes). Un attaquant qui arrive à compromettre une application mobile liée à un objet connecté est en mesure de le contrôler. Il existe des attaques facilement réalisables permettant de modifier le code source de l'application comme le démontre Torano [70] lors d'une attaque sur la poupée My Friend Cayla. Afin de compliquer le reverse engineering effectué lors de cette attaque, il est possible d'utiliser des solutions d'obfuscation d'application. Il est également possible de mettre en place des procédures de vérification d'intégrité de l'application avant toute utilisation.

En se basant sur la connaissance des menaces liées aux différentes fonctionnalités des objets IoT, la conception d'un système d'authentification multidevices sécurisé peut être à priori possible. Cependant, le contrôle des fonctionnalités dans une solution IoT reste indispensable afin d'en maîtriser ses menaces. Godquin et al. [63] propose également une sécurisation progressive de l'objet en parallèle de sa conception (security by design). De la sorte, les concepteurs peuvent évaluer le coût de développement d'une solution, accroître son taux d'acceptation à grande échelle, tout en laissant la possibilité d'introduire de nouvelles fonctionnalités ultérieurement.

Dans le but de mieux répondre aux exigences d'un système d'authentification multidevices, il serait mieux de disposer d'une sécurisation adaptative servant au plus près les besoins de sécurité des objets connectés, en prenant en considération les techniques de sécurisation logicielle telles que l'obfuscation et la cryptographie dite whitebox. Une variante de ce type de méthodes pourrait être explorée en combinant les solutions logicielles avec des éléments issus de l'environnement hardware de chacun des équipements IoT.

2.4 Objectifs de la thèse

La tâche d'authentification, élément essentiel de la sécurisation de nos données et de nos terminaux, est devenue pénible et risquée alors que nous convergeons vers un monde presque totalement numérique, où les interactions sont de plus en plus fréquentes avec des appareils électroniques de plus en plus présents (véhicule connecté, habitat connecté), et des services en ligne incontournables et à forte valeur ajoutée (démarches administratives, impôts, banques et assurances, santé...). Il est donc absolument nécessaire d'utiliser des moyens d'authentification à la fois plus sûrs et plus simples d'utilisation que le vieillissant couple identifiant/mot de passe.

Cette thèse vise à améliorer les systèmes d'authentification existants, par proposer une solution d'authentification multidevices, transparente et continue, basée principalement sur des modalités biométriques comportementales et qui prend en considération l'environnement numérique ubiquitaire autour de l'utilisateur. Cette approche a pour objectifs de réduire la charge de l'utilisateur et faciliter l'usage de ses appareils électroniques, tout en assurant la sécurité et la protection de ses données personnelles.

2.5 Conclusion

Dans ce chapitre, nous avons positionné la problématique de cette thèse. Une définition précise de l'authentification a été proposé, ainsi qu'un état de l'art sur les systèmes d'authentification existants, leurs critères et exigences en terme de sécurité, usabilité et vie privée. Nous avons aussi évoqué le principe de protection de données personnelles sensibles tout en présentant un état de l'art sur les algorithmes de protection des données biométriques.

Pour permettre de proposer une alternative solide aux systèmes d'authentification existants, en tenant compte de son environnement numérique ubiquitaire, il a été judicieux de citer les objets connectés, leurs fonctionnalités et une taxonomie des menaces liées aux différentes fonctionnalités. La conception d'un système d'authentification multidevices sécurisé peut être possible. Cependant, le contrôle des fonctionnalités dans une solution IoT reste indispensable afin d'en maîtriser ses menaces.

Dans le chapitre suivant, nous allons proposer une solution générique d'authentification transparente basée sur l'utilisation d'un unique objet intelligent (ordinateur ou smartphone) respectant la vie privée de l'utilisateur.

Chapitre 3

Authentification transparente via un unique objet intelligent

Resume: *Ce chapitre présente une nouvelle méthode d'authentification transparente via un unique objet connecté. L'approche proposée est illustrée avec deux modalités biométriques différentes. La première est comportementale, elle exploite les données des habitudes d'appels sur le smartphone et seconde, les données de visage sur un ordinateur. Les résultats expérimentaux montrent l'intérêt en usage et sécurité pour l'utilisateur.*

3.1 Introduction

Les objets connectés sont devenus omniprésents dans notre vie quotidienne : les smartphones, les ordinateurs portables, tablettes, montres connectées, ... Ils stockent et génèrent une énorme quantité de données personnelles sensibles qui les rendent vulnérables aux menaces en terme de sécurité et de vie privée. La protection de ces objets connectés est devenue une nécessité absolue. En contrepartie, les méthodes d'authentification traditionnelles, qui sont principalement les codes PIN et les mots de passe, présentent des inconvénients remarquables (usage, vulnérabilité). Afin de renforcer la sécurité de ces dispositifs, l'authentification basée sur la biométrie a été adoptée comme une alternative pour assurer une meilleure protection.

Cependant, aucune méthode n'est sans limite, il a été démontré que la biométrie présente également des inconvénients et peut être attaquée [71] et est susceptible de présenter des risques pour la vie privée (traçabilité, impossibilité de révoquer une donnée biométrique). Comme indiqué dans [72], les données biométriques sont des identifiants uniques mais elles ne sont pas secrètes. Elles peuvent engendrer pour l'utilisateur des actions répétitives lors des phases d'authentification ayant parfois comme conséquence la désactivation de l'authentification. Dans le but d'améliorer les méthodes d'authentification, de nouvelles solutions ont été proposées permettant

une authentification transparente et sécurisée de l'utilisateur. L'idée principale est de mettre à jour continuellement la confiance dans l'identité d'un individu à partir d'informations collectées (géo-localisation, biométrie, habitudes, traces de navigateur,...).

Le concept d'authentification transparente a été initialement proposé par Nathan Clarke [18] comme étant l'authentification qui peut être réalisée par toute méthode d'authentification capable d'obtenir l'échantillon nécessaire à la vérification de manière non intrusive.
Les solutions d'authentification comportementale sont souvent utilisées dans des approches d'authentification transparente et représentent un domaine en pleine expansion. Cela est dû en particulier au projet d'authentification active [73]. L'Agence des projets de recherche avancée de la Défense propose d'aller au-delà du mot de passe en utilisant un mécanisme d'authentification transparent. Cela signifie que la plupart des utilisateurs s'authentifient à l'aide de capteurs biométriques. Google a annoncé en mai 2016 le projet Abacus [6], un système d'authentification multimodal, transparent et continu destiné à remplacer le couple login/mot de passe.

Avant de présenter notre contribution sur le sujet, nous réalisons un état de l'art des solutions de la littérature.

3.2 Etat de l'art

Dans le but d'assurer une authentification transparente de l'utilisateur, Clarke et al [74] ont fourni une architecture qui fonctionne sur toute la gamme des appareils mobiles, en tenant compte des différentes configurations matérielles, des capacités de traitement et des divers niveaux de connectivité au réseau. Ils ont étudié différents systèmes biométriques qui augmentent la sécurité de l'authentification au-delà des approches basées sur des mots de passe. Des méthodes basées notamment sur la biométrie comportementale comme la façon d'écrire (handwriting) [75], la façon de taper sur le clavier(Keystroke) [76] et la reconnaissance faciale [77].

Ils ont également proposé un framework basé sur la combinaison des modalités biométriques appelé NICA (Non-Intrusive Continuous Authentication) [78] pour fournir une authentification forte, transparente et continue. NICA utilise la dynamique de frappe au clavier, la reconnaissance faciale et la reconnaissance vocale pour informer le niveau d'alerte pendant que l'utilisateur interagit avec son appareil. Ce framework est basé sur la confiance d'authentification, qui est associée à chaque service afin de permettre à l'utilisateur d'accéder à un service si le niveau de confiance est supérieur au niveau d'alerte.

Plus tard, les auteurs de [79] ont montré que la combinaison de la localisation avec une authentification standard augmente la confiance globale dans cette authentification. En outre, cet article montre que les deux principaux lieux de résidence d'un utilisateur sont son domicile et son lieu de travail. Cela implique de savoir en

permanence où se trouve l'utilisateur et donc de compromettre sa vie privée. La propriété de localisation, et notamment celle offerte par les capteurs du système GPS intégrés dans les smartphones modernes, offre des caractéristiques pertinentes. Dans [80], une solution a été proposée pour authentifier les utilisateurs en utilisant la géo-localisation et les appels téléphoniques. Ils obtiennent un EER de 5,4% avec les 6 derniers appels téléphoniques. Dans [81], les auteurs combinent différentes modalités d'authentification et incluent également le contenu du SMS. Pour procéder à la lecture des informations contenues dans les SMS, les messages doivent être lus. Cela implique une fuite de confidentialité.

Des données moins sensibles peuvent être exploitées pour procéder à une authentification comportementale. C'est le cas de la reconnaissance de la démarche [82]. Cependant, les auteurs de [83] ont montré que la combinaison des informations de localisation avec la reconnaissance de la démarche augmente les performances globales du système. En combinant ces données, ils ont obtenu un EER de 10% sur un ensemble de données de 13 utilisateurs. En outre, l'authentification par dynamique de toucher sur les smartphones a récemment connu un intérêt considérable, initialement introduit par [84]. En effet, cette modalité fait appel aux mouvements de la main de l'utilisateur tout en tenant son smartphone, et à la durée de la frappe au toucher. Les auteurs proposent un système biométrique bi-modal pour l'authentification des utilisateurs basé sur des modèles de mouvements du smartphone et des modèles de texte libre à 4 chiffres de type toucher, mis en œuvre et évalué sur les systèmes Android. Afin de détailler des travaux plus connexes, le tableau3.1 présente une revue de la littérature récente sur les solutions d'authentification comportementale sur les smartphones, classées selon les types de modalités. Aucune de des solutions proposées n'a étudié la vie privée et la protection des données personnelles sensibles.

À notre connaissance, il existe peu de solutions proposées dans la littérature traitant des questions de protection de la vie privée. Les auteurs de [85] utilisent un système de chiffrement homomorphe. Dans [86], les auteurs abordent le problème de l'authentification en ligne en utilisant des informations implicites et en stockant les données directement sur le smartphone en lui déléguant ainsi le rôle de serveur d'autorisation. Cela permet d'atténuer le problème de protection de la vie privée mais ne résout pas la question de la révocabilité.

Dans [87], les auteurs ont pris en compte les questions de vie privée dans la conception d'un système d'authentification basé sur des caractéristiques biométriques. Cette solution permet de résoudre à la fois le problème du respect de la vie privée et celui de la révocabilité.

Dans ce travail, nous souhaitons améliorer l'étude de cette solution basée sur les habitudes d'appels sur smartphone, et proposer une solution basée sur la reconnaissance faciale sur ordinateur, permettant une authentification transparente, sécurisée, et en respectant la vie privée.

TABLE 3.1: Travaux connexes de l'authentification comportementale sur smartphone

Modalities	**References**	**Classification**	**Dataset (users)**	**EER(%)**
Touchstroke	Ooi and Teoh[88]	Temporal Regression Forest	Serwadda (190), Frank(41)	4, 2.5
	Soni et al.[89]	SVM, Tree Bagger,Artificial Neural Network	10	0, 40, 10.6
	Attaullah et al.[90]	BayesNET	12	18
Gait recognition	Mufandaidza et al.[91]	Dynamic Time Warping+Neural Network	N/A	22
	Sun et al.[92]	Pearson Correlation Coefficient	ZJU-GaitAcc(153)	3.1
	Bours and Denzer[93]	Euclidean Distance	13	N/A
Behavioural profiling	Nejr et al.[94]	SVM, Random Forest, gradient boosting	76	30.5, 28.7, 27
	Ashibani and Mahmoud[95]	better results with Random Forest	10	4.4
	Goni et al.[96]	logistic regression	1000	3
	Li and Bours.[97]	Random Forest	304	22.7
	Ryu et al.[98]	Dynamic Time Warping	22	6.3
Voice recognition	Lu et al.[99]	SVM	48	9.79
	Wang et al.[100]	SVM	18	5.4

3.3 Méthode proposée

Nous souhaitons développer une solution d'authentification transparente générique pouvant utiliser des données issues d'un smartphone et d'un ordinateur. Nous souhaitons remplir des enjeux d'usage (faible interaction avec l'utilisateur), de sécurité et de protection de la vie privée. Après la collecte et le traitement de données (incluant leurs protections), nous appliquons des algorithmes d'apprentissage machine pour faire évoluer la confiance dans l'identité d'un utilisateur par un tiers de confiance. Nous évaluons les performances de notre système dans 2 scénarios différents : le premier sans tenir compte de la protection de données et le deuxième étudie la performance des données avec protection. Nous étudions par la suite l'évolution de la confiance de l'usage du smartphone et du laptop dans un contexte d'authentification transparente. La figure 3.1 présente la chaîne de traitement adoptée dans les 2 cas, sans et avec protection de données.

3.3.1 Collecte de données

La collecte de données est une étape importante et nécessaire dans un système d'authentification. Aujourd'hui, nous sommes entourés par une énorme quantité de données de types différents, et l'accès à ces données est devenu de plus en plus facile et disponible. Cependant, il est nécessaire d'adapter les données collectées au besoin de chaque système, pour mieux alimenter les modèles d'apprentissage et mieux analyser les résultats. Pour ce fait, nous avons choisi de classifier les données qu'il est possible de collecter en 3 catégories : les données de type image, les données de type série temporelle et celles de type texte, comme le montre la figure 3.2. Nous donnons dans ce paragraphe, des exemples de données de chaque type, qui peuvent être collectées sur smartphone et sur ordinateur.

Image

Les données biométriques représentées sous forme d'images sont nombreuses. Nous citons notamment la reconnaissance faciale basée sur l'extraction et l'analyse des caractéristiques faciales à partir des images en adoptant des méthodes d'apprentissage profond [22]. Les empreintes digitales sont aussi collectées sous formes d'image contenant les minuties permettant d'identifier la personne [21]. Considérons aussi les images de la palme de la main qui sont utilisées dans des systèmes d'authentification basés sur la reconnaissance de la forme de la main [24]. Nous pouvons également citer la reconnaissance de l'iris qui est considérée comme une des modalités biométriques les plus fiables et qui adopte des images de l'iris pour authentifier correctement les utilisateurs. Un état de l'art des méthodes d'analyse d'iris est donné dans ce papier [23].

Série temporelle

Les données de type série temporelle sont beaucoup utilisées dans les systèmes d'authentification biométrique comportementale afin d'analyser l'évolution du com-

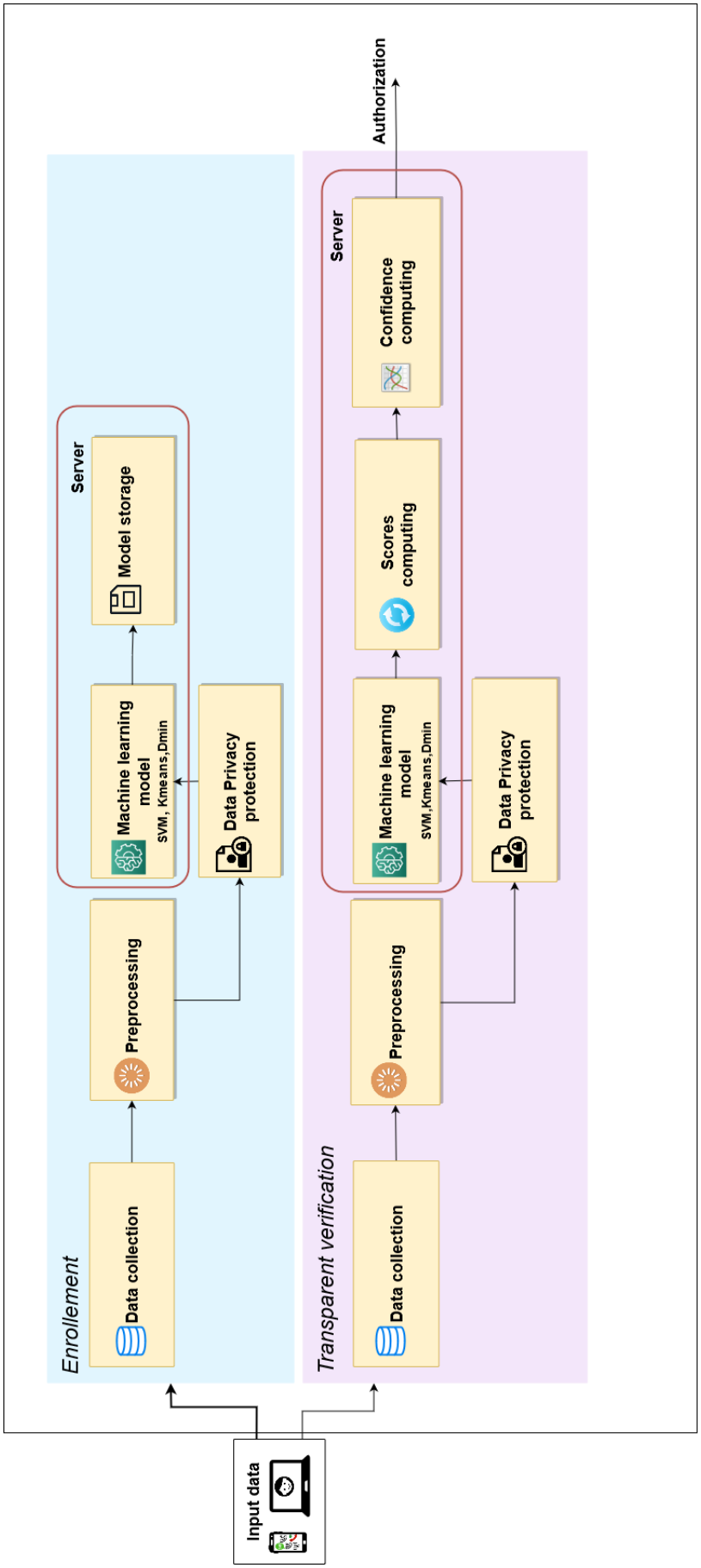

FIGURE 3.1: Chaîne de traitement de la solution proposée

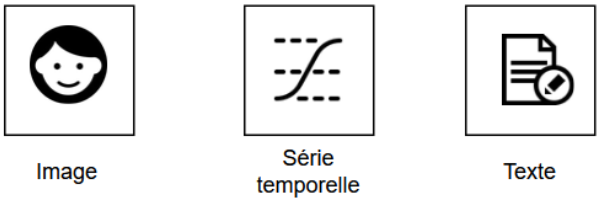

FIGURE 3.2: Les types de données à utiliser

portement au cours du temps. À titre d'exemple, nous référons à la dynamique de frappe sur clavier, basée sur des caractéristiques uniques tels que la vitesse de frappe, le temps entre deux frappes, la fréquence des erreurs de frappe, le temps d'appui sur une touche [26]. Nous pouvons également citer les habitudes d'appels téléphoniques, la dynamique d'interaction tactile (le swipe) [27] et la reconnaissance vocale [31] (voir chapitre 2).

Données textuelles

Les données textuelles peuvent en général être définies comme tous les documents écrits et conversations transcrites. Les textes peuvent être représentés par des mots individuels, des phrases entières, des chiffres, des caractères, etc. Différentes analyses du texte sont possibles [101]. Nous citons par exemple, les habitudes d'appels téléphoniques qui contiennent des données textuelles comme le numéro de téléphone, le nombre de SMS envoyés, le contenu d'un SMS, la durée d'un appel, etc. Aussi, les adresses IP, les coordonnées GPS, la liste de WIFI, les empreintes de navigateur web, etc.

Dans la suite de ce travail, nous allons étudier les données des habitudes d'appel issues d'un smartphone et les données de type Image pour la reconnaissance faciale sur Ordinateur. Dans la prochaine section, nous décrivons les pré-traitements des données collectées.

3.3.2 Pré-traitement

Le pré-traitement de données est une étape importante afin de préparer et analyser correctement les données brutes et les rendre utiles par le modèle d'apprentissage. Les techniques de pré-traitement sont diverses et choisies selon le type de données. La figure 3.3 montre un exemple de pré-traitements appliqués sur chaque type de données pour construire en sortie un vecteur de donnée qui alimente par la suite le modèle d'apprentissage.
Quand la donnée d'entrée est une image, les pré-traitements utilisés généralement sont de type filtrage et segmentation, pour faciliter la détection et l'extraction des caractéristiques spécifique à chaque image. Lorsqu'il s'agit d'une image de visage, beaucoup d'informations sont extraites comme l'âge, le genre, l'état émotionnel,etc...

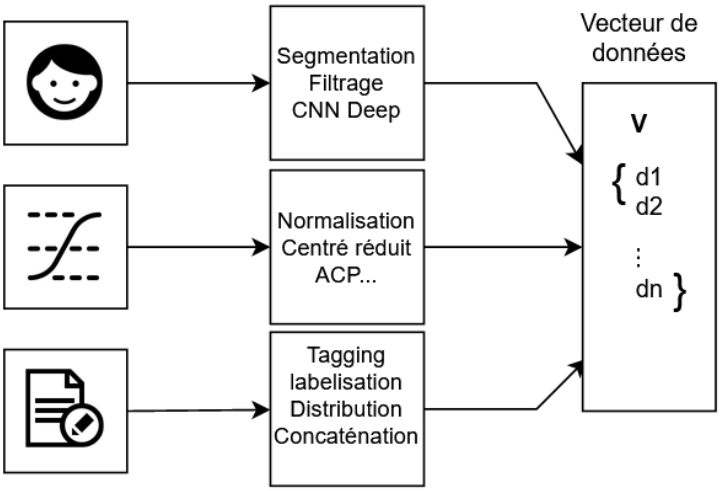

FIGURE 3.3: Les techniques de pré-traitement pour chaque type de de données.

Les réseaux de neurones convolutifs CNN ont connu un succès considérable dans la reconnaissance des visages et l'extraction des caractéristiques. Ils sont des modèles de programmation puissants permettant notamment la reconnaissance d'images en attribuant automatiquement à chaque image fournie en entrée, une étiquette correspondant à sa classe d'appartenance. En plus de la partie classification, l'architecture du CNN dispose en amont d'une partie convolutive qui a pour objectif d'extraire des caractéristiques propres à chaque image en les compressant de façon à réduire leur taille initiale. En résumé, l'image fournie en entrée passe à travers une succession de filtres, créant par la même occasion de nouvelles images appelées cartes de convolutions. Enfin, les cartes de convolution obtenues sont concaténées dans un vecteur de caractéristiques appelé code CNN.

En ce qui concerne le type de données série temporelle, les pré-traitements sont divers, notamment la normalisation des données, les réduire et les centrer, l'analyse en composante principale ACP, etc... Pour les données textuelles, la labelisation et l'étiquetage (tagging) des données est obligatoire pour faciliter le traitement, la numérisation des valeurs, la distribution de la taille des données, la concaténation des données , etc...

3.3.3 Protection des données

Comme le processus de vérification pourrait être effectué par un serveur considéré comme honnête mais curieux, une protection de la vie privée des données collectées est nécessaire. Nous avons appliqué l'algorithme de Biohashing décrit en détail dans la section 2.2.6 du chapitre 2. La clé secrète utilisée peut être un aléa stocké sur l'objet intelligent (constitue un facteur d'authentification supplémentaire).

3.3.4 Modèle d'apprentissage

Afin d'établir un système d'authentification capable de prendre une décision correcte sur l'identité de l'utilisateur, nous définissons le modèle d'apprentissage comme suit. Soit l'ensemble de données (x_i,u_i), i=1 :n avec u_i le label de la classe définie (identité de l'individu), et n le nombre des échantillons. Soit la fonction d'apprentissage définie par :

$$\mathbf{A} = \{(x_i, u_i), i = 1 : n\} \tag{3.1}$$

Il faut trouver la sortie du système y_i telle que :

$$\mathbf{f}(A) = y_i \tag{3.2}$$

avec **f** étant la fonction de classification (le classifieur) à déterminer. Nous utilisons 3 classifieurs classiques pour mesure un score d'authentification : la minimisation de distance, K-means et SVM, que nous détaillerons par la suite.

Minimisation De Distance

Il s'agit d'une technique simple consistant à affecter l'individu inconnu à la classe ayant les individus les plus semblables. En considérant un ensemble de données (x_i,u_i), i=1 :n avec u_i ayant K valeurs différentes (K classes). Un individu inconnu est affecté à la classe J si la distance entre ses paramètres et un individu appartenant à la classe J est minimale parmi tous les individus. Plus la distance est faible, plus on est sûr qu'il s'agit d'une tentative légitime.

K-means

C'est l'un des algorithmes non supervisés les plus populaires. En général, les algorithmes non supervisés prennent des décisions à partir d'ensembles de données en utilisant uniquement des vecteurs d'entrée sans impliquer de classes connues ou étiquetées. L'algorithme K-means génère k centroïdes utilisés comme modèle de comportement. Un comportement inconnu calculé à partir de données est affecté à l'utilisateur légitime s'il est proche d'un centroïde du modèle de l'utilisateur. Basé sur [102], dans le problème de regroupement(clustering), un ensemble de formation x^1,...,x^m est donné, et le but est de regrouper les données en quelques clusters cohérents. Principalement, pour un vecteur de caractéristiques donné pour chaque point de données $x^i \in \mathrm{R}_n$, l'intention est de prédire k centroïdes et une étiquette C^i pour chaque point de données. L'algorithme est le suivant :

— Initialiser les centroïdes des clusters u_1,u_2,...,$u_k \in \mathrm{R}^n$ aléatoirement

— Répéter jusqu'à la convergence :
Pour chaque i, définir

$$C^i = \arg\min_j(||x^i - u_j||^2) \tag{3.3}$$

Pour chaque j, définir

$$u_j = \frac{\sum_{i=1}^{m} 1\{C^i = j\}x^i}{\sum_{i=1}^{m} 1\{C^i = j\}} \tag{3.4}$$

Support Vector Machine SVM

L'algorithme de classification SVM est l'un des meilleurs algorithmes d'apprentissage supervisé et il est largement utilisé dans de nombreux types d'applications [103]. Dans notre étude, nous avons utilisé un SVM à classe unique avec un noyau de type RBF. Étant donné un ensemble de données d'apprentissage (x_i, y_i) pour i=1..n (n = la taille des données), avec $x_i \in \mathrm{R}^d$ et $y_i \in \{-1, 1\}$, apprendre à un classifieur à trouver : $\mathbf{f}(x_i) = \begin{cases} < 0 & \text{if } y_i = \text{-1} \\ \geq 0 & \text{if } y_i = 1 \end{cases}$ (3.5)La fonction de décision f(x_i) dépend du fait que les données sont linéairement séparables ou non. Lorsque les données sont linéairement séparables, la fonction de décision est la suivante :

$$\mathbf{f}(x_i) = \sum_{i=1}^{n} w_i x_i + b \tag{3.6}$$

où $w_i \in \mathrm{R}^n$ est le vecteur de poids, b le biais, et x_i la variable de données.
Dans la plupart des problèmes de classification, les données ne sont pas linéairement séparables. La solution pour classifier ces données est donc de les projeter dans un espace plus grand où les données deviennent linéairement séparables à l'aide d'une fonction appelée noyau (ou Kernel) K et la fonction de décision devient :

$$\mathbf{f}(x_i) = \sum_{i=1}^{n} w_i K(x_i, x) + b \tag{3.7}$$

Le noyau dépend du nombre de données et de la complexité du problème présenté et il peut être linéaire, sigmoïde, polynomial ou RBF comme indiqué dans le tableau suivant 3.2 :

TABLE 3.2: Les types de noyaux SVM

Type du noyau	Formule de calcul
linéaire	K(x,y)= x.y
sigmoïde	K(x,y)= $\tanh(\sigma(x.y) + b)$
polynomial	K(x,y)= $(\sigma(x.y) + b)^d$
RBF	K(x,y)= $\exp(-\sigma\|\|x - y\|\|^2 + C)$

3.3.5 Evolution de la confiance

A chaque intervalle de temps, des données collectées sur l'individu permettent de générer un score d'authentification (basée sur une distance). Dans un premier temps, nous traduisons ce score par une confiance ponctuelle d'authentification par la formule suivante :

$$C = e^{-\frac{(score/\sigma)^2}{2}} \tag{3.8}$$

La valeur C traduit la confiance dans la vérification d'identité ponctuelle de l'individu à l'aide d'une loi normale centrée paramétrée avec l'écart-type σ. L'utilisation

d'une loi normale pour l'estimation de la confiance permet de paramétrer simplement la probabilité que l'utilisateur de l'objet intelligent soit la personne légitime via un unique paramètre σ.

La confiance ponctuelle va nous permettre de faire évoluer la confiance globale dans l'identité de l'individu. Cette tâche est réalisée par le serveur de confiance. Lors de l'usage de l'objet intelligent par l'utilisateur, l'accès à un service numérique nécessitera un niveau de confiance que le tiers de confiance pourra ou pas garantir de façon transparente pour l'utilisateur.

Dans le cas d'une tentative d'usurpation, la confiance calculée C doit être faible si l'authentification est performante. Nous mettons la confiance globale C_G à zéro pour éviter tout accès illégitime par un imposteur. Dans le cas contraire, la confiance globale est mise à jour à chaque instant t de telle sorte à bénéficier de la confiance obtenue par la dernière authentification ponctuelle :

$$C_G(t) = \begin{cases} min(C_G(t-1) + C, 100) & \text{si } C > Cthreshold \\ 0 & \text{sinon.} \end{cases} \tag{3.9}$$

La confiance globale est évidemment majorée à 100%. La dernière étape est la décroissance de la confiance au cours du temps. Ceci permet de garantir qu'un objet intelligent ne puisse conserver une confiance élevée lorsqu'il n'est pas utilisé. La décroissance est paramétrée par la fonction suivante :

$$C_G(t) = max(0, C_G(t) \times e^{-1/2}) \tag{3.10}$$

La figure 3.4 résume l'évolution de la confiance à partir du score calculé par un des classifieurs présentés dans la section précédente. Dans la section suivante, nous validons le système proposé par des simulations avec des données biométriques réelles dans différents scénarios.

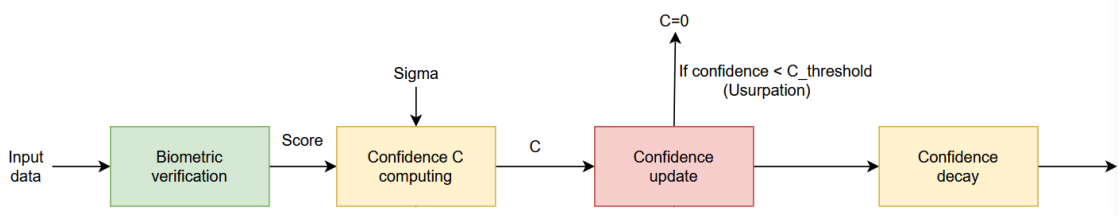

FIGURE 3.4: Schéma explicatif du calcul de la confiance de l'individu via un unique objet intelligent.

3.4 Protocole expérimental

Nous détaillons dans cette partie le protocole expérimental utilisé dans cette thèse pour la validation de la méthode proposée. Nous considérons dans les expériences une solution d'authentification transparente sur ordinateur par reconnaissance faciale et un autre sur smartphone en utilisant les habitudes d'appels de l'individu. Nous présentons tout d'abord les bases de données utilisées pour l'authentification sur smartphone et sur ordinateur ainsi que l'ensemble de pré-traitements appliqués sur ces données. Nous présentons également les métriques de performance employées pour l'évaluation de la confiance.

3.4.1 Bases de données

Nous utilisons deux bases de données. La première interne à Orange contient des données sur les habitudes d'appel d'individu. La seconde est une base de visages de la littérature.

Données collectées sur un smartphone

Un ensemble de données réelles et privées est collecté, incluant le comportement de communication de 93 utilisateurs enregistrés pendant un mois avec plus de 16 000 échantillons. Chaque utilisateur dispose de 8 informations collectées comme suit :

— La date et l'heure de début d'un appel ;
— Le numéro de téléphone de l'appelant ;
— Le numéro de téléphone de la personne appelée ;
— Le type de communication (message textuel SMS ou appel téléphonique) ;
— Le nombre d'unités consommées dans une communication (secondes pour un appel téléphonique, ou nombre de caractères du SMS) ;
— Le type d'appel (appel sortant, appel entrant) ;
— La latitude d'une cellule ;
— La longitude d'une cellule.

Un ensemble de 8 valeurs définit chaque profil d'utilisateur représentant son comportement basé sur une communication téléphonique, comme le montre la figure 3.5. Le nombre d'échantillons est différent d'un utilisateur à un autre, car les données sont collectées de manière réelle et unique pour chaque utilisateur (différents nombres d'appels, différents appelants, etc.). Le nombre total d'échantillons est de 16143. Les échantillons sont ensuite divisés et échangés de manière aléatoire en deux ensembles : l'ensemble de données de référence, dédié à l'entraînement du modèle et l'ensemble de données de test, utilisé pour fournir une évaluation du modèle.

Comme pré-traitement des données, nous transformons toutes les données collectées en valeurs numériques. Par exemple, nous transformons l'heure et la date d'un appel en un nombre réel de secondes calculé depuis une heure prédéfinie, dans notre cas

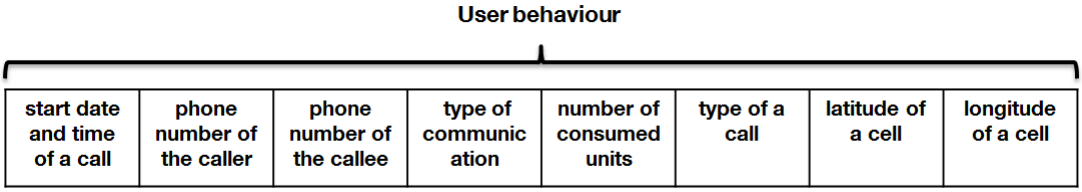

FIGURE 3.5: Comportement de l'utilisateur basé sur les données de communication du smartphone

depuis minuit pour tous les appels. En plus, nous avons traduit toutes les données en digits. Par exemple, la distribution en 10 digits la taille du numéro appelé et le numéro appelant. La concaténation de toutes ces données constitue un vecteur modèle de taille m=53, divisé en ensembles de données de référence et de test.

Données collectées sur un PC

Dans cette partie, nous utilisons les données image de la base publique AR (AR Face Database), qui a été créée par Aleix Martinez et Robert Benavente au Computer Vision Center (CVC) [104]. Cette base contient plus que 4000 images de visages en couleur de 126 personnes (70 hommes et 56 femmes). Chaque personne a 26 échantillons présentant des visages en vue frontale avec différentes expressions faciales, différentes conditions d'éclairage et d'occultations (lunettes de soleil et foulard). Quelques exemples d'images de visage utilisées sont présentés dans la figure 3.6.

FIGURE 3.6: Quelques exemples d'images de visages de la base de données AR

Dans un contexte opérationnel sur la reconnaissance faciale, il est nécessaire de réaliser une détection du visage dans l'image. Depuis, la méthode pionnière de détection de visages par Viola et Jones en 2004 [105], de nombreuses méthodes ont été proposées. Dans cette thèse, nous utilisons une méthode basée sur un apprentissage profond [106]. Cette méthode donne d'excellents résultats comme le montre la figure 3.7.

(a) Examples of results on FDDB

(b) Examples of results on WIDER FACE

FIGURE 3.7: Illustration de l'efficacité de la méthode utilisée de détection de visages (source [106]).

Il existe plusieurs modèles CNN qui ont été entraînés avec succès pour la tâche de reconnaissance des visages. Dans ce travail, nous utilisons le modèle VGG-face proposé par [107], pour extraire les caractéristiques de visage de la base AR. VGG-Face se compose de 11 couches, 8 couches convolutionnelles et 3 couches entièrement connectées. Le nombre de caractéristiques des couches connectées est 4096.

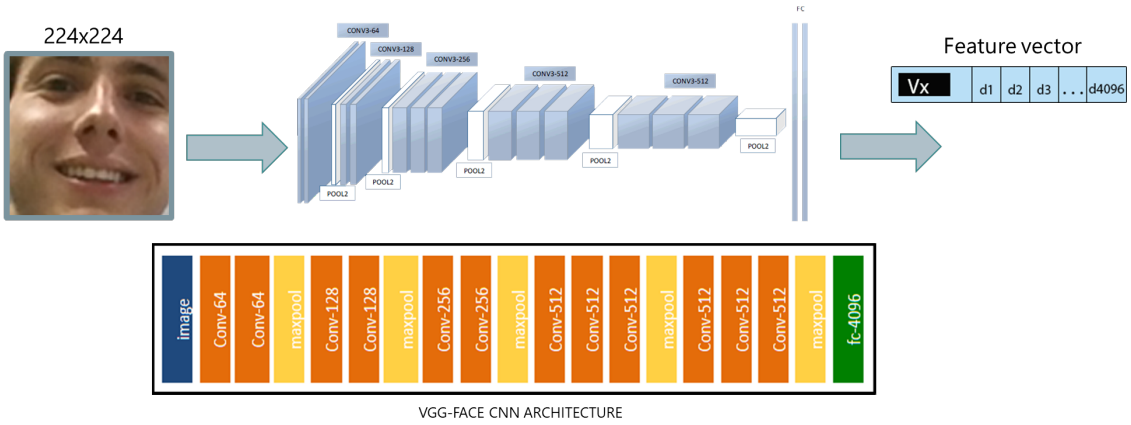

FIGURE 3.8: Illustration de la caractérisation de visages par VGG-Face (à partir de [107]).

3.4.2 Métriques de performance

Il existe 3 protocoles d'évaluation de systèmes biométriques :

— L'évaluation technologique : consiste à tester seulement un composant de la chaîne d'authentification, comme par exemple, un algorithme de matching ou d'extraction de caractéristiques

— L'évaluation par scénario : consiste à tester l'ensemble du système
— L'évaluation opérationnelle : similaire à l'évaluation par scénario, cependant le système est intégré à une application réelle, avec utilisateurs finaux réels.

Dans ce chapitre, nous réalisons une évaluation technologique pour estimer la performance des paramètres biométriques sur des jeux de données dans un premier temps. Nous réalisons une évaluation par scénario dans un second temps du système d'authentification transparente. L'objectif étant de quantifier la performance des systèmes biométriques, nous utilisons les métriques classiques en biométrie ci-dessous :

— **FAR : taux de fausse acceptation (False Acceptance Rate en anglais)** Ce taux représente le pourcentage d'imposteurs acceptés à tort par le système.
— **FRR : taux de faux rejet (False Rejection Rate en anglais)** Ce taux représente le pourcentage d'utilisateur rejeté à tort.
— **EER : taux d'égale erreur (Equal Error Rate en anglais)** Cette métrique représente le taux d'erreur correspondant un paramétrage du seuil de décision du système biométrique pour que la valeur du FAR soit égale à FRR (voir figure 3.9).

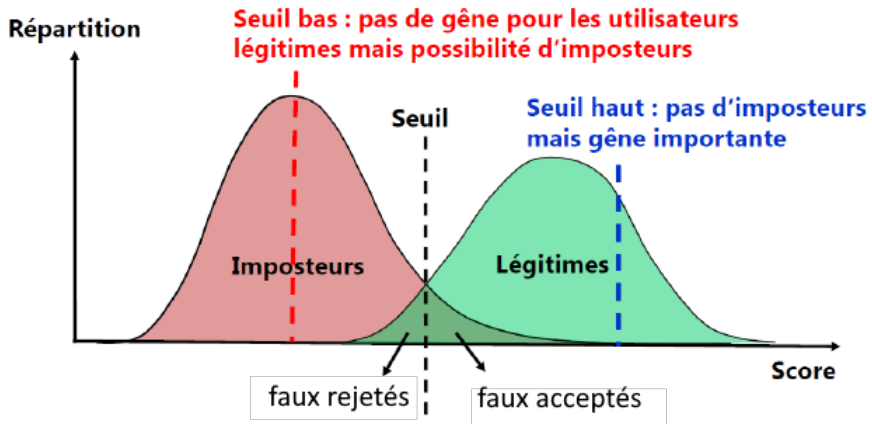

FIGURE 3.9: Distribution des scores légitimes (utilisés pour le calcul du FRR) et des d'imposture (utilisés pour le calcul du FAR) en fonction du seuil de décision du système biométrique.

La performance d'un système biométrique est mesurée via les métriques FAR et du FRR. Ces deux mesures peuvent être contrôlées en ajustant le seuil de décision du système biométrique. Dans un monde idéal, un système parfait a une valeur de l'EER à zéro. En pratique, cela est presque impossible puisqu'il est compliqué d'obtenir un FAR et un FRR proches de zéro compte tenu de la variabilité intrinsèque de la capture des données biométriques.

La courbe DET (souvent appelée ROC) est utilisée pour représenter l'efficacité d'un système biométrique quelque soit le paramétrage du seuil de décision (voir Figure

3.10). Elle représente l'évolution du FAR en fonction du FRR. L'EER est l'intersection de la courbe et de la diagonale passant par l'origine. Plus l'aire sous cette courbe tend vers zéro, plus le système est performant.

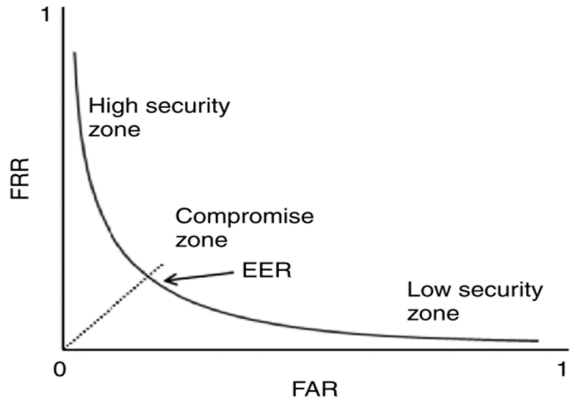

FIGURE 3.10: Courbe DET décrivant la performance d'un système biométrique.

3.5 Résultats expérimentaux

Cette section présente les résultats expérimentaux obtenus afin d'évaluer la performance du système proposé. Dans un premier temps, nous évaluons la performance des données biométriques dans un contexte classique d'authentification biométrique statique. La performance de la chaîne de traitement est analysée en testant l'impact de l'étape de protection des données pour la vie privée et en implémentant une attaque pour la sécurité.

3.5.1 Évaluation de la performance des données

Nous évaluons d'abord la performance des données collectées pour déduire sur leur fiabilité d'usage et de sécurité dans un système d'authentification. La performance du système d'authentification est évaluée par trois mesures principales qui sont le FAR, le FRR et le EER. Nous adoptons 2 scénarios de test avec ou sans protection des données. Comme la base de données de référence et la base de données de test sont construites aléatoirement pour un ratio donné, il est fortement recommandé de faire la moyenne de la valeur du EER résultant de chaque tirage aléatoire. Le ratio est le pourcentage de division des données de référence et des données de test. Ainsi, pour une valeur fixe d'un ratio, la valeur du EER est calculée 10 fois et ensuite la moyenne des valeurs obtenues est considérée comme la valeur finale du EER attribuée à un ratio fixe.

Évaluation de la performance des données sur les habitudes d'appel

Nous présentons dans cette partie les résultats expérimentaux trouvés en utilisant les données des habitudes d'appels issues d'un smartphone. Les trois méthodes

de classification sont appliquées séparément afin de comparer leurs performances en utilisant la base de données comportementales en termes de valeur du EER (qui doit être minimisée). Nous allons également estimer l'impact de l'étape de protection des données sur la performance du système et sa sécurité.

Scénario 1 : Performances du système sans protection des données

Ce premier scénario évalue essentiellement les performances des différents algorithmes de classification sans aucune protection des données, afin de déterminer la meilleure approche pour une meilleure authentification. La minimisation de la distance est implémentée par un simple calcul de la distance euclidienne entre l'ensemble des données de test et le modèle décrit dans le protocole proposé. Un SVM à classe unique est utilisé avec le noyau RBF qui donne de meilleurs résultats que les autres types de noyaux. L'algorithme K-means est utilisé avec K=3 classes. Les résultats sont présentés en termes de valeurs EER pour un rapport = 80% (ainsi, 20% des données sont utilisées pour les tests) et sont résumés dans le tableau 3.3. Les courbes DET présentant les valeurs FRR par rapport aux valeurs FAR pour les 3 classifieurs utilisés sont exposées dans la figure 3.11.

TABLE 3.3: Comparaison des valeurs de l'EER obtenues par les 3 classifieurs des données biométriques basées sur les habitudes d'appel sans protection.

Méthode de classification	EER (%)
Minimisation de distance	23.9
K-means	39
One class SVM	10.4

Des meilleurs résultats sont obtenus avec le classifieur SVM à classe unique avec un EER= 10,4%. De plus, la valeur de l'EER dépend de la quantité de données utilisée pour l'enrôlement et donc de la valeur du ratio utilisé. Pour cette raison, les figures 3.12, 3.13, 3.14 montrent respectivement la l'évolution de l'EER en fonction de la quantité de données biométriques utilisées pour la référence de l'utilisateur. Il est clair que pour les 3 classifieurs, plus le rapport est élevé, plus le EER est faible.

Scénario 2 : Performances du système avec protection des données

Dans ce second scénario, la protection des données est prise en compte. Les données collectées sont protégées par l'algorithme de Biohashing décrit dans le chapitre 2. L'identifiant de l'utilisateur a été utilisé comme clé secrète dans l'algorithme du BioHashing et le BioCode est de longueur 50 bits. Dans un contexte opérationnel, on peut utiliser un code PIN comme clé secrète pour protéger les données collectées afin de générer le BioCode. Les BioCodes des différents utilisateurs sont présentés sous la forme d'un code barres comme le montre la figure 3.15. Les résultats expérimentaux des 3 méthodes de classification sont présentés en termes de valeurs de

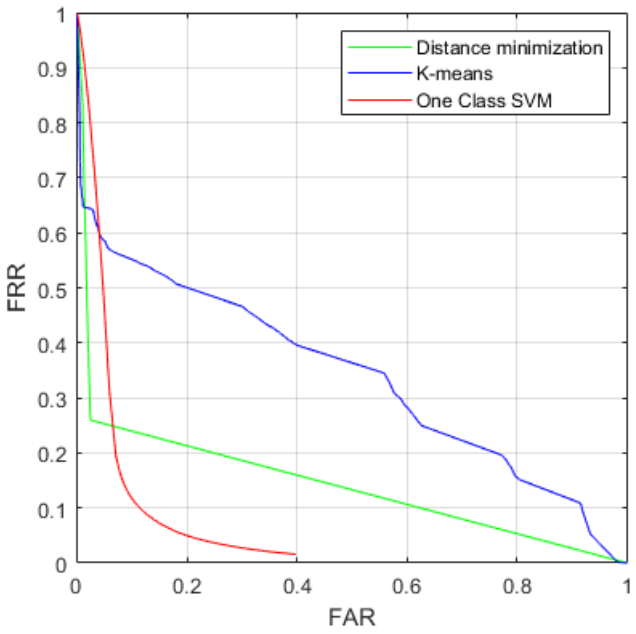

FIGURE 3.11: Courbes ROC pour les 3 classifieurs (ratio=80%) sur les données biométriques basées sur les habitudes d'appel sans protection.

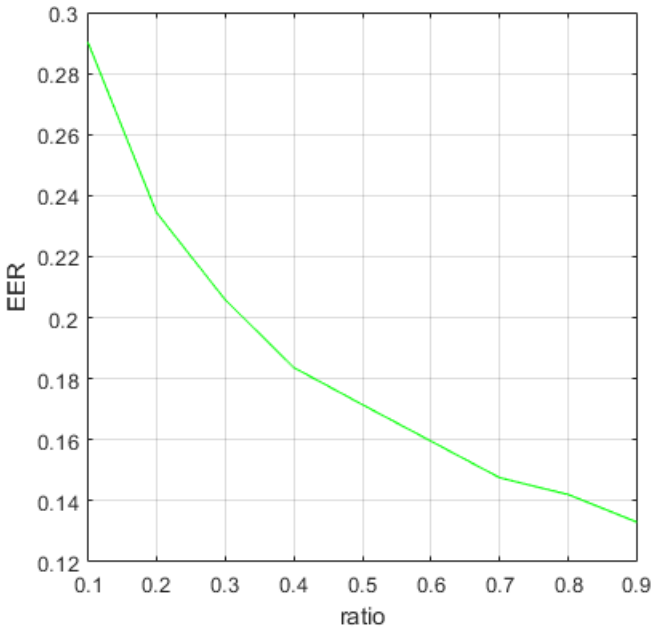

FIGURE 3.12: Evolution de l'EER pour la minimisation de distance des données biométriques basées sur les habitudes d'appel sans protection.

l'EER pour un ratio = 80%, ils sont résumés dans le tableau 3.4.

Les courbes ROC présentant les valeurs FRR par rapport aux valeurs FAR pour les 3 classifieurs évaluées avec l'algorithme BioHashing sont exposées dans la figure

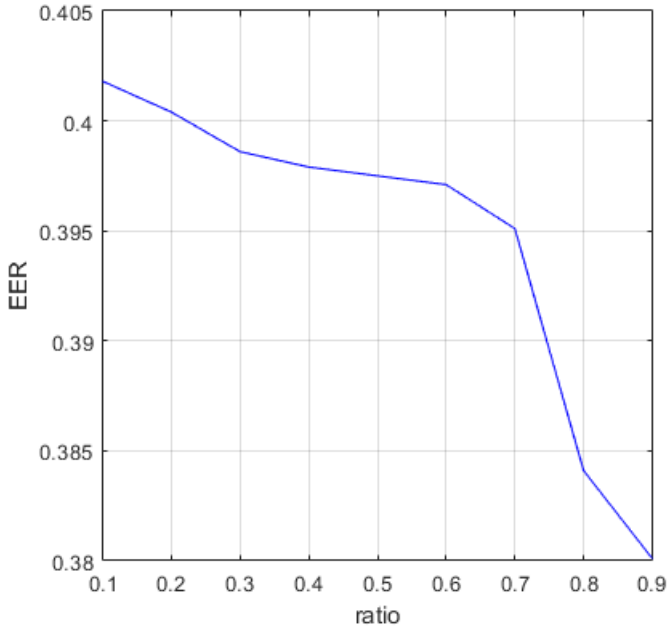

FIGURE 3.13: Evolution de l'EER pour la méthode K-means (K=3) des données biométriques basées sur les habitudes d'appel sans protection.

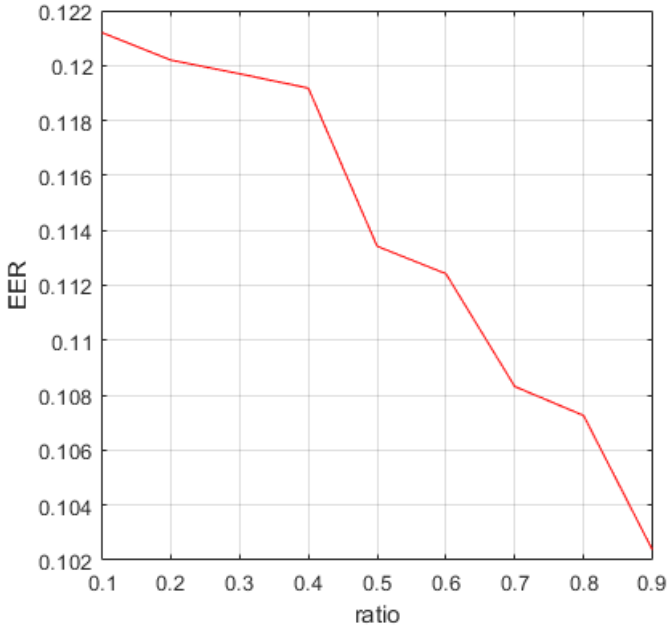

FIGURE 3.14: Evolution de l'EER avec la méthode SVM sur les données biométriques basées sur les habitudes d'appel sans protection.

3.16. Il est clairement visible que la performance des 3 méthodes de classification avec l'algorithme BioHashing est améliorée par rapport au cas non protégé. Le SVM à classe unique reste le meilleur classifieur avec un EER= 0,02%, mais on peut également noter que la performance de l'algorithme K-means avec protection de la vie

FIGURE 3.15: Codes-barres de différents BioCodes d'utilisateurs.

TABLE 3.4: Comparaison des valeurs d'EER obtenues par les 3 classifieurs avec protection des données biométriques basées sur les habitudes d'appel.

Méthode de classification	EER (%)
Minimisation de distance	0.02
K-means	0.14
SVM	0.02

privée s'est améliorée de manière très importante et l'EER passe d'une valeur de 37% à 0,14% , ce qui n'est pas le meilleur résultat mais reste intéressant. De plus, comme étudié dans le premier scénario, les valeurs du EER varient en fonction des valeurs de la quantité de données en apprentissage. Pour cela, à titre d'exemple, la figure 3.20 montre la variation du EER en fonction des valeurs du ratio avec protection des données pour la méthode SVM à classe unique. Les figures 3.17, 3.18 et 3.19 montrent respectivement les performances avec et sans protection de la vie privée des 3 classifieurs. On peut constater que la performance de reconnaissance est stable lorsque les données sont protégées.

Impacts des paramètres

Dans cette partie, nous étudiants l'impact des paramètres de la chaîne de traitement des données biométriques dans un contexte d'authentification statique :

— **Type de noyau SVM** : Afin de déterminer quel type de noyau SVM est approprié à nos données, nous avons testé 3 types de noyaux différents comme indiqué dans le tableau 3.5. Pour cette expérience, nous avons utilisé les données d'appel. Il est clair que de meilleurs résultats sont obtenus avec le noyau de type RBF.

TABLE 3.5: Valeurs de l'EER en fonction des types de noyau du SVM pour les données biométriques basées sur les habitudes d'appel.

Type du noyau	EER(%) sans protection	EER(%) avec protection
Linéaire	51.9	47.1
Polynomial	77.1	53.4
RBF	10.4	0.02

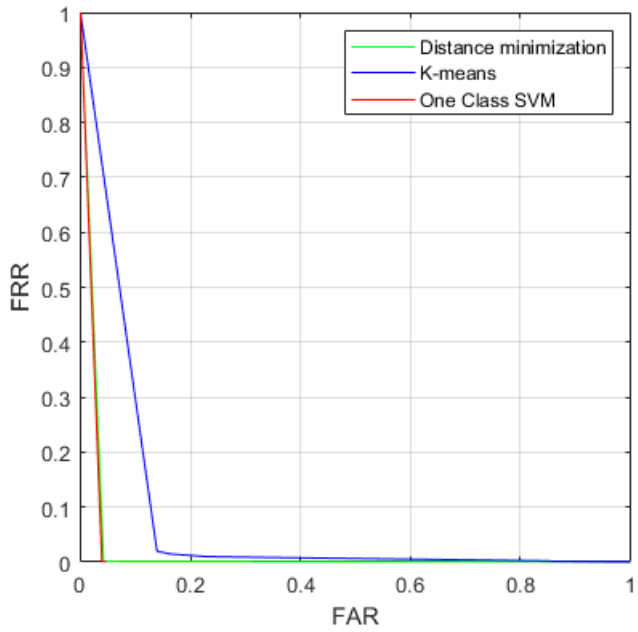

FIGURE 3.16: Courbes ROC pour les 3 classfieurs (ratio=80%) sur les données biométriques basées sur les habitudes d'appel avec protection.

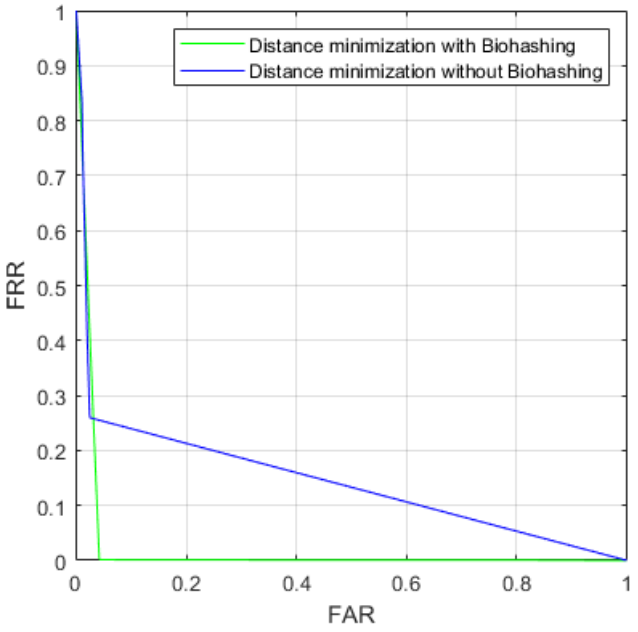

FIGURE 3.17: courbes ROC pour la minimisation de distance avec et sans BioHashing (ratio=80%) sur les données biométriques basées sur les habitudes d'appel.

— **Valeur de K** : Nous avons testé différentes valeurs de K (nombre de centroides dans l'algorithme de Kmeans) afin de déterminer la valeur la plus appropriée pour traiter nos données. En représentant K par rapport à l'EER, comme le montre la figure 3.21, nous constatons que l'erreur diminue à fur et à me-

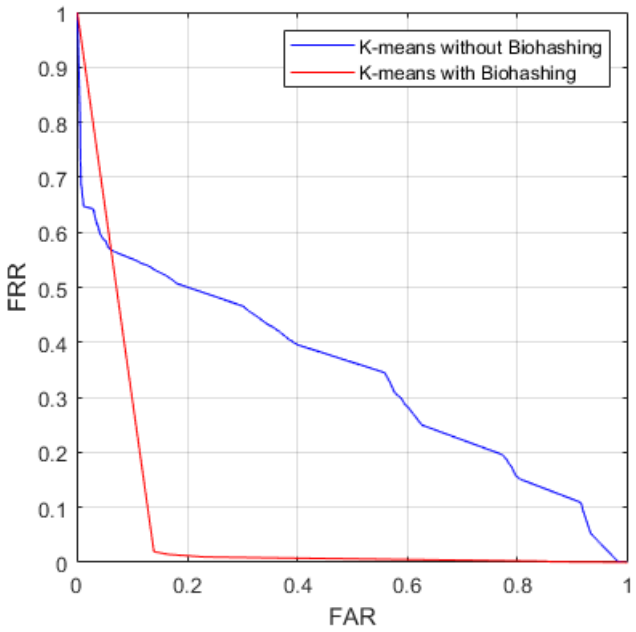

FIGURE 3.18: Courbes ROC pour le K-means avec et sans BioHashing (ratio=80%) sur les données biométriques basées sur les habitudes d'appel.

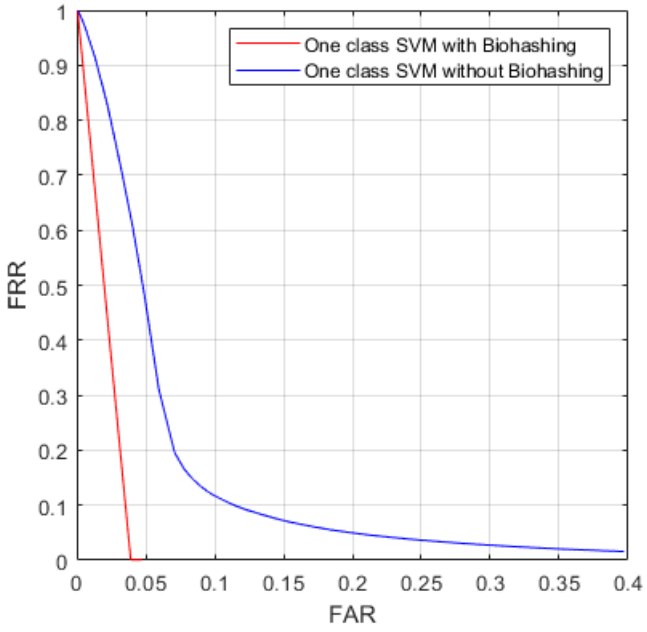

FIGURE 3.19: Courbes ROC pour le SVM avec et sans BioHashing (ratio=80%) sur les données biométriques basées sur les habitudes d'appel.

sure que K augmente. Cela s'explique par le fait que les clusters deviennent plus petits lorsque leur nombre augmente, et donc que la corrélation est également plus faible. Nous choisissons la valeur K à laquelle le EER diminue brusquement et reste presque stable après. Dans notre cas, K=3 est la valeur

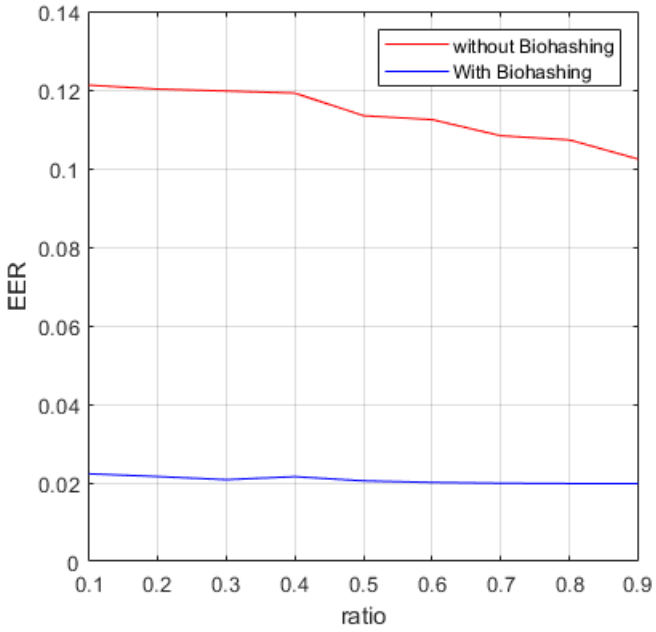

FIGURE 3.20: Evolution de la valeur de l'EER avec la méthode SVM avec et sans BioHashing en faisant évoluer le ratio de la base d'apprentissage.

appropriée.

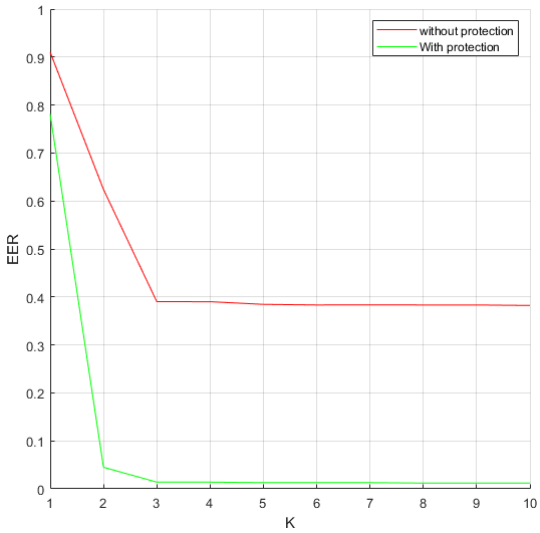

FIGURE 3.21: Les impacts de la valeur de K sur les résultats avec K-means

— **Valeur du Ratio (nombre d'échantillons dans la référence)** : Les valeurs de l'EER dépendent des valeurs des ratios impactant le nombre d'échantillons dans la référence (données biométriques sur les habitudes d'appel). Pour cette raison, différentes valeurs de ratio sont testées. La figure 3.22 montre la variation de l'EER en fonction de la variation du ratio pour les 3 classifieurs. Il est clair que pour les 3 méthodes, plus le ratio est élevé, plus le EER est faible.

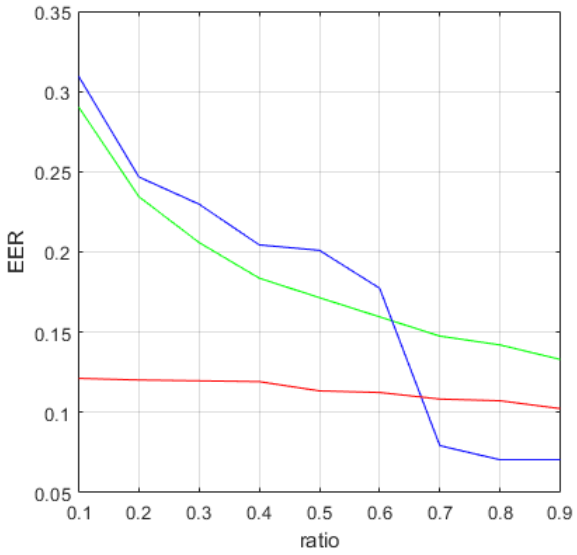

FIGURE 3.22: Variation de l'EER pour minimisation de la distance, K-means et One class SVM

— **Révélation de la clé secrète de protection** : En étudiant la protection des données, nous avons d'abord évalué les performances du système avec l'algorithme BioHashing en utilisant une clé secrète unique pour chaque utilisateur. Ici, nous voulons étudier les impacts si la clé était révélée sur la performance. Nous avons donc assigné à chaque utilisateur la même clé. Nous avons utilisé les données biométriques basées sur les habitudes d'appel. Le tableau 3.6 présente les valeurs de l'EER des données protégées avec l'algorithme BioHashing en utilisant une clé secrète unique et une même clé révélée, pour les 3 classifieurs. Avec une clé secrète unique pour chaque utilisateur, les résultats de la classification sont bien meilleurs que l'utilisation d'une clé révélée pour tous les utilisateurs, ce qui correspond aux conditions d'usage.

TABLE 3.6: Impacts des différentes clés de l'algorithme BioHashing sur les valeurs de l'EER des données biométriques basées sur les habitudes d'appel protégées pour les 3 classifieurs.

Méthodes de classification	Clé révélée	Clé secrète
Minimisation de la distance	19.94	0.02
K-means	27.15	0.14
One class SVM	6.73	0.02

Évaluation de la performance des données faciales

Nous présentons dans cette partie les résultats expérimentaux en utilisant les images du visage capturées par une webcam (contexte PC portable ou fixe). Nous adoptons les 2 mêmes scénarios à savoir l'analyse de la performance sans et avec

protection de données.

Scénario sans protection de données

La minimisation de distance est implémentée par un simple calcul de la distance euclidienne entre l'ensemble des données de test et le modèle décrit dans le protocole proposé. Pour les données du visage, une seule image de chaque personne est suffisante comme référence.La courbe ROC présentant les performances du système sans protection de données est représentée par la figure 3.23. Pour un ratio de 80%, nous obtenons une valeur de EER = 13%.

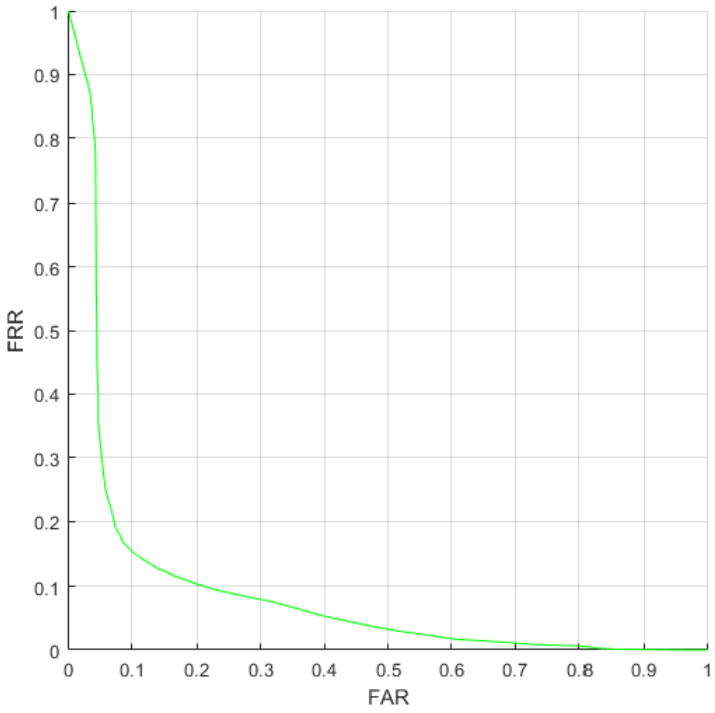

FIGURE 3.23: Courbe ROC pour la minimisation de distance (ratio=0.8) sur les données faciales sans protection.

Scénario avec protection de données

Dans ce scénario, nous appliquons l'algorithme de Biohashing sur les données faciales. Nous avons choisi une taille de BioCode de 512 bits. Nous avons généré la courbe ROC dans la figure 3.24. Dans un premier temps, nous utilisons une clé secrète pour l'algorithme BioHashing, qui est l'identifiant de l'utilisateur. Nous remarquons que la performance avec protection clé secrète est meilleure avec un EER = 0.02%. Dans le cas où la clé est révélée, nous obtenons une performance proche à celle sans protection de donnée dans la figure 3.25 avec une valeur d'EER = 0.09%.

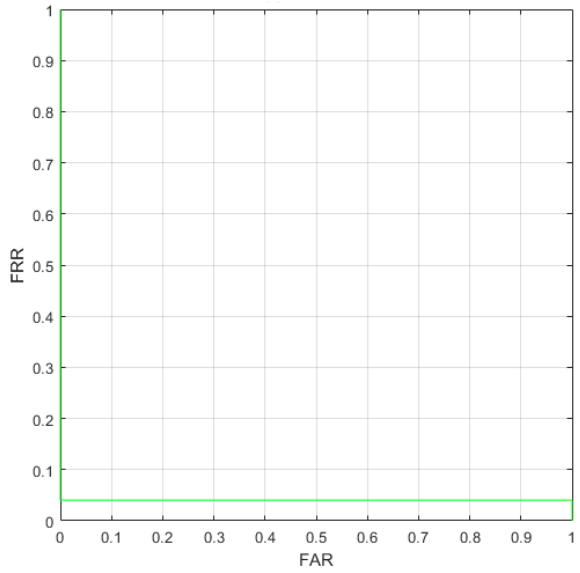

FIGURE 3.24: Courbe ROC pour la minimisation de distance (ratio=80%) de données faciales protégées (clé secrète).

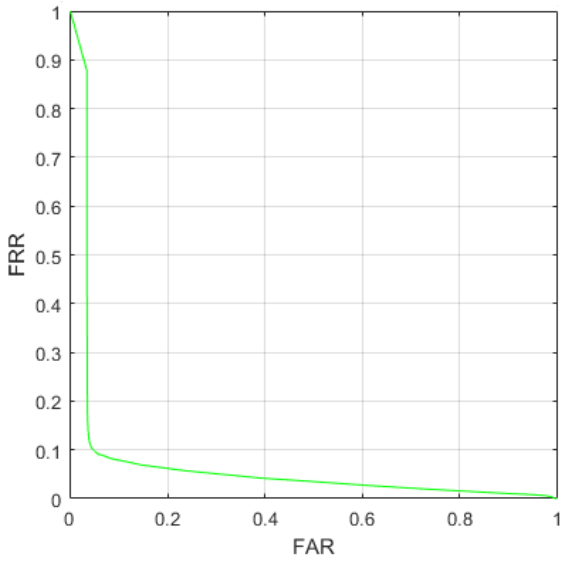

FIGURE 3.25: Courbe ROC pour la minimisation de distance (ratio=80%) des données faciales protégées (clé révélée).

Discussion

Comme le montrent les résultats expérimentaux, la chaîne de traitement allant de la collecte de données biométriques (comportementales ou morphologiques), le prétraitement, la protection par l'algorithme de BioHashing et la classification donne de très bons résultats. Si la clé secrète n'est pas compromise, nous obtenons pour les

2 modalités biométriques, un EER de 0.02%. Ceci démontre l'intérêt de la chaîne de traitement proposée pour une application en authentification biométrique statique (ou classique). Dans la section suivante, nous allons estimer la performance de cette chaîne de traitement dans un contexte d'authentification transparente.

3.5.2 Évaluation de l'authentification transparente

Dans cette partie, nous souhaitons évaluer la solution d'authentification transparente proposée via un unique objet intelligent. Nous allons étudier l'évolution de la confiance de l'usage d'un device au cours d'une journée. Nous présentons dans la suite l'étude de l'authentification transparente basées sur les données biométriques faciales. Nous allons choisir arbitrairement des périodes d'usages légitimes de l'objet intelligent (qui sera représentée en vert) et des tentatives d'imposture (représentées en rouge). Nous allons pouvoir observer l'évolution de la confiance globale définie dans l'équation 3.9. Le seuil utilisé pour la détection d'une usurpation d'identité a été fixé à 50% dans toutes les simulations.

Nous allons simuler l'usage d'un objet intelligent pendant 24 heures. Nous mesurons une authentification ponctuelle toutes les 3 minutes, ce qui fait 20 mesures par heure. Chaque mesure est réalisée à partir d'une donnée tirée aléatoirement dans les bases de données biométriques décrites dans la section précédente. Ce scénario permet de tester à la fois des tentatives légitimes et d'imposture afin d'observer l'évolution de la confiance au cours du temps. Nous débutons par la la méthode d'authentification transparente basée sur les données faciales.

Authentification transparente basée sur la reconnaissance faciale

Dans ce contexte, nous simulons le fait que lorsqu'un individu utilise un ordinateur portable ou fixe, il est possible de prendre de façon transparente pour lui, des photos de son visage toutes les 3 minutes. Dans des conditions réelles, le visage de l'utilisateur est susceptible de changer pour différentes raisons : des variations d'éclairage, des changements d'expressions, le visage peut être caché par un masque, des lunettes, etc... Pour ce fait, nous souhaitons présenter les résultats de cette partie en prenant en compte l'impact de ces différents changements sur l'évolution de la confiance lors d'une authentification transparente.

Nous devons fixer quelques paramètres notamment σ qui permet de convertir un score de comparaison de 2 visages en une confiance entre 0 et 100%. Nous fixons la valeur de σ fixé à 0.5 pour les données brutes et σ=0.25 pour les données protégées (cette valeur dépend de la distance utilisée). Nous montrons l'impact du changement de contexte sur l'évolution de la confiance, pour les 2 scénarios : avec et sans protection de données :

— **Quel est l'impact du changement d'expressions sur l'évolution de la confiance ?**

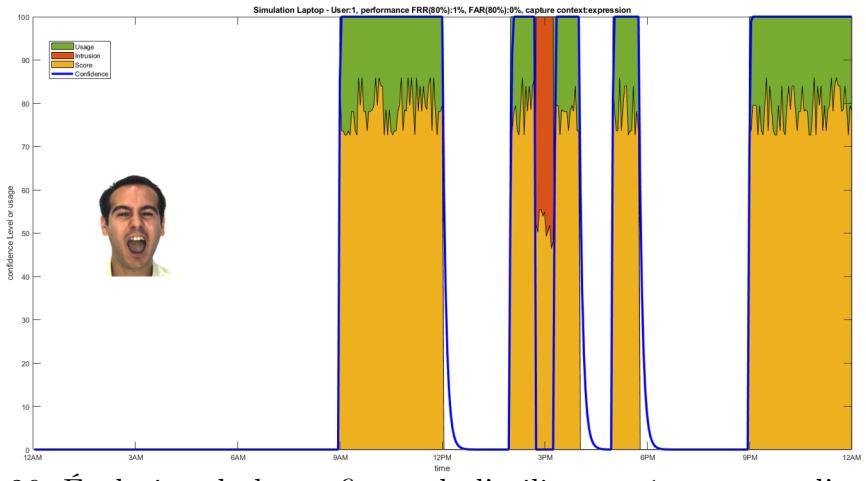

FIGURE 3.26: Évolution de la confiance de l'utilisateur 1 au cours d'une journée pendant l'utilisation d'un ordinateur en considérant le contexte du changement d'expressions.

La figure 3.26 montre l'évolution de la confiance de l'utilisation d'un ordinateur de l'utilisateur 1 au cours de la journée en considérant le contexte du changement d'expressions (sans protection). En abscisse, nous représentons les heures d'usage de l'ordinateur (fixées arbitrairement) et en ordonnée, les scores d'authentification ponctuelle (en jaune) toutes les 3 minutes et la confiance globale calculée (en bleu). En vert, nous simulons des tentatives légitimes et en rouge des tentatives d'imposture. La valeur des scores d'authentification ponctuelle dans ces deux contextes sont très différents.

Selon le scénario étudié, on a une utilisation légitime de l'ordinateur de 9h à midi par exemple ce qui est traduit par une confiance à 100%. La confiance décroît exponentiellement jusqu'à 0% quand l'utilisateur n'utilise pas son ordinateur, de midi à 14h par exemple. En effet, à chaque instant, la confiance est décrémentée à chaque intervalle de temps (voir équation 3.10) jusqu'à 0%. Un score d'authentification ponctuelle est calculé toutes les 3 minutes en cas de l'usage de l'ordinateur et permet de mettre à jour la confiance globale.

Dans le cas où on détecte une tentative d'intrusion, quand un imposteur essaye d'utiliser le l'ordinateur de l'utilisateur 1 (la zone rouge sur la figure 3.26), le score est trop faible et la confiance associée est mise à 0%. Pour évaluer la performance de l'authentification transparente, nous avons fixé un seuil de décision à 80%. Ce seuil nous permet par la suite de calculer les valeurs du taux de faux rejet FRR et de fausse acceptation FAR. Ceci nous permet d'évaluer la performance (à partir de quelle valeur de seuil, on considère que la confiance est acceptée). Pour ce seuil de décision, le FRR traduit combien de fois la confiance n'est pas au moins 80% quand on est dans un cas d'usage légitime. Le FAR mesure combien de fois la confiance est au delà de 80% quand on est

dans une tentative d'intrusion. En considérant uniquement des images avec des changements d'expressions pour l'utilisateur 1 dans la figure 3.26, pour un seuil = 80%, nous obtenons des valeurs du FRR(80%) = 0% et un FAR(80%) = 0%.

Dans un second scénario, la protection de données est prise en compte. Les données collectées sont protégées par l'algorithme de Biohashing. Nous présentons d'abord les résultats en appliquant l'algorithme de BioHashing avec clé secrète, et avec clé révélée par la suite. Les figures 3.27 et 3.28 montrent respectivement l'évolution de la confiance de l'utilisateur 1 au cours de la journée en considérant le contexte du changement d'expressions avec protection des données faciales dans le cas où la clé est secrète puis révélée.

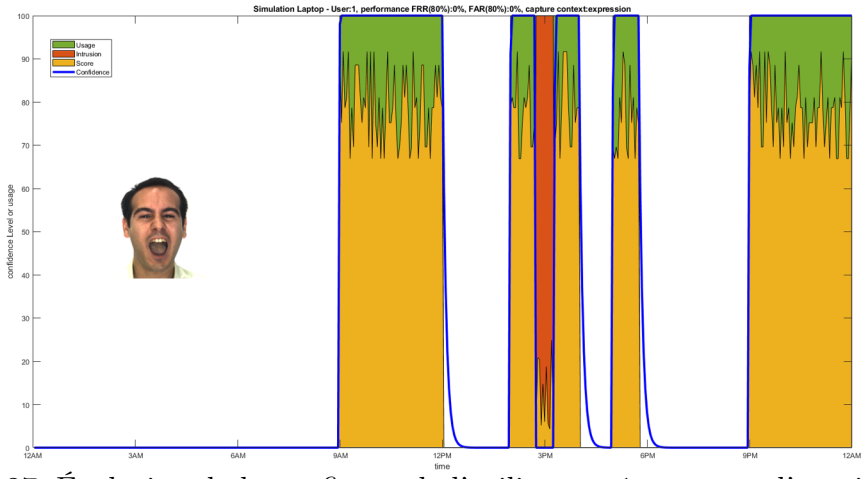

FIGURE 3.27: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement d'expressions avec protection (clé secrète).

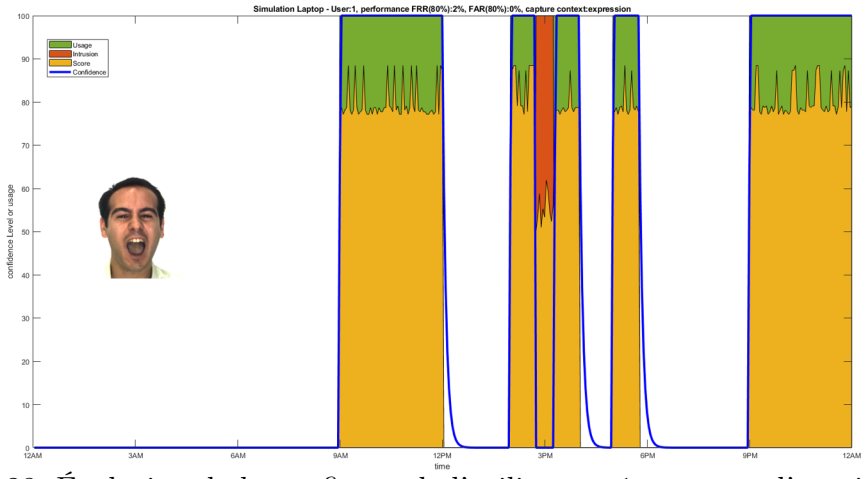

FIGURE 3.28: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement d'expressions avec protection (clé révélée).

Nous remarquons sur la figure 3.27 que dans un cas d'attaque, le score calculé avec des données protégées à clé secrète est plus faible qu'avec des données brutes. Par conséquent, l'application du BioHashing à clé secrète améliore les performances du système et protège l'accès en cas d'intrusion. Pour un seuil = 80%, nous obtenons également un FRR(80%) = 0% et un FAR(80%) = 0%. Même dans le cas d'une clé révélée, l'évolution de la confiance reste proche à celle des données brutes avec un FRR(80%) = 2% et un FAR(80%) = 0%. Nous pouvons donc déduire que le changement d'expression n'a pas d'impact sur l'évolution de la confiance et le système est capable de reconnaître correctement le visage de la personne légitime avec les changements d'expressions possibles, pour assurer son authentification.

Afin de mieux appuyer cette conclusion, nous montrons dans le même contexte les résultats obtenus pour les utilisateurs 11 et 86 de la base AR. Les figures 3.29, 3.30 et 3.31 montrent respectivement l'évolution de la confiance de l'utilisateur 11 au cours de la journée en considérant le contexte du changement d'expressions sans protection de données, avec protection de données (clé secrète) et avec protection de données (clé révélée).

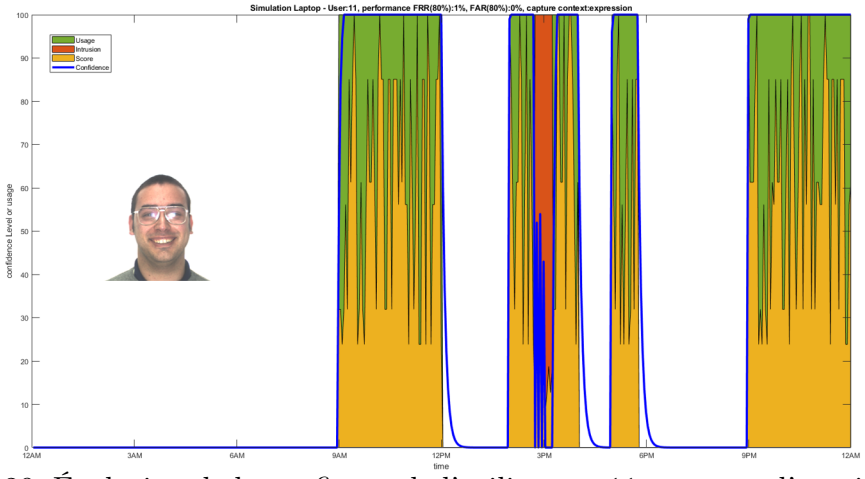

FIGURE 3.29: Évolution de la confiance de l'utilisateur 11 au cours d'une journée en considérant le contexte du changement d'expressions

Pour l'utilisateur 11, dans le scénario de données brutes, nous obtenons un FRR(80%) = 1% et un FAR(80%) = 0% et dans le scénario de données protégées à clé secrète , nous obtenons un FRR(80%) = 2% et un FAR(80%) = 0%, et à clé révélée, nous obtenons un FRR(80%) = 1% et un FAR(80%) = 1%. Les figures 3.32, 3.33 et 3.34 montrent respectivement l'évolution de la confiance de l'utilisateur 86 au cours de la journée en considérant le contexte du changement d'expressions sans protection de données, avec protection de données (clé secrète) et avec protection de données (clé révélée). Pour l'utilisateur 86, nous obtenons un FRR(80%) = 0% et un FAR(80%) = 0% pour les 2 scénarios, sans et avec protection de données. En conclusion, ces résultats

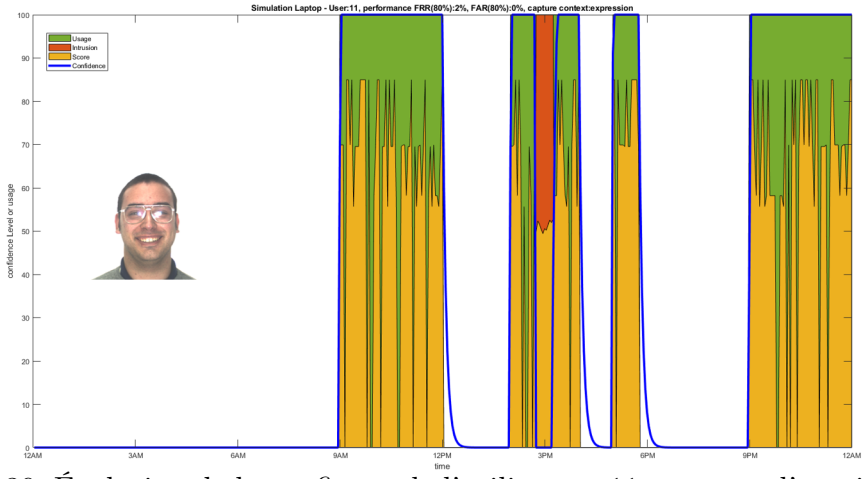

FIGURE 3.30: Évolution de la confiance de l'utilisateur 11 au cours d'une journée en considérant le contexte du changement d'expressions avec protection (clé secrète).

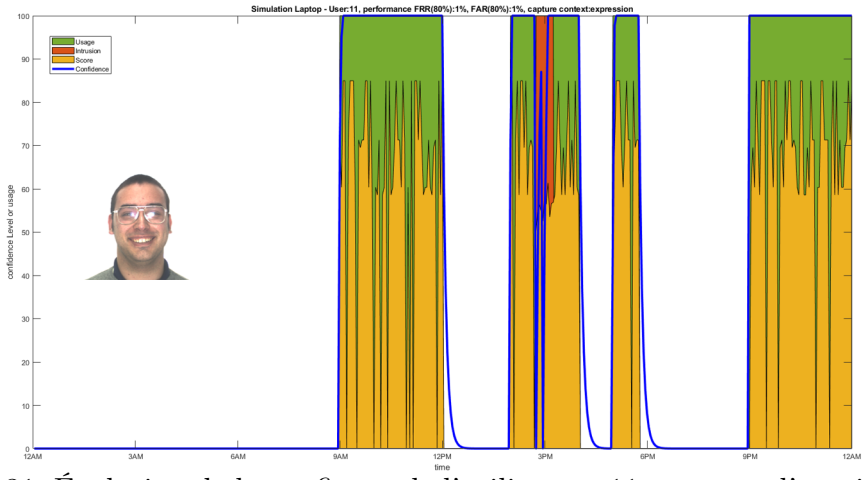

FIGURE 3.31: Évolution de la confiance de l'utilisateur 11 au cours d'une journée en considérant le contexte du changement d'expressions avec protection (clé révélée).

montrent une excellente performance dans un contexte d'acquisition avec uniquement des changements d'expression.

— **Quel est l'impact du changement de lumière sur l'évolution de la confiance ?**

De la même façon que le changement d'expressions, nous étudions l'impact de variation de lumière sur l'évolution de la confiance. Nous générons la courbe de confiance de l'utilisateur 1 au cours d'une journée en tenant compte que des images de changements de lumière de la base AR, donnée par la figure 3.35. Dans ce contexte, nous obtenons un $\text{FRR}(80\%) = 11\%$ et un $\text{FAR}(80\%) = 0\%$. Les figures 3.36 et 3.37 montrent respectivement l'évolution de la confiance de l'utilisateur 1 au cours de la journée en considérant le contexte du changement

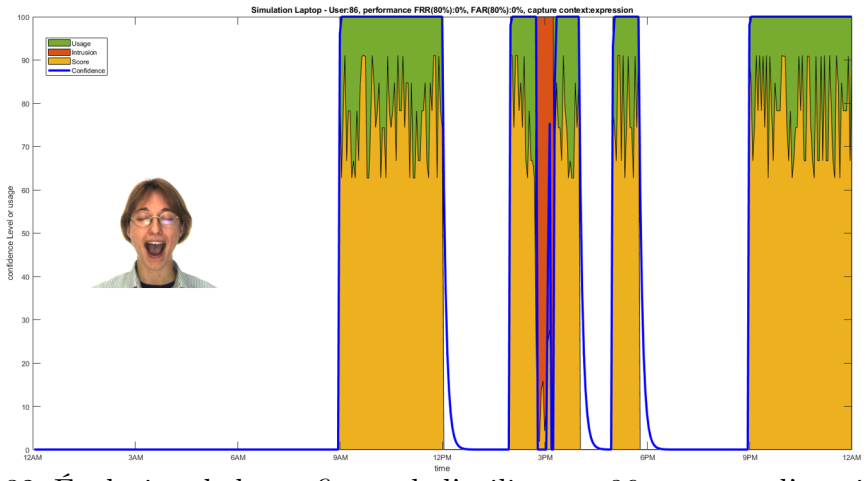

FIGURE 3.32: Évolution de la confiance de l'utilisateur 86 au cours d'une journée en considérant le contexte du changement d'expressions.

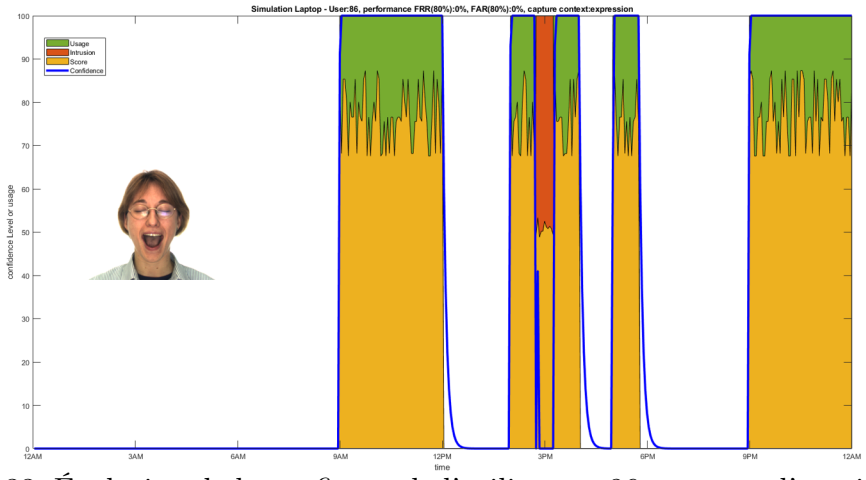

FIGURE 3.33: Évolution de la confiance de l'utilisateur 86 au cours d'une journée en considérant le contexte du changement d'expressions avec protection (clé secrète).

de lumière avec protection de données (clé secrète) et avec protection de données (clé révélée). Avec protection des données à clé secrète, nous obtenons un FRR(80%) = 10% et un FAR(80%) = 0% et à clé révélée, un FRR(80%) = 16% et un FAR(80%) = 0%. L'impact de la variation de lumière sur l'évolution de la confiance est plus important que l'impact du changement d'expressions, mais il est considéré faible (FRR(80%) = 10%) pour affecter l'authentification transparente de l'utilisateur.

— **Quel est l'impact du changement d'occultations sur l'évolution de la confiance ?**

Pour le contexte d'occultations partielles du visage (lunettes de soleil, cache-col, masque...), ceci pourrait, à priori, affecter la performance du système.

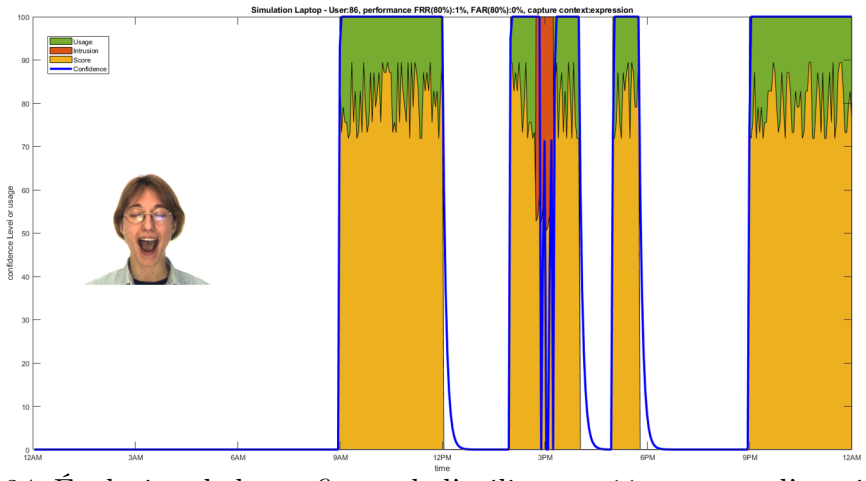

FIGURE 3.34: Évolution de la confiance de l'utilisateur 11 au cours d'une journée en considérant le contexte du changement d'expressions avec protection (clé révélée).

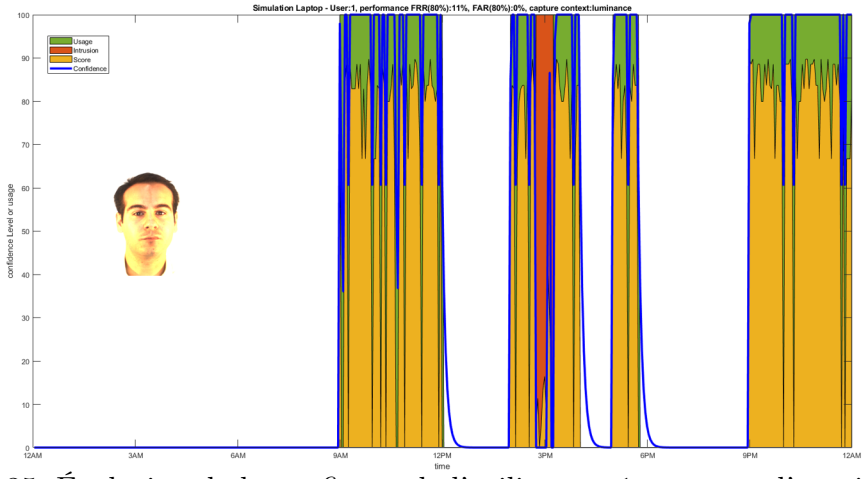

FIGURE 3.35: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement de lumière (sans protection).

Après génération des résultats avec les images contenant des occultations de visage, nous obtenons dans le scénario non protégé, un FRR(80%) = 15% et un FAR(80%) = 0%. L'évolution de la courbe de confiance de l'utilisateur 1, dans un contexte de changement d'occultations, est donnée dans la figure 3.38. Dans le scénario protégé à clé secrète, nous obtenons un FRR(80%) = 1% et un FAR(80%) = 0%, et à clé révélée, nous obtenons un FRR(80%) = 20% et un FAR(80%) = 0%. Les figures 3.39 et 3.40 montrent respectivement l'évolution de la confiance de l'utilisateur 1 au cours de la journée en considérant le contexte du changement d'occultation avec protection de données (cas d'une clé secrète et révélée).

Nous déduisons que l'occultation partielle du visage de l'utilisateur affecte l'évolution de la confiance dans une authentification transparente sans protec-

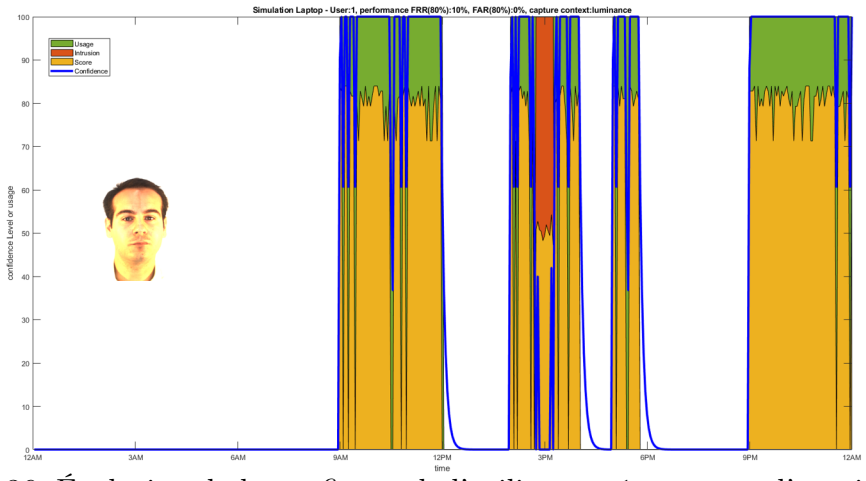

FIGURE 3.36: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement de lumière avec protection de données (clé secrète).

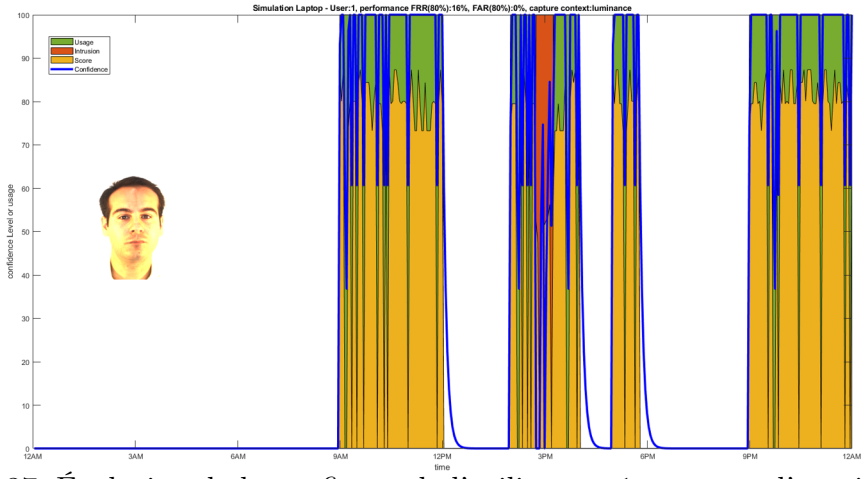

FIGURE 3.37: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement de lumière avec protection de données (clé révélée).

tion (FRR=15%). En appliquant la protection de données avec clé secrète, cet impact s'annule (FRR=1%) même dans ces conditions extrêmes.

— **Quel est l'impact de tous les changements sur l'évolution de la confiance ?**

Maintenant, considérons tous les changements possibles, on prend la première image de l'utilisateur dans la base AR comme référence, et on tire aléatoirement un échantillon des 25 images restantes pour calculer le score de l'authentification ponctuelle. L'évolution de la courbe de confiance en considérant tous les changements possibles pour l'utilisateur 1 est donnée dans la figure 3.41.

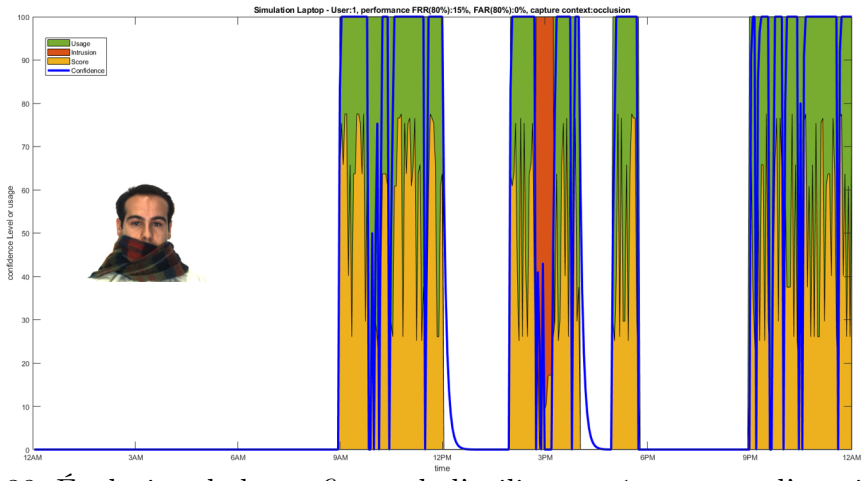

FIGURE 3.38: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement d'occultations (sans protection).

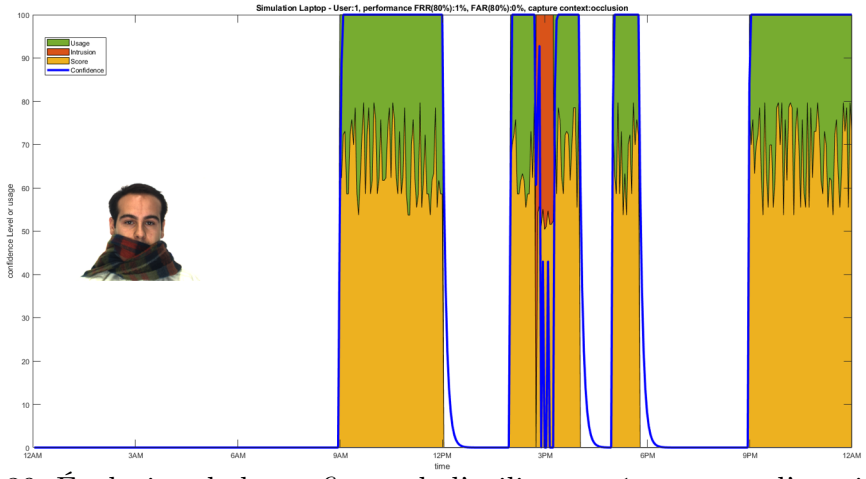

FIGURE 3.39: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement d'occultations avec protection de données (clé secrète).

Nous obtenons un FRR(80%) = 8% et un FAR(80%) = 0% sans protection. Les figures 3.42 et 3.43 montrent respectivement l'évolution de la confiance de l'utilisateur 1 au cours de la journée en considérant tous les changements de contexte avec protection de données (cas d'un clé secrète et révélée). Dans le scénario protégé à clé secrète, nous obtenons un FRR(80%) = 12% et un FAR(80%) = 0%, et à clé révélée, nous obtenons un FRR(80%) = 15% et un FAR(80%) = 0%.

Nous remarquons alors, qu'en considérant qu'un seul changement spécifique (expressions, lumière, occultation) ou en tenant compte de tous les changements possibles, l'impact de ces changements n'affecte pas l'évolution de la confiance même lors d'une tentative d'intrusion.

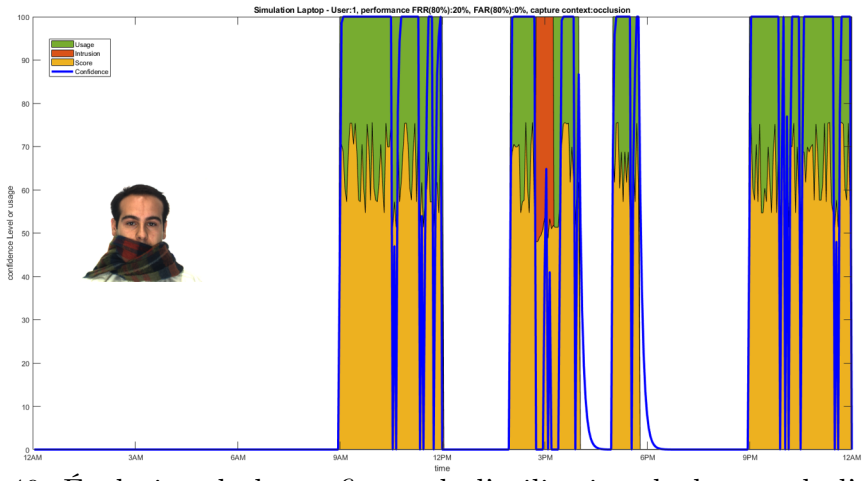

FIGURE 3.40: Évolution de la confiance de l'utilisation du laptop de l'utilisateur 1 au cours d'une journée en considérant le contexte du changement d'occultations avec protection de données (clé révélée).

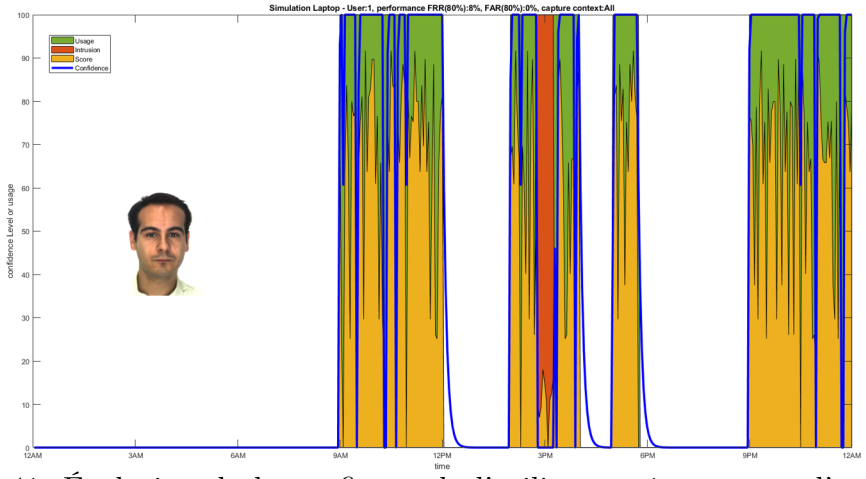

FIGURE 3.41: Évolution de la confiance de l'utilisateur 1 au cours d'une journée dans un contexte de tous les changements (sans protection).

— **Quel est l'impact de σ sur l'évolution de la confiance ?**

La valeur σ permet de convertir un score en une confiance de l'authentification ponctuelle. Le choix de σ dans ce travail a été fait d'une manière empirique en testant différentes valeurs pour décider sur la meilleure valeur qui donne les meilleures performances. Nous montrons l'impact de différentes valeurs de σ sur l'évolution de la confiance pour un contexte donné, soit le contexte de changement d'expressions, à un seuil fixe (seuil=80%). La figure 3.44 montre les courbes de l'évolution de la confiance de l'utilisateur 1 au cours de la journée en considérant le contexte du changement d'expressions sans protection de données avec σ_1=0.5, σ_2=0.8 et σ_3=0.2. Il est clair sur la figure 3.44 que la

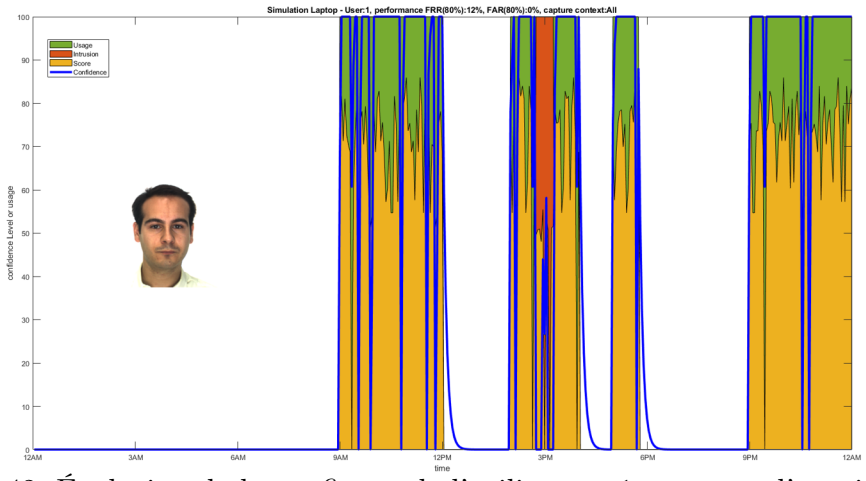

FIGURE 3.42: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant tous les changements avec protection de données (clé secrète).

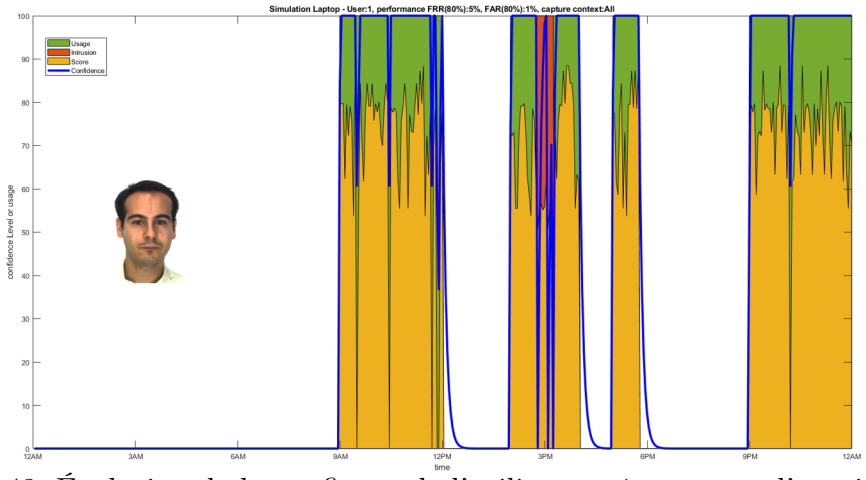

FIGURE 3.43: Évolution de la confiance de l'utilisateur 1 au cours d'une journée en considérant tous les changements avec protection de données (clé révélée).

valeur de σ a un impact important sur l'évolution de la confiance. Par exemple, pour σ_2=0.8, nous remarquons que la courbe de confiance est à 100% quand il s'agit d'une tentative d'intrusion. La confiance est mal générée quand σ_2=0.2, on voit sur la courbe qu'elle est à moins de 50% quand il s'agit bien d'une authentification légitime ou parfois à 0%. La valeur de σ dépend de l'évolution des scores, il pourrait être fixé en analysant sa distribution d'évolution entre les scores légitimes et d'imposture, ceci constitue une perspective de ce travail.

Les résultats obtenus sont très bons même dans des conditions extrêmes (personne ne masquera une partie de son visage dans un contexte d'authentification). Nous allons regarder dans la section suivante si l'approche conserve cette efficacité dans un contexte de données biométriques comportementales.

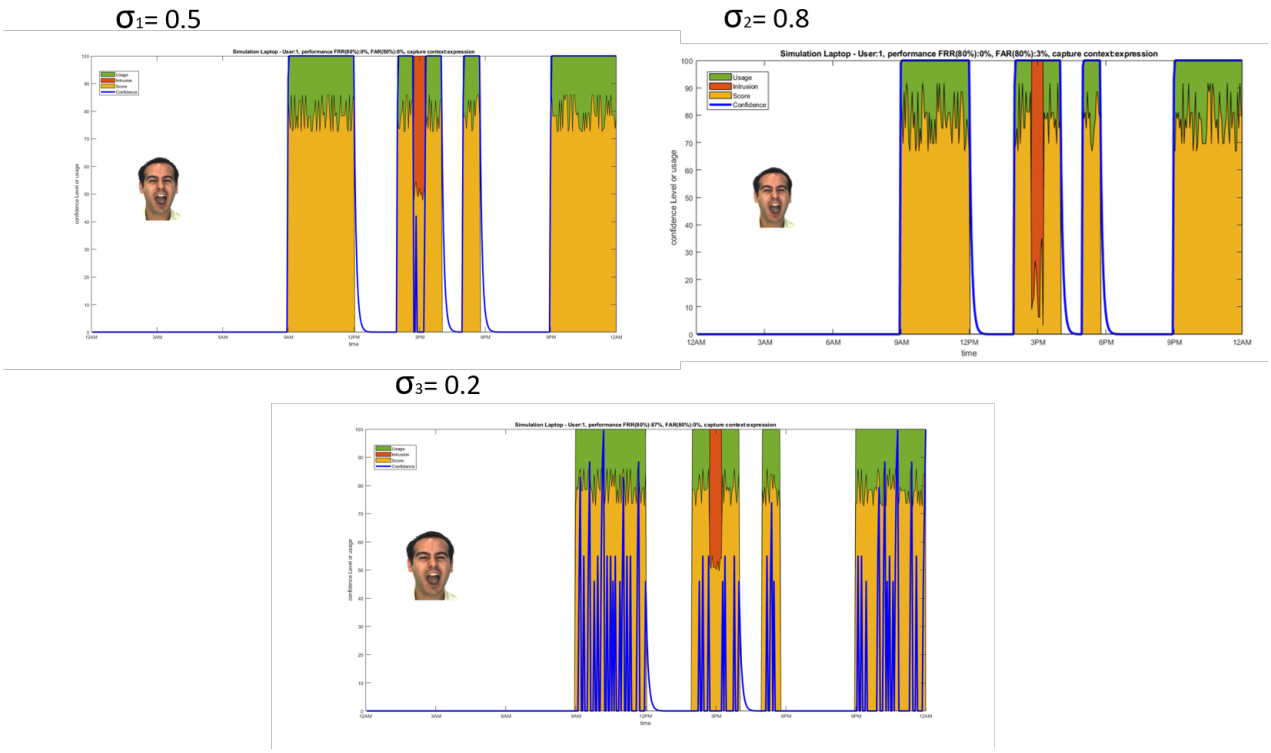

FIGURE 3.44: Impact des valeurs de σ sur l'évolution de la confiance de l'utilisateur 1 au cours de la journée en considérant le contexte du changement d'expressions sans protection de données.

Évaluation de l'authentification transparente sur smartphone

Dans cette partie, nous étudions l'évolution de la confiance d'un individu lors de l'usage du smartphone au cours d'une journée. Nous utilisons pour cela les données d'habitudes d'appels. Nous définissons NB le nombre d'échantillons utilisés pour construire la base de référence. On tire aléatoirement un échantillon de test parmi tous les échantillons hors la base de référence. Dans un premier temps, nous générons les courbes de confiance avec un σ fixé à 0.12 pour les données brutes et un σ fixé à 0.25 pour les données protégées. Le choix de σ est empirique, et ces 2 valeurs choisies donnent les meilleurs performances de la confiance. Ensuite, nous étudions l'impact de la valeur de σ et l'impact de NB sur l'évolution de la confiance.

La figure 3.45 montre l'évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours d'une journée. Nous constatons que la confiance est à 100% quand il s'agit d'un usage légitime et bien à 0% lors de la tentative d'intrusion entre 17h30 et 17h55. Nous montrons également les courbes de l'évolution de confiance au cours d'une journée de l'utilisateur 30 et l'utilisateur 80, données respectivement par les figure 3.46 et 3.47.

Sans protection de données, nous obtenons un FRR(80%) = 16% et un FAR(80%)

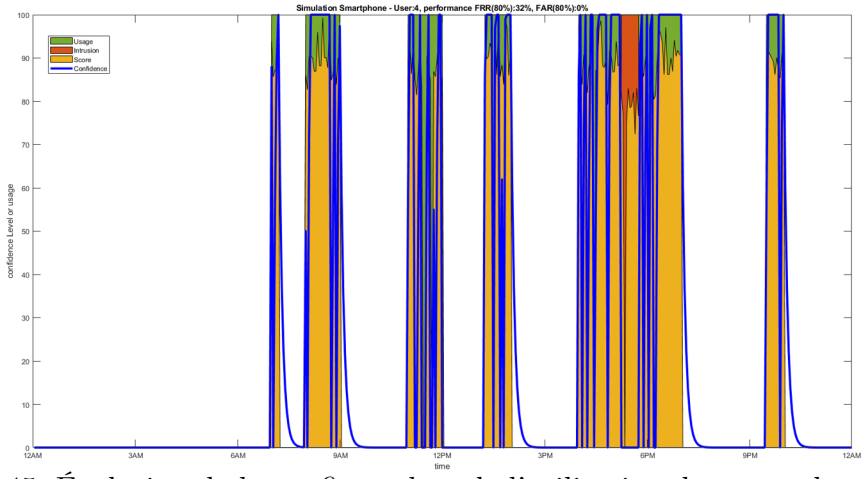

FIGURE 3.45: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours d'une journée, NB=5, σ =0.12 (sans protection).

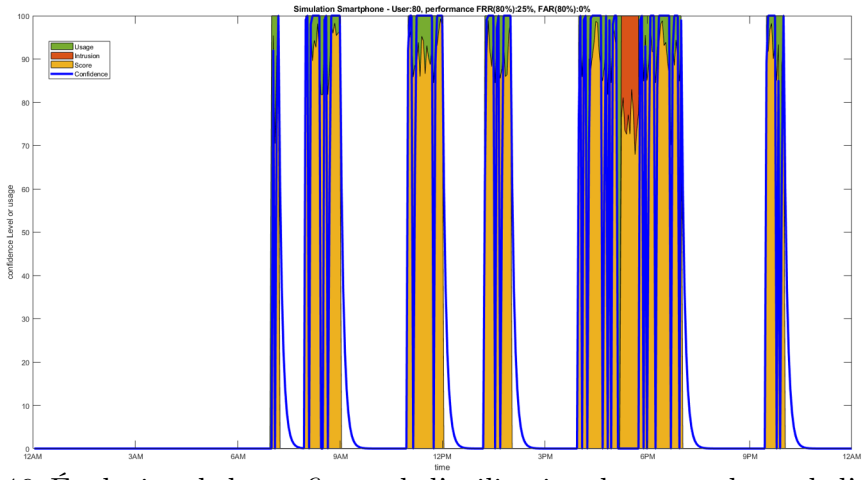

FIGURE 3.46: Évolution de la confiance de l'utilisation du smartphone de l'utilisateur 30 au cours d'une journée, NB=5, σ =0.12 (sans protection).

= 0% pour l'utilisateur 4 et un FRR(80%) =25% et un FAR(80%) = 0% pour l'utilisateur 30 et 80. Nous constatons sur la figure 3.48 que l'évolution de la courbe de confiance est améliorée avec les données protégées à clé secrète. Nous avons également étudié l'authentification transparente sur smartphone avec les données protégées.

Les figures 3.48 et 3.49 montrent respectivement l'évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours d'une journée avec protection de données lorsque la clé est secrète et révélée. Nous obtenons un FRR(80%) = 16% et un FAR(80%) = 0% à clé secrète, et un FRR(80%) = 18% et un FAR(80%) = 0% à clé révélée.

Les figures 3.50 et 3.51 montrent respectivement l'évolution de la confiance de l'uti-

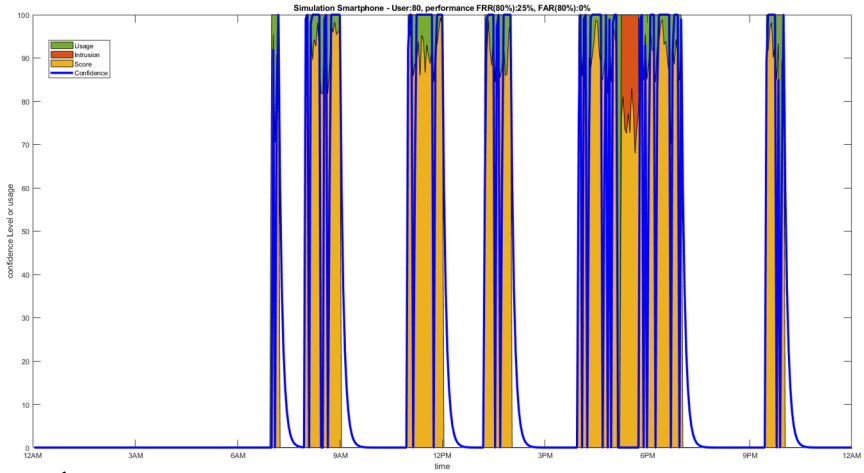

FIGURE 3.47: Évolution de la confiance de l'utilisation du smartphone de l'utilisateur 80 au cours d'une journée, NB=5, σ =0.12 (sans protection).

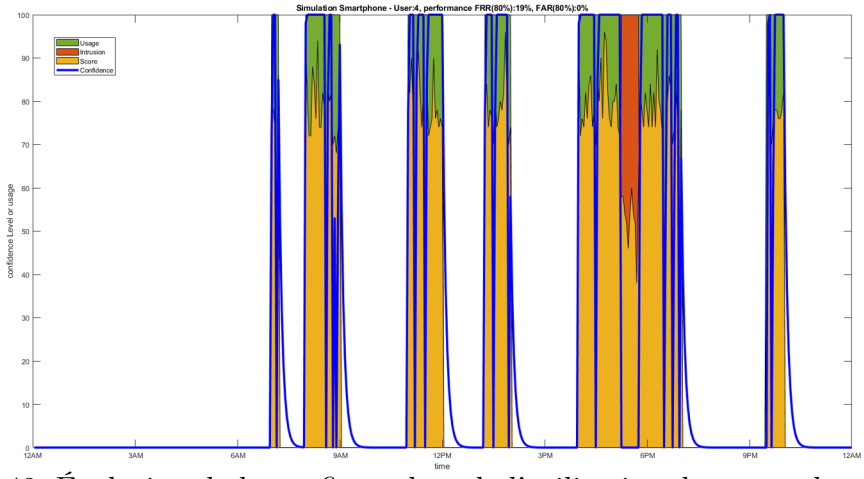

FIGURE 3.48: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours d'une journée avec protection de données (clé secrète), NB=5, σ =0.25

lisateur 30 et 80 au cours de la journée avec protection de données à clé secrète. Nous obtenons un FRR(80%) = 2% et un FAR(80%) = 0% pour les 2 utilisateurs. Les figures 3.52 et 3.53 montrent respectivement l'évolution de la confiance de l'utilisateur 30 et 80 au cours de la journée avec protection de données à clé révélée. Nous obtenons un FRR(80%) = 19% et un FAR(80%) = 0% pour les 2 utilisateurs.

— **Quel est l'impact de σ sur l'évolution de la confiance sur smartphone ?**

Dans l'étude précédente de l'authentification transparente basée sur la reconnaissance faciale, nous avons montré l'impact important de σ sur l'évolution de la confiance des utilisateurs. De la même façon, nous souhaitons étudier l'im-

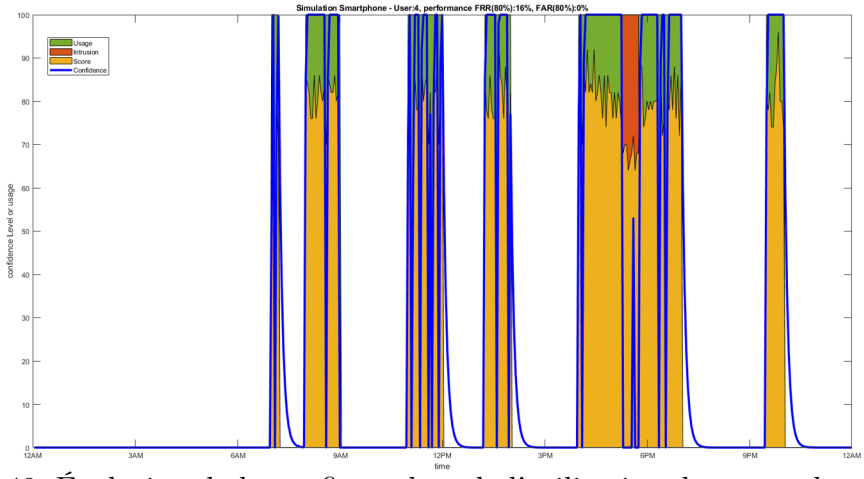

FIGURE 3.49: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours d'une journée avec protection de données (clé révélée), NB=5, σ =0.25

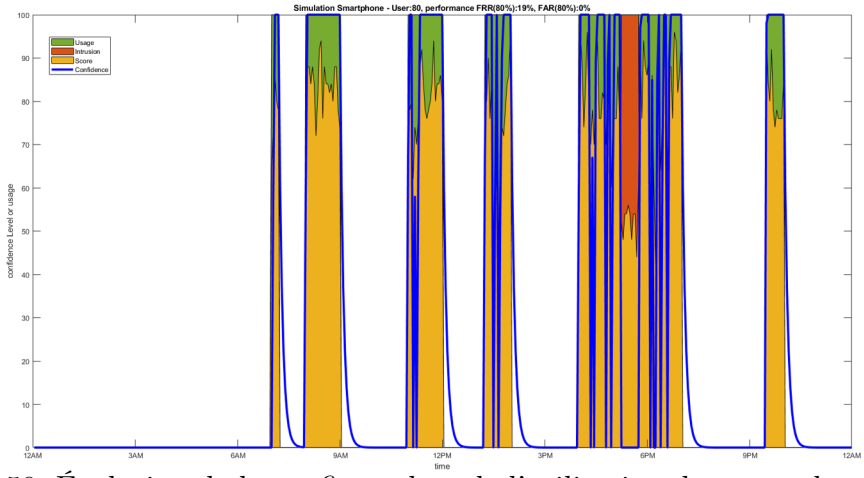

FIGURE 3.50: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 30 au cours d'une journée avec protection de données à clé secrète, NB=5, σ =0.25

pact de σ sur l'évolution de la confiance lors de l'usage du smartphone pour un nombre d'échantillons NB fixé (NB=5, soit 5 échantillons dans la référence). La figure 3.54 montre les courbes d'évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours de la journée sans protection de données avec σ_1=0.12, σ_2=0.3 et σ_3=0.5. Quand σ_2=0.3 et σ_2=0.5, nous remarquons que la confiance est impactée par rapport à celle quand σ_1=0.12. La confiance ne se met pas à 0% par exemple quand il s'agit d'une attaque. Nous retrouvons cette dépendance à la distribution des scores déjà mentionnée précédemment.

— **Quel est l'impact de NB sur l'évolution de la confiance ?**

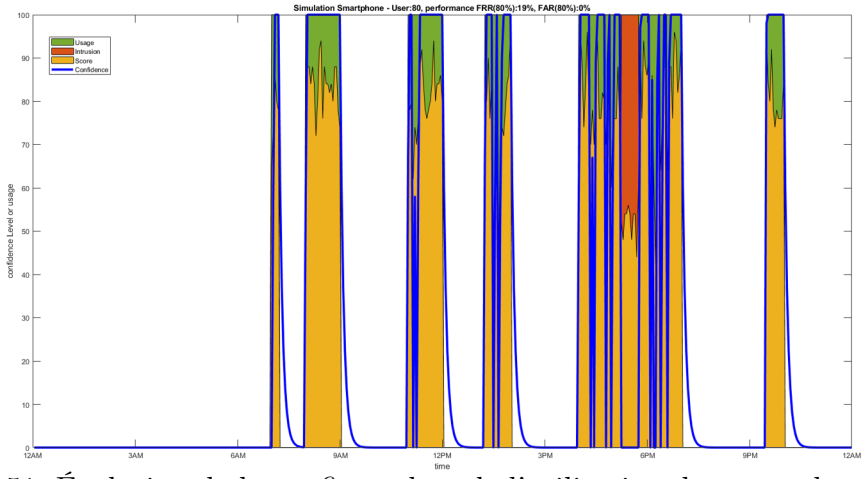

FIGURE 3.51: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 80 au cours d'une journée avec protection de données à clé secrète, NB=5, σ =0.25

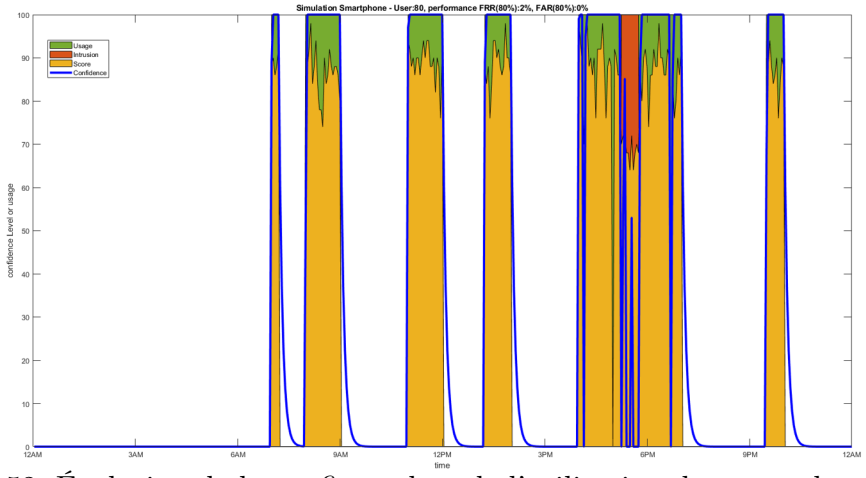

FIGURE 3.52: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 30 au cours d'une journée avec protection de données à clé révélée, NB=5, σ =0.25

Nous souhaitons étudier l'impact de NB sur l'évolution de la confiance lors de l'usage du smartphone pour un σ fixé (avec σ=0.12 sans protection et σ=0.25 avec protection). La figure 3.55 montre respectivement l'évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours de la journée sans protection de données avec σ_1=0.12, pour NB=5, NB=3 et NB=8. Nous remarquons que le choix de la valeur de NB n'a pas un impact significatif sur l'évolution de la confiance. Pour toutes les valeurs testées, la confiance est générée correctement.

Nous avons obtenu des bons résultats avec les données des habitudes d'appels assurant une authentification transparente sur le smartphone. La valeur de σ a un

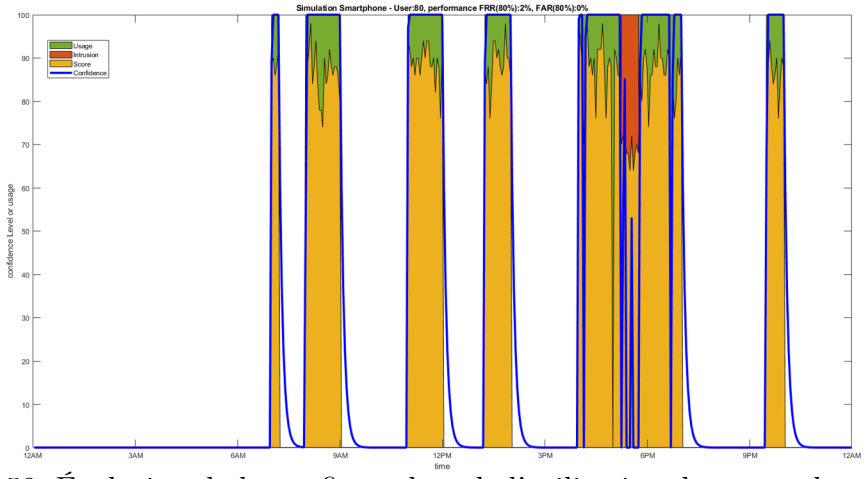

FIGURE 3.53: Évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 80 au cours d'une journée avec protection de données à clé révélée, NB=5, σ =0.25

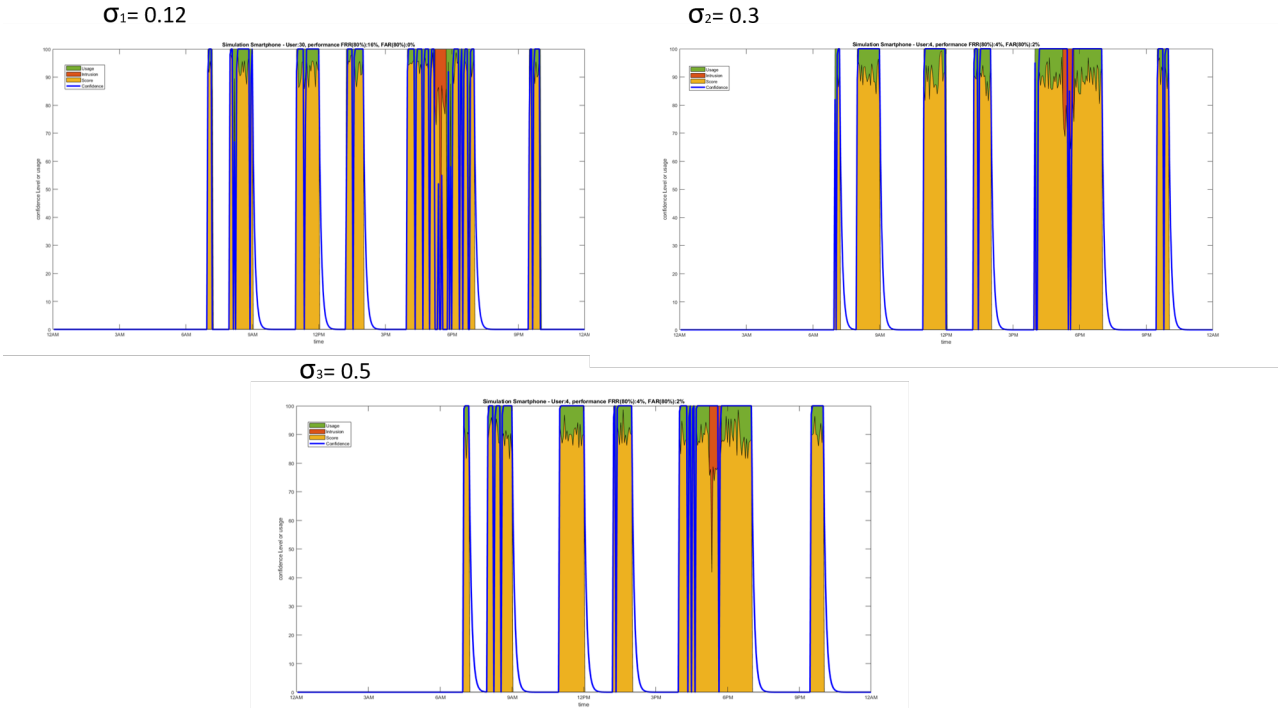

FIGURE 3.54: Impact des valeurs de σ sur l'évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours de la journée sans protection de données, NB=5.

impact important sur l'évolution de la confiance de l'usage du smartphone. Pour NB=5, nous avons fixé la valeur de σ à 0.12 pour les données brutes et σ =0.25 pour les données protégées.

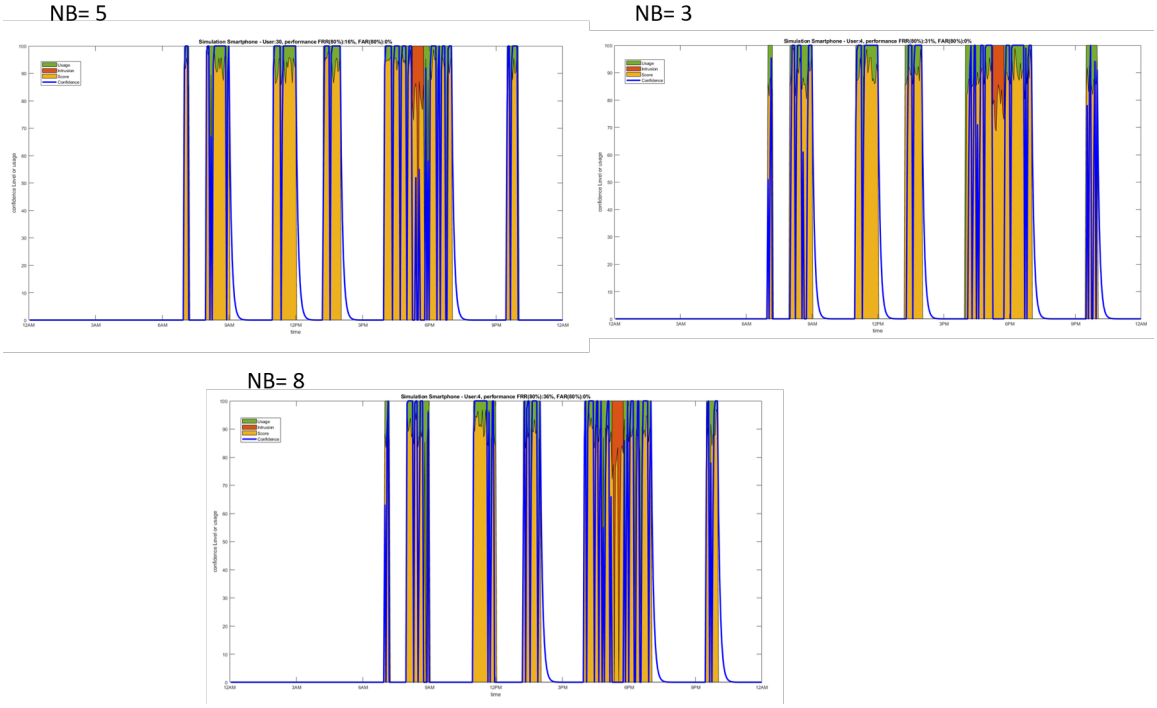

FIGURE 3.55: Impact des valeurs de NB sur l'évolution de la confiance lors de l'utilisation du smartphone de l'utilisateur 4 au cours de la journée sans protection de données, σ =0.12

3.5.3 Temps de calcul

Afin de calculer la confiance de l'usage d'un objet intelligent et de générer les courbes, nous avons utilisé le logiciel Matlab pour l'implémentation. Les calculs ont été réalisés sur un ordinateur avec un processeur Intel(R) Core(TM) i5-6300U CPU @2.50GHz. Le temps d'exécution est de 60 secondes pour le calcul de confiance sur smartphone et de 12 minutes pour le calcul de confiance sur ordinateur avec les données de visage pour simuler une journée complète. Dans un contexte opérationnel, la génération du BioCode (facial ou sur les habitudes d'appel) doit être réalisée sur le client et l'évolution de la confiance par le tiers de confiance à distance.

3.5.4 Analyse des propriétés

Nous avons défini dans la partie 2.2.4 les exigences d'un système d'authentification transparente, qui sont : l'usabilité, la sécurité et la vie privée. Dans cette partie, nous montrons à quel point nos solutions d'authentification transparente sur smartphone et sur ordinateur respectent ces propriétés :

— **Vie privée**
Notre approche prend en considération le respect de la vie privée des utilisateurs dans sa conception. Toutes les données personnelles et sensibles utilisées dans notre système sont protégées par l'algorithme de BioHashing et la clé

secrète de l'utilisateur. Nous avons montré dans la partie de résultats expérimentaux, que les performances du système sont meilleures avec protection des données à clé secrète, ce qui empêche un imposteur en cas d'attaque, de récupérer les données brutes. Le tiers de confiance est en mesure de mettre à jour la confiance dans l'identité de l'individu utilisant l'objet intelligent sans connaître le contenu des données utilisées. Le tiers de confiance peut être vu comme un fournisseur d'identité pouvant fournir des pseudo-anonymes à des services numériques pour éviter de lier les différentes identités de l'utilisateur. Enfin, cette solution évite que chaque service utilisé n'ait à stocker des données personnelles de l'individu pouvant potentiellement atteinte à la vie privée de la personne en cas de compromission du service.

— **Sécurité**
L'authentification est la première barrière contre l'usurpation d'identité. Une fois une identité numérique usurpée, l'attaquant peut accéder à des données personnelles de l'utilisateur. La sécurité de système est donc indispensable. Nous avons conçu une solution d'authentification transparente sécurisée et robuste aux attaques. Nous avons étudié la fonction de décroissance de la confiance au cours du temps pour ne pas laisser des failles aux attaques lors de l'absence d'utilisation de l'objet intelligent par la personne légitime. Nous avons également étudié des scénarios d'attaques et nous avons montré que la confiance du système est nulle quand une tentative d'intrusion est détectée. Néanmoins, le système que nous avons proposé reste sensible aux attaques par rejeu. Une attaque par rejeu est une attaque réseau dans laquelle une transmission est malicieusement répétée par un attaquant qui a intercepté la transmission. Afin d'éviter ce type d'attaque, il est possible de chiffrer le canal de communication entre le client et le tiers de confiance (en SSL). Il est également possible d'utiliser une clé dynamique (basée sur un mot de passe à usage unique) qui change après chaque utilisation et qui est générée aléatoirement. Dans ce cas, l'attaquant ne peut pas utiliser dans une nouvelle session les informations de l'utilisateur légitime d'une session précédente parce que sa clé de session est différente.

— **Usabilité**
La solution proposée demande moins d'intervention explicite par l'utilisateur pour prouver son identité au système, ce qui le rend facile à utiliser. En plus, la collecte de données pour entraîner le système est faite d'une manière simple et surtout implicite, lui permettant d'être efficace et plaisant à utiliser. Nous avons obtenu des résultats de performances intéressants avec des faibles valeurs d'EER, traduisant son efficacité d'utilisation. Le faux rejet qui peut traduire une gêne utilisateur est très faible au vu des expériences. Le bémol de cette solution est qu'aucune confiance d'un objet intelligent n'est transféré à un autre objet (par exemple du smartphone à l'ordinateur portable) occasionnant des besoins d'authentification statique. Le chapitre 4 tente de répondre à cette problématique.

3.6 Conclusion

Nous avons présenté dans ce chapitre une approche d'authentification transparente via un unique objet connecté. Nous avons évalué la performance de cette approche pour deux modalités biométriques. La première est basée sur les données d'habitudes d'appels issues d'un smartphone et la seconde utilise des images du visage de l'individu potentiellement acquise par une webcam d'un ordinateur.

Dans un premier temps, après la collecte et le traitement de données, nous avons appliqué des algorithmes d'apprentissage machine qui sont, la minimisation de distance, K-means et SVM à classe unique, pour évaluer la performance de la chaîne de traitement dans un contexte d'authentification statique (classique). Dans un second temps, la protection des données par l'algorithme BioHashing est prise en compte pour évaluer les performances du système en matière de protection des données.

Les résultats expérimentaux obtenus sont convaincants. Pour les données du smartphone, nous avons obtenu une valeur du EER de 10% pour les SVM à classe unique sans protection des données, ce qui en fait le meilleur classificateur pour une meilleure authentification par rapport aux autres classifieurs adoptés. La méthode K-means donne une valeur EER de 37%, ce qui est considéré comme élevé par rapport aux techniques de minimisation de la distance et du SVM. Toutefois, lors de la mise en œuvre de l'algorithme BioHashing pour la protection de la vie privée des données, les valeurs EER pour les 3 méthodes sont inférieures à 1% et le classifieur SVM à classe unique reste le meilleur avec une valeur EER de 0,1%. Pour les données faciales, nous avons estimé que l'application de la méthode de minimisation de distance pour étudier les performances de données était suffisante car une seule image a été utilisée comme référence (enrôlement unique). Nous avons obtenu une valeur d'EER égale à 13% pour les données brutes. Les performances s'améliorent en appliquant l'algorithme de BioHashing avec clé secrète avec une valeur d'EER égale à 0,02%. Dans le cas où clé était révélée, nous avons obtenu une valeur d'EER égale à 0,09%. De grandes améliorations sont observées lors de l'utilisation de l'algorithme BioHashing avec les données d'habitudes d'appels sur smartphone et les données faciales sur PC, ce qui s'avère une approche intéressante pour une authentification robuste des utilisateurs qui respecte leur vie privée.

Dans une seconde étape, nous avons évalué la confiance de l'usage du smartphone et de l'ordinateur au cours de la journée dans un contexte d'authentification transparente. Nous avons défini la fonction de calcul de confiance en fonction d'un score d'authentification généré à partir des données collectées sur l'individu, et à l'aide d'une loi normale centrée et paramétrée avec l'écart-type σ. Les résultats obtenus montrent l'efficacité de l'approche proposée en assurant une authentification transparente de l'utilisateur. La confiance est maximale quand il s'agit d'un usage légitime de l'objet intelligent et elle est nulle en cas d'intrusion, avec des faibles valeurs du FAR et du FRR associées à un seuil fixe (égale à 0% dans la plupart des cas). Nous avons opté pour un choix empirique de la valeur de σ et nous avons fait une

étude sur l'impact de cette valeur sur l'évolution de la confiance. Il s'est avéré que σ affecte l'évolution de la confiance si sa valeur est mal choisie. Dans un scénario plus précis, il faudrait générer les distributions de tous les scores légitimes et d'impostures de la base de données, pour pouvoir identifier la valeur exacte et précise de σ donnant les meilleures performances. Nous avons également analysé l'impact du nombre d'échantillons NB sur l'évolution de la confiance de l'usage du smartphone et nous avons déduis que NB n'affecte pas l'évolution de la confiance d'une manière significative. Pour les données faciales sur ordinateur, nous avons étudié l'impact des changements d'expressions, de lumière et d'occultations sur l'évolution de la confiance. Notre approche est robuste face aux différents changements, et la reconnaissance faciale est assurée correctement pour la génération de la confiance.

Maintenant, après avoir évalué et validé l'évolution de la confiance de l'usage d'un objet intelligent dans un contexte d'authentification transparente, une question légitime se pose : comment peut-on transférer la confiance d'un objet intelligent à un autre ? En effet, dans un environnement numérique ubiquitaire, l'authentification via plusieurs objets intelligents devient la solution pour une authentification transparente de l'utilisateur, plus sûre et mieux protégée. Dans une telle solution, les objets intelligents d'un même utilisateur doivent interagir mutuellement afin de garantir un score de confiance élevé, suffisant pour pouvoir communiquer entre eux la confiance d'un objet intelligent.

Pour cette raison, les résultats obtenus avec l'ensemble de données personnelles d'habitudes d'appels sur smartphone et les données faciales sur ordinateur, figurant dans le présent chapitre seront exploités dans une approche d'authentification transparente via plusieurs objets connectés dans le chapitre suivant. Le but est de garantir une authentification plus robuste en terme de sécurité, d'usabilité et de vie privée, dans un environnement numérique ubiquitaire.

Chapitre 4

Authentification transparente via plusieurs objets intelligents

Resume: *Ce chapitre présente un nouveau concept d'authentification transparente via plusieurs objets connectés, qu'on appelle "Aura d'authentification" où la confiance est transférée d'un objet intelligent à un autre, pour faciliter l'authentification de l'utilisateur en toute sécurité et dans le respect de sa vie privée. Un état de l'art est élaboré pour définir et positionner notre solution dans la littérature. Une formulation détaillée du concept est proposée. La validation des résultats est réalisée sur des données réelles issues d'un smartphone, un ordinateur portable et un ordinateur fixe.*

United we stand, devided we fall

4.1 Introduction

La recherche de solutions innovantes pour l'authentification des utilisateurs est très active. Cependant, bien que les gens soient quotidiennement entourés de différents et multiples objets connectés, ils utilisent toujours un seul et même appareil pour accéder à leurs équipements électroniques ou pour les sécuriser. Malgré l'augmentation significative des objets connectés et de l'Internet des objets (IoT), peu de gens s'intéressent à l'expérience utilisateur prise dans la totalité de ses interactions numériques. Si un appareil est compromis par des attaquants, la sécurité n'est plus garantie et les attaquants peuvent facilement accéder aux données personnelles de l'utilisateur. De plus, la tâche d'authentification nécessite des interventions répétitives de la part de l'utilisateur car il doit agir avec différents appareils dans son voisinage afin de prouver son identité à chacun. Avec la multiplicité des facteurs d'authentification et la diversité des terminaux possédés, cette action devient pénible et gênante, crée du stress, fait perdre du temps et encombre notre vie quotidienne

de tâches inutiles.

Dans l'intention de se débarrasser du superflu au quotidien et de garantir une meilleure sécurité et une meilleure protection de la vie privée, nous proposons de prendre en compte les multiples appareils de l'utilisateur dans le processus d'authentification et de déléguer la tâche d'authentification à chacun d'entre eux. Ainsi, en cas de compromission d'appareil, tous les appareils créent ensemble un solide cercle de confiance afin de protéger les données personnelles de l'utilisateur. Par la suite, ce cercle multi-devices est appelé Aura d'Authentification. Ce concept n'est pas nouveau puisqu'il a été proposé dans [108], nous proposons dans cette thèse d'étendre certaines notions en mettant particulièrement l'accent sur la vie privée. À cet égard, dans un environnement numérique ubiquitaire, nous proposons de fournir une authentification transparente à l'utilisateur tout au long de sa journée tout en assurant une meilleure protection de la vie privée.

4.2 Notions préliminaires

Tout d'abord, nous présentons dans cette partie, une définition de l'Aura humaine et de l'Aura numérique pour introduire par la suite l'Aura d'authentification. Dans la mesure où dans ce travail nous considérons plusieurs objets intelligents de l'utilisateur, nous donnons également un aperçu sur les technologies existantes permettant d'évaluer la proximité entre deux objets connectés.

Aura humaine

Dans plusieurs traditions mystiques, l'aura est un concept ésotérique qui désigne un contour coloré, comme un « halo de lumière » qui rayonnerait autour du corps ou de la tête d'un être vivant et serait la manifestation d'une force vitale créant des vibrations entourant le corps physique de différentes couleurs, voir Figure 4.1. En littérature, le mot est utilisé dans un sens métaphorique, dérivé du sens mystique. Il désigne l'atmosphère qui entoure ou semble entourer une personnalité qui s'impose fortement à l'attention d'autrui par sa présence, une œuvre qui marque son époque d'un rayonnement particulier.

Aura Numérique

La présence permanente de terminaux et d'objets connectés dans l'environnement proche de l'individu et en interaction avec celui-ci, il est possible par analogie de définir une véritable aura numérique, traduisant la communication entre l'utilisateur et ses devices numériques présentée par la figure 4.2. Dans cette aura numérique, l'authentification de l'utilisateur représente un aspect essentiel dans l'interconnexion des objets connectés. L'interaction entre l'utilisateur et ses devices devrait être sécurisée. D'une part, l'utilisateur vérifie si le device qu'il souhaite utiliser est bien le sien et d'autre part, le device lui aussi, doit vérifier si l'utilisateur qui demande l'accès

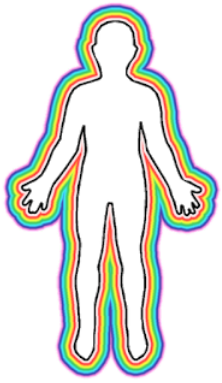

FIGURE 4.1: Aura Humaine

à un service numérique est bien le propriétaire légitime de ce device. À ce stade, la tâche d'authentification a lieu au moment où, après son identification auprès de son device, l'utilisateur doit donner une preuve, généralement un code secret ou un mot de passe, afin de s'authentifier correctement et autoriser son accès. Dans ce scénario, l'utilisateur devrait à chaque fois s'authentifier auprès de tous ses devices, en présentant un code différent pour chaque device, ce qui fait une dizaine voire une vingtaine de codes à apprendre et à taper une centaine de fois par jour, afin d'autoriser l'accès. La tâche d'authentification est répétitive et pénible et demande toujours une action explicite de l'utilisateur même au sein d'un même device (Authentification non continue). Afin de résoudre ce problème, nous introduisons le concept d'Aura d'authentification dans la section suivante qui tient compte de la proximité des objets connectés pour transférer la confiance entre eux et assurer une authentification transparente. Nous présentons par la suite les technologies existantes pour évaluer la proximité entre deux objets connectés.

La proximité entre deux objets connectés

Plusieurs technologies et de réseaux sans fil permettent la géo-localisation des appareils électroniques et la communication entre eux. Le choix de la méthode de communication appropriée est basé sur plusieurs facteurs, notamment la distance ou la portée de la connexion, la vitesse de connexion et le taux de consommation d'énergie de l'appareil. La figure 4.3 [109] montre la classification des technologies de communications entre les appareils IoT en fonction du débit de données et la consommation d'énergie des connexions par rapport à la distance de connexion.

RFID
La technologie RFID consiste à utiliser des radio-fréquences pour transférer des données afin d'identifier et de suivre automatiquement les étiquettes fixées aux objets. Un système RFID se compose de deux parties : une étiquette et un lecteur. La lecture d'un tag RFID permettra d'identifier une proximité avec un lecteur en général à 15cm mais il existe des tag RFID lisibles à des distances plus élevées (cas d'utili-

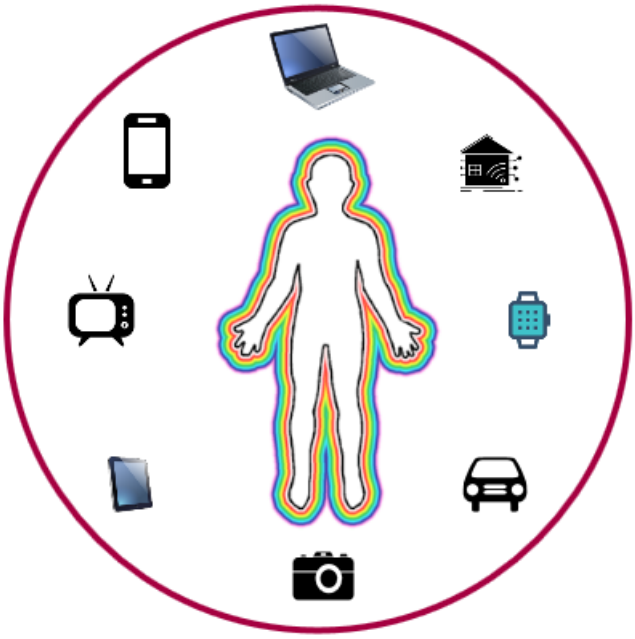

FIGURE 4.2: Aura Numérique

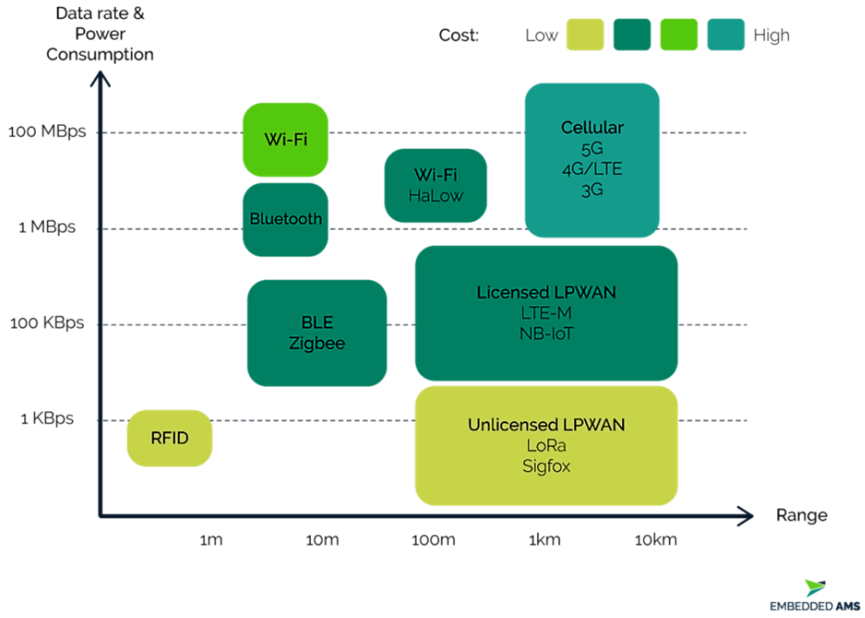

FIGURE 4.3: Le débit de données et la consommation d'énergie des connexions des appareils IoT par rapport à la distance de connexion [109]

sation en logisitique).

NFC
Le NFC est une technologie de communication sans contact de courte portée. Elle permet aux périphériques équipés de communiquer avec n'importe quel terminal mobile compatible en rapprochant simplement les deux supports. Le NFC est un dérivé de la technologie RFID et aujourd'hui elle équipe la plupart des smartphones. Cette technologie étend les modes de communication et peut être utilisée pour estimer une distance entre objets NFC.

Bluetooth
Le Bluetooth est une technologie sans fil utilisée pour transférer des données sur de courtes distances via les ondes radio UHF. Cette technologie est fréquemment utilisée dans des appareils de petite taille qui se connectent aux téléphones et aux tablettes des utilisateurs. Par exemple, la technologie est utilisée dans de nombreux systèmes de haut-parleurs, les traqueurs de fitness, les montres intelligentes et autres appareils connectés afin de transmettre des données sans fil sans compromettre fortement la puissance de la batterie du téléphone de l'utilisateur. La portée est limitée à quelques mètres seulement (10 à 15 mètres). Cela peut permettre par exemple d'identifier si deux objets sont dans la même pièce.

WiFi
Le WiFi utilise les ondes radio-fréquences pour permettre à deux appareils de communiquer entre eux. Cette technologie est le plus souvent utilisée pour connecter des routeurs Internet à des appareils tels que des ordinateurs, des tablettes et des téléphones ; elle peut cependant être utilisée pour relier entre eux deux composants matériels quelconques. Le WiFi est un réseau local sans fil qui fonctionne selon les normes 802.11 établies par l'Institute of Electrical and Electronics Engineers (IEEE). La portée est de plusieurs dizaines de mètres, voire de plusieurs centaines de mètres en extérieur.

ZigBee
ZigBee est une technologie sans fil conçue spécialement pour être utilisée dans les réseaux Machine to Machine (M2M). Cette technologie est peu coûteuse et ne nécessite pas beaucoup d'énergie, ce qui en fait une solution idéale pour de nombreuses applications industrielles. La technologie a une faible latence et un faible cycle d'utilisation, ce qui permet aux produits de maximiser la durée de vie des batteries (portée sur une surface de 10m^2).

Géolocalisation par GPS
Le GPS est un système de positionnement par satellite permettant de connaître sa position ou celle d'un objet ainsi que l'heure actuelle à tout instant, par tous les temps et n'importe où sur Terre ou dans son voisinage. La précision du système GPS pour le grand public est de l'ordre de 2 à 10 mètres.

Ces protocoles de communication peuvent permettre d'estimer la proximité entre deux objets intelligents. Dans la suite, nous utiliserons la localisation GPS et le

protocole wifi (appartenance à un hotspot). Dans la section suivante, nous nous intéressons aux solutions d'authentification multi-objets dans la littérature.

4.3 État de l'art

Au cours des dernières années, il a été montré que l'internet des objets (IoT) a le potentiel d'avoir un impact sur l'ensemble de la société en changeant divers secteurs mais aussi notre vie quotidienne. Malgré la croissance impressionnante du nombre d'objets connectés, comme l'indique la dernière étude de Juniper Research [110], qui estime que le nombre des objets connectés atteindra 83 milliards d'ici 2024, contre 35 milliards de connexions en 2020, peu de travaux de recherche focalisent sur les solutions d'authentification basées sur plusieurs objets connectés. L'authentification via plusieurs objets connectés, en particulier en définissant le concept Aura d'authentification, a été présentée pour la première fois par Hocking et al.[108] en 2011. Cette nouvelle approche de l'authentification d'identité sur des appareils mobiles est basée sur un framework qui peut améliorer de manière transparente la confiance de l'utilisateur en matière de sécurité. Les informations relatives à l'authentification de l'utilisateur sont partagées entre les objets intelligents de l'utilisateur, permettant ensemble d'établir et de maintenir un cercle de confiance autour de l'utilisateur. Cependant, nous pouvons souligner que la protection de la vie privée n'a pas été considérée dans ce travail, ce qui est un inconvénient important une fois que le Règlement Général sur la Protection des Données (RGPD) est maintenant en place pour les citoyens européens.

Le tableau 4.1 présente une revue des travaux connexes sur les solutions d'authentification via plusieurs objets connectés. Riva et al.[111] ont présenté une authentification progressive basée sur l'association de plusieurs sources de données d'authentification. Cha et al. [112] proposent un modèle d'authentification à deux objets connectés pour les systèmes de micro-paiement utilisant des appareils mobiles portables. Xu [113] se focalise sur l'authentification biométrique à l'aide de objets intelligents portables (wearables), notamment sur la reconnaissance du visage à l'aide de smart-glass et la reconnaissance de la démarche à l'aide d'une montre intelligente.

Toutefois, toutes ces approches manquent de détails sur la protection de la vie privée des utilisateurs. En outre, afin de contrôler et sécuriser l'accès aux services de stockage basés sur le cloud, Gonzalez et al. [114] proposent une solution multi-devices avec un schéma cryptographique symétrique. Cependant, il n'y a pas d'analyse concernant la protection de la vie privée et les détails de sécurité discutés. Hajny et al. [115] ont également présenté un schéma cryptographique assurant une authentification multi-device utilisant les objets connectés portables. Move2auth, une solution basée sur la proximité entre les objets connectés assurant leur authentification, a été présenté par Zhang et al. [116]. Il repose sur le mouvement de gestes de la main (rapprochement, éloignement et rotation) pour détecter la proximité et authentifier les objets connectés. Néanmoins, les résultats présentés sont limités à

un seul appareil,qui est le smartphone, et il n'y a pas d'analyse plus avancée sur d'autres objets connectés, et la protection de la vie privée n'est pas prise en compte.

Sur la base de cet aperçu de la littérature, nous pouvons donc admettre que les solutions d'authentification via plusieurs objets connectés sont à la fois limitées en nombre et considèrent peu la protection de la vie privée. Les solutions existantes ne se focalisent pas à la fois sur la facilité d'utilisation, la sécurité et le respect de la vie privée. Dans la section suivante, nous proposons une nouvelle solution étendant le concept Aura par [108]. La solution proposée définit un nouveau service de tiers de confiance pour les utilisateurs d'Internet, respectant les exigences du RGPD et mettant l'accent sur la facilité d'utilisation. Nous décrivons le concept et toutes les étapes de son utilisation.

4.4 Méthode proposée

Nous présentons dans cette partie une formulation détaillée du concept d'Aura d'authentification et nous décrivons le procédé de transfert de confiance entre les objets connectés.

4.4.1 Aura d'authentification : concept et formulation

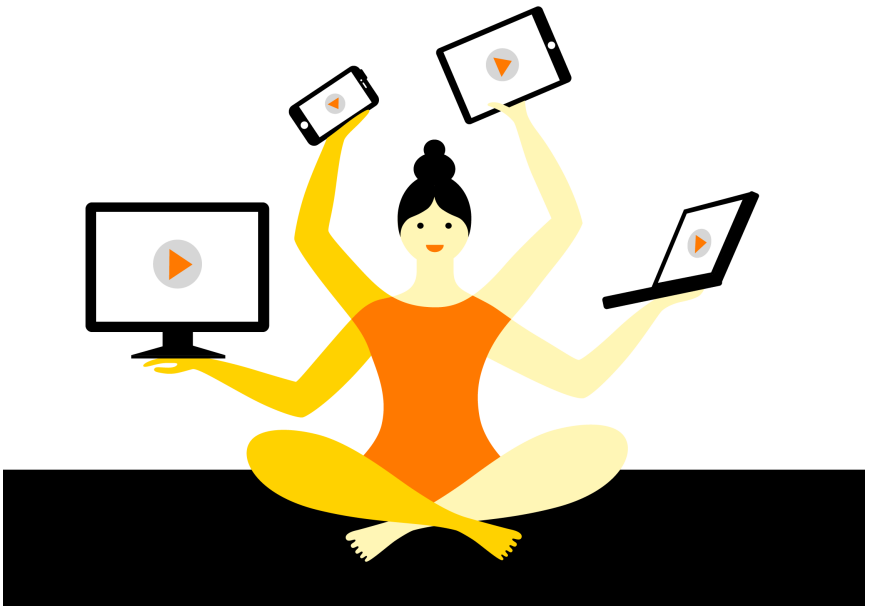

FIGURE 4.4: Aura Numérique

Notre approche consiste à créer une procédure d'authentification multi-devices basée sur une communication mutuelle entre les objets connectés possédés par un utilisateur, grâce au service du tiers de confiance (voir figure 4.5). Il s'agit de définir une architecture de confiance entre différents terminaux et objets connectés

TABLE 4.1: Aperçu des Travaux connexes sur l'authentification via plusieurs objets connectés

References	**Multidevice solution**	**Modalities**	**IoT devices**	**Privacy protection**
Hocking et al.[108]	Authentication Aura	Face recognition, keystroke, tokens	Smartphone, Laptop	No
Riva et al.[111]	Progressive Authentication	Biometrics, Possessions, PINS	Smartphone,tablets	No
Cha et al.[112]	Micro Payment System based multidevice authentication	Fingerprint,Password,OTP	Mobile and Wearable devices	No
Xu [113]	Biometric Authentication	Face recognition, gait recognition	Smart glass, Smart watch	No
Xu[113]	Biometric Authentication	Face recognition, gait recognition	Smart glass, Smart watch	No
Gonzalez et al. [114]	Storage services access control in miltidevice authentication	Symetric cryptography	smartphone,laptop	No
Hajny et al. [115]	Multidevice cryptographic scheme	N/A	Wearable devices	No
Zhang et al. [116]	Move2auth : proximity based authentication	hand gestures	Smartphone, connected IoT devices	No

capables de collecter en continu des données comportementales ou morphologiques sur le propriétaire des objets intelligents. Ces données ensuite agrégées et protégées doivent permettre l'authentification en continu de l'utilisateur. En d'autres termes, nous souhaitons assurer des interactions suffisantes entre un utilisateur et ses appareils pour garantir un niveau de confiance élevé qui pourrait être transféré entre eux. Dans la suite, nous considérons les termes suivants : Appareil électronique, objet connecté et device comme des synonymes ayant la même définition pour décrire les terminaux d'un utilisateur.

Considérons les définitions suivantes :

— A : l'Aura d'authentification de l'utilisateur U,

— O_i : un objet connecté $\in A$, avec i $\in [1, n_d]$, n_d est le nombre des objets connectés de l'utilisateur U,

— $C(O_i)$: Confiance dans un objet connecté O_i. Elle est calculée à tout moment avec une solution d'authentification transparente basée sur des authentifications ponctuelles basées sur de nombreux facteurs (mots de passe, biométrie, géolocalisation).

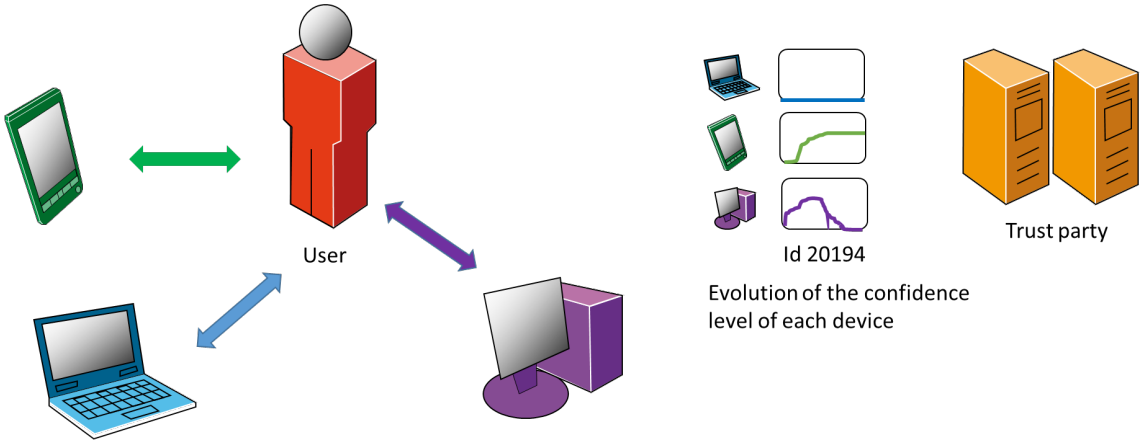

FIGURE 4.5: Principe de la méthode proposée : Le tiers de confiance réalise la mis à jour du niveau de confiance de chaque objet connecté à tout moment.

Nous considérons que chaque objet connecté a sa propre Aura, que nous appelons Aura Objet AO_i, tel que : A= $\bigcap_{i=1}^{n_d} AO_i$, en se basant sur ses différentes composantes. L'intersection entre 2 Auras d'objets AO_i et AO_j traduit la communication entre eux et peut être représentée par $P_{i/j} = AO_i \bigcap AO_j$.

Une Aura Objet peut-être constituée par :

— les capteurs : Accéléromètre, Gyroscope, Caméra, Microphone...

— Les technologies à base de signaux RF (radiofréquence) : WIFI, Bluetooth, NFC, RFID, Réseau Téléphonique ;

— Les satellites : GPS ;

— L'utilisateur propriétaire du device ;

— les données et les habitudes de l'utilisateur.

Afin de pouvoir déterminer l'intersection $P_{i/j}$ entre les auras d'objets et leurs confiances par rapport à l'Aura d'authentification A, on se base sur une analogie avec le fonctionnement du système GPS. D'abord, commençons par expliquer le principe de fonctionnement du GPS. Le système GPS permet de se situer où qu'on soit dans le monde. Il fonctionne avec une constellation de 30 satellites en orbite autour de la Terre. Chaque satellite envoie sur Terre des signaux qui contiennent :

— la position dans l'espace du satellite ;

— l'heure et la date d'émission du signal.

La puce GPS, qu'elle soit dans un smartphone ou un boîtier GPS, se contente de capter ces signaux. Quand un device reçoit les signaux de minimum 3 satellites, il est alors en mesure de calculer sa propre latitude, longitude et altitude et donc de vous dire où vous êtes. Maintenant supposons de se placer dans le plan et pas dans l'espace. Imaginons que le capteur GPS de ton smartphone reçoive le signal d'un premier satellite. Il connaît la date d'émission et de réception du signal, il connaît donc la durée du parcours du signal. Le signal voyage à la vitesse de la lumière, on en déduit qu'on se trouve à une distance d du satellite, autrement dit, sur un cercle centré sur le satellite comme l'indique la figure 4.6.

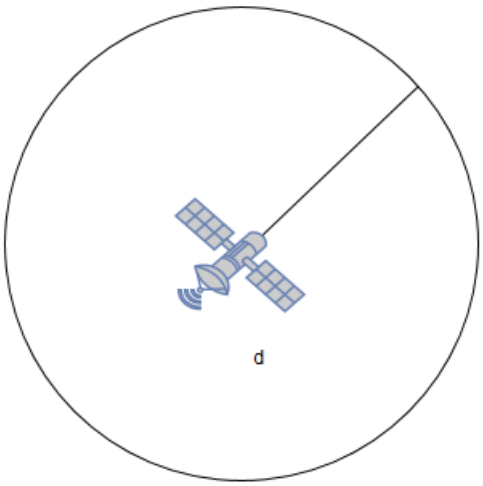

FIGURE 4.6: Signal émis par un satellite

Maintenant, ajoutons un 2ème signal provenant d'un 2ème satellite. On sait qu'on se trouve en même temps sur les 2 cercles, sur l'un des points où les cercles se coupent (voir figure 4.7).

Pour savoir lequel des 2 points est localisé l'objet, il nous faut un 3ème satellite donc 3ème cercle (voir figure 4.8).

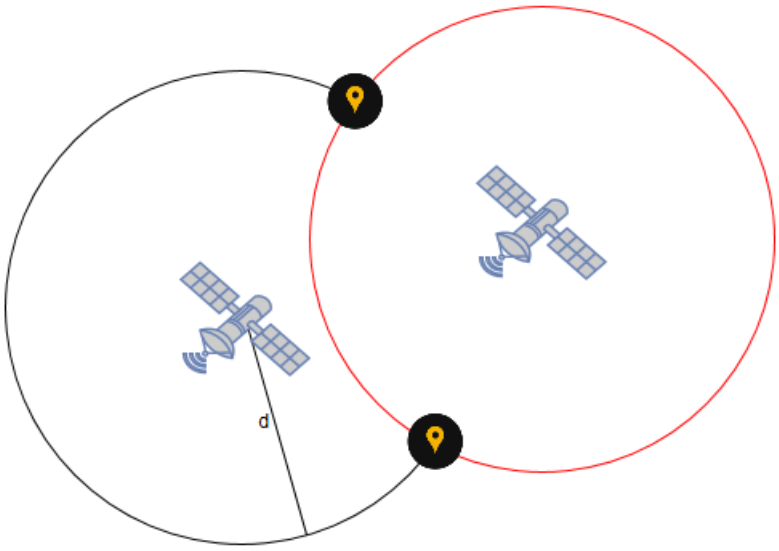

FIGURE 4.7: Intersection entre 2 Signaux émis par 2 satellites

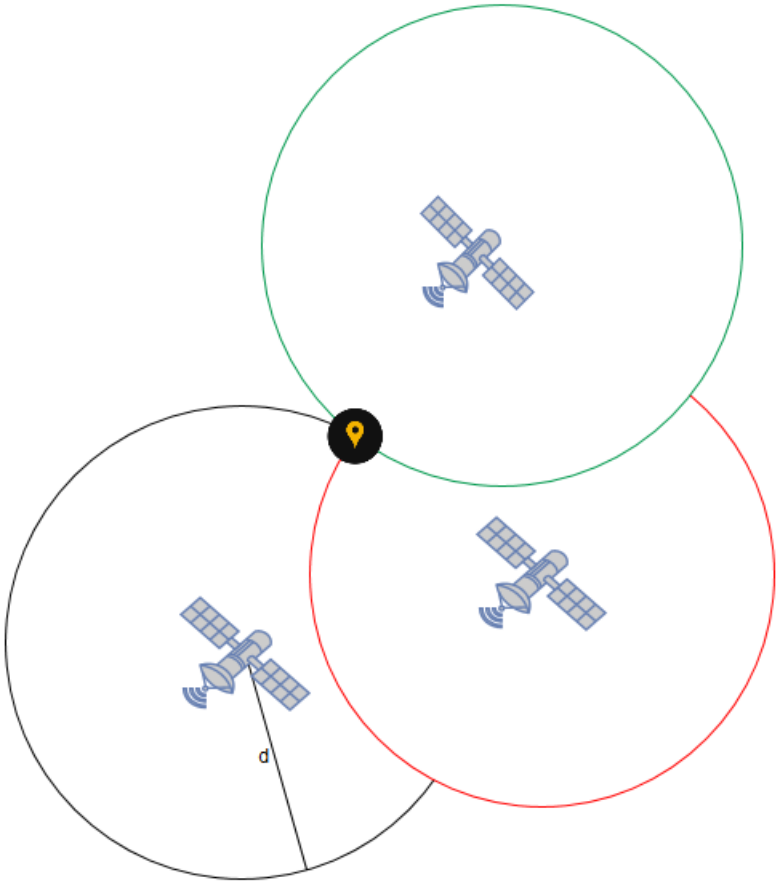

FIGURE 4.8: Point d'intersection entre 3 Signaux émis par 3 satellites

Par analogie au fonctionnement du GPS et si on prend en considération les hypothèses suivantes : La terre = un utilisateur U Un satellite = un objet (un device) O_i. Par analogie à la constellation des satellites en orbite autour de la terre, on peut considérer l'Aura d'authentification comme la constellation des devices en orbite autour d'un utilisateur. Chaque utilisateur délègue un device maître, dans ce cas le smartphone, qui remplace par la suite le rôle de l'utilisateur à gérer l'authentification des devices. Chaque device est représenté par son Aura Objet AO_i. Soit AO_1 l'Aura du smartphone. Imaginons que le smartphone reçoive un signal de proximité d'un 2ème device, donc on peut déduire qu'on se trouve à une distance d du 2ème device donc sur un cercle centré sur ce dernier. Ce cercle représente l'Aura Objet du 2ème device, soit AO_2. Soit $C_{1/2}$ la confiance calculée au point d'intersection de AO_1 du smartphone et AO_2 du 2ème device. Ajoutons un 3ème device qui envoie lui aussi un signal de proximité au smartphone, on aura donc une intersection entre les 3 Auras de devices.

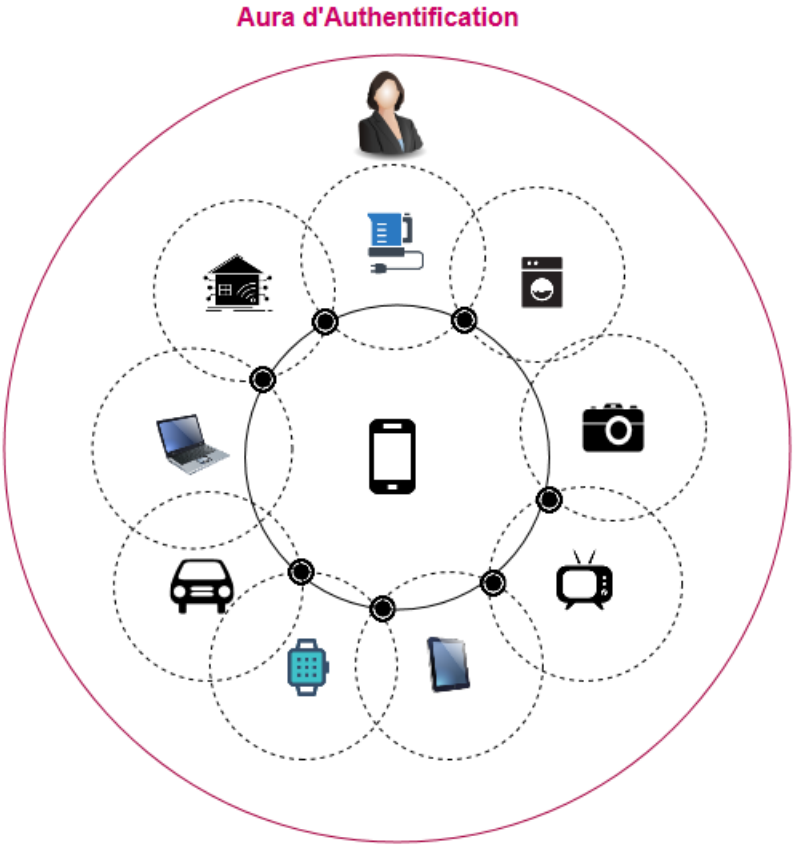

FIGURE 4.9: schéma explicatif de l'Aura d'Authentification d'une personne.

Plus loin, les satellites aussi communiquent entre eux et peuvent se localiser grâce aux éphémérides. En langage courant, une éphéméride désigne ce qui se passe quotidiennement. Une éphéméride du jour est la liste des évènements marquants de ce

jour. Disposant de leurs éphémérides, les satellites connaissent leurs propres positions sur leur orbite. Par exemple, ils savent qu'ils vont passer au-dessus de Caen tel jour à telle heure. Ils gardent ça en mémoire. Les éphémérides sont envoyées au capteur GPS dans le signal GPS. Le capteur dispose alors de la trajectoire du satellite pour les heures qui viennent. Dans le même contexte et par analogie aux satellites, les devices peuvent échanger entre eux des « éphémérides ». Comment peut-on définir les « éphémérides » des devices ? Les « éphémérides » des devices peuvent êtres des signaux partagés entre eux comme par exemple :

— Un signal pour dire que les devices appartiennent au même utilisateur (vérifier l'appartenance à la même aura d'authentification d'un utilisateur) ;

— un signal de proximité entre 2 devices ;

— un signal de proximité de l'utilisateur ;

— les comptes-rendus applications, des historiques d'appels et de connexion.

À partir du moment où ces signaux sont détectés et vérifient certaines conditions, on peut transférer de la confiance émanant d'un device à un autre device. Dans ce cas, l'intervention explicite de l'utilisateur (mot de passe, empreinte digitale, reconnaissance faciale...) afin d'assurer l'accès aux devices n'est plus nécessaire. Maintenant, la tâche d'authentification est déléguée aux devices et c'est entre eux qu'ils décident l'accès ou non d'un nouveau device.

Deux types d'informations sont envoyées au service du tiers de confiance. Premièrement, certaines données sont transmises pour calculer le niveau de confiance associé à un objet connecté (Authentification transparente mono-device). Deuxièmement, des informations de géolocalisation sont également transmises afin de mettre à jour le niveau de confiance d'un objet connecté O_i s'il se trouve dans l'Aura d'un objet connecté O_j ($i \neq j$). Le niveau de confiance transféré à l'objet O_j pourrait être un ratio du niveau de confiance associé à O_i et dépend également de la proximité des deux objets. La figure 4.10 illustre ce processus.

L'envoi de ces informations pourrait constituer un problème de sécurité en cas d'interception par un attaquant ou si le service est considéré honnête mais curieux. Nous proposons un système de protection de la vie privée qui permet à la fois de protéger les informations personnelles de l'utilisateur et de permettre au service de calculer le niveau de confiance (sans savoir quel type d'information a été utilisé pour l'authentification transparente).

4.4.2 Procédé de transfert de confiance

Estimation de la confiance d'Aura d'authentification

Considérons un utilisateur U, ayant un nombre de devices nd et un nombre de hotspots nh définis à priori. Un hotspot est une zone de confiance comme le domicile ou le lieu de travail de l'individu. Cet utilisateur a la possibilité de déclarer ses devices et ses Hotspots via une application fournie par un tiers de confiance, comme

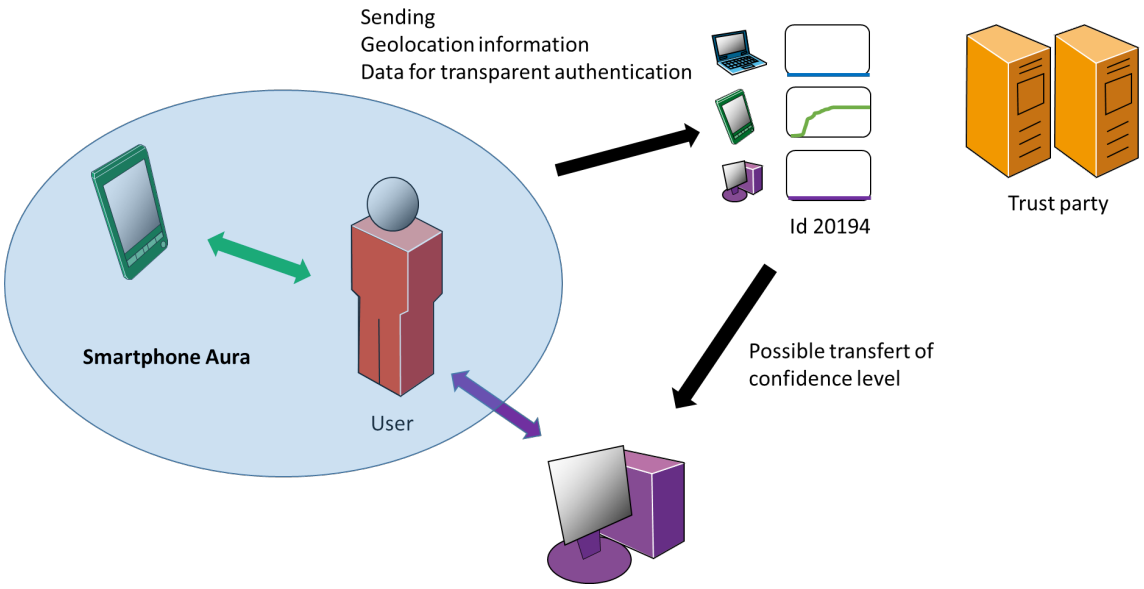

FIGURE 4.10: Illustration : Le smartphone de l'utilisateur possède une Aura où la confiance est contrôlée par le tiers de confiance. Si l'ordinateur est situé dans l'Aura du smartphone (pas très loin dans un certain sens), une partie du niveau de confiance associé au smartphone peut être transférée à celui de l'ordinateur.

présentée dans la figure 4.11. L'utilisateur enregistre l'adresse IP de chacun de ses devices et les coordonnées GPS de ses hotspots.

FIGURE 4.11: Application Orange pour ajouter les devices et les Hotspots de confiance d'un utilisateur

Une fois enregistrés sur cette application, ils sont considérés comme devices et Hotspots de confiance. La confiance sur un hotspot (désigné par NAH) est variable et est fixée par l'utilisateur. Par exemple, dans le cas d'un hotspot de confiance élevée (comme chez moi par exemple), NAH peut être fixé à 100 %. Nous supposons que l'utilisateur dispose sur chacun de ses appareils d'une solution d'authentification transparente qui permet à un tiers de confiance de calculer un niveau de confiance garantissant qu'à tout moment, l'objet connecté est utilisé par le bon utilisateur. Nous avons détaillée cette solution dans le chapitre 3 et nous pouvons citer aussi notre papier [117].

Ce niveau de confiance évolue avec le temps, c'est-à-dire que l'appareil envoie à chaque intervalle de temps défini (par exemple, toutes les 3 minutes) des informations protégées (à partir de son comportement, données biométriques du visage, ...) au service de tiers de confiance pour le calcul de la confiance. Cette valeur de confiance diminue automatiquement à chaque intervalle de temps pour garantir que si l'objet connecté n'est pas utilisé par l'utilisateur légitime, il ne puisse pas être utilisé par un imposteur. Nous voulons, à tout moment, déterminer le niveau de confiance de chaque objet connecté en tenant compte de son Aura. Ce niveau peut être amélioré sur la base de deux cas d'utilisation :

— Les hotspots de confiance où se trouvent l'utilisateur et l'objet connecté,

— La proximité des objets connectés de confiance appartenant à l'utilisateur.

Nous étudions chacun de ces cas séparément. À cette fin, nous illustrons le processus en envisageant le scénario suivant : Alice possède 3 objets connectés : un smartphone (S), un ordinateur portable (L) et un ordinateur de bureau (PC). Afin de s'authentifier correctement sur ses appareils, Alice utilise des mots de passe, soit le même mot de passe pour les 3 appareils, soit des mots de passe différents pour chacun, voir figure 4.12. Dans les deux cas, cette méthode d'authentification est faible et vulnérable à différentes attaques. Nous souhaitons établir une connexion entre les objets connectés d'Alice afin de permettre le transfert d'une partie du niveau de confiance sur l'authentification d'un appareil à l'autre sans l'intervention d'Alice. Alice crée avec ses objets connectés une aura d'authentification A. Soit AO_1 l'aura du smartphone, AO_2 l'aura de l'ordinateur portable et AO_3 l'aura de l'ordinateur.

Dans ce scénario, nous définissons les constituants de chaque aura et nous les détaillons dans la figure 4.13.

Cas 1 : Aura avec appartenance à un Hotspot de confiance

Soit A_T l'aura d'Alice dans des Hotspots de confiance. Afin de calculer la confiance d'un device d'Alice (le smartphone par exemple), on veut d'abord vérifier s'il appartient à A_T, donc vérifier si ce device est dans un Hotspot de confiance. Soit $C(O_i)$ la confiance d'un objet O_i calculée individuellement (pas dans l'Aura). Nous définissons la confiance $C_H(O_i)$ d'un device O_i appartenant à un Hotspot par le minimum de la somme des produits de confiance d'un device O_j (pour tout i$\neq$ j) apparte-

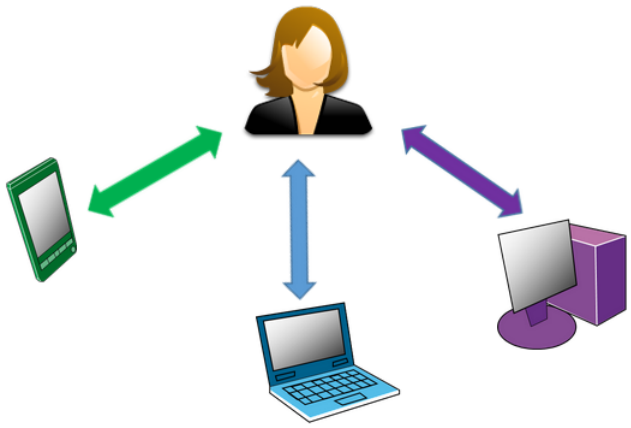

FIGURE 4.12: Procédure classique d'authentification

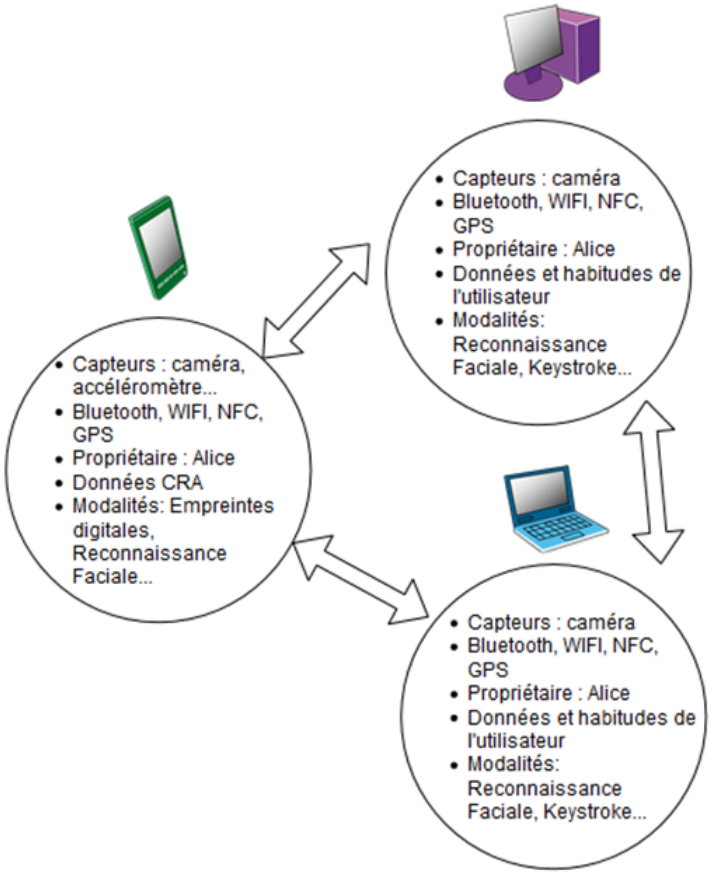

FIGURE 4.13: Définition des constitutions des auras d'objets.

nant au même Hotspot (ayant la confiance NAH) et 100 (la valeur maximale de la confiance), donnée par l'équation suivante :

$$C_H(O_i) = min(\sum_{i \neq j}^{n_H} NAH \times C(O_j) + C(O_i), 100) \tag{4.1}$$

avec n_H le nombre de devices d'Alice appartenant au même hotspot.

Cette valeur est mise à jour à chaque intervalle de temps fixée par le service tiers de confiance ou l'utilisateur. Nous supposons que la confiance initiale d'un nouvel appareil auquel Alice souhaite s'authentifier est de zéro. Afin de déterminer le transfert éventuel du niveau de confiance entre les objets connectés, nous devons vérifier s'ils sont localisés sur le même hotspot. Pour atteindre cet objectif, nous pouvons mesurer leur géolocalisation à l'aide de nombreuses données telles que les coordonnées GPS, l'adresse IP ou via la liste WIFI. Soit (g_1, g_2,...,g_n) les données de géolocalisation du smartphone d'Alice. Pour assurer la sécurité et la protection de la vie privée des données de géolocalisation, nous appliquons l'algorithme BioHashing pour générer un Biocode de géolocalisation avec la clé secrète d'Alice (ici, une valeur de germe aléatoire). Afin de décider si l'appareil est situé dans un hotspot de confiance connu, nous calculons la distance de Hamming dist_H entre les Biocodes des données de géolocalisation de l'appareil et le hotspot. Le niveau de proximité d'un hotspot est défini par la valeur de la distance de Hamming. Parmi un seuil de décision fixé par le tiers de confiance (gestion des risques), le service de confiance peut décider si l'appareil est situé dans l'un des hotspots de confiance d'Alice. Notez que le service de confiance n'est pas en mesure de savoir où se trouve ce hotspot car il ne connaît que son BioCode de géolocalisation et ne connaît pas la clé secrète d'Alice. La figure 4.14, illustre le scénario adopté.

Pour un nombre de device $n_H = 2$, imaginons une journée typique de Alice décrite par le scénario suivant : Alice utilise chaque matin son smartphone pour lire les actualités et consulter ses réseaux sociaux. Avec le temps de connexion passé, le smartphone gagne en confiance (Authentification Transparente). Sur son chemin au travail, elle continue à utiliser son smartphone pour appeler sa maman. Maintenant, en arrivant à son bureau, qui est déclaré déjà comme un Hotspot de confiance élevée (illustré sur la figure 4.15), elle veut s'authentifier auprès de son PC. Ayant un niveau de confiance assez élevé sur son smartphone, elle peut utiliser son PC sans se ré-authentifier. Le niveau de confiance de son PC, la confiance $C(L)$ peut-être calculée par la formule donnée ci-dessous 4.1 :

$$C_H(L) = min(NAH \times C(S) + C(L), 100) \tag{4.2}$$

Dans le cas où Alice utilise son smartphone mais pas assez suffisamment pour établir une confiance pour être partagée avec d'autres devices, une autre preuve d'authentification est demandée (un code PIN).

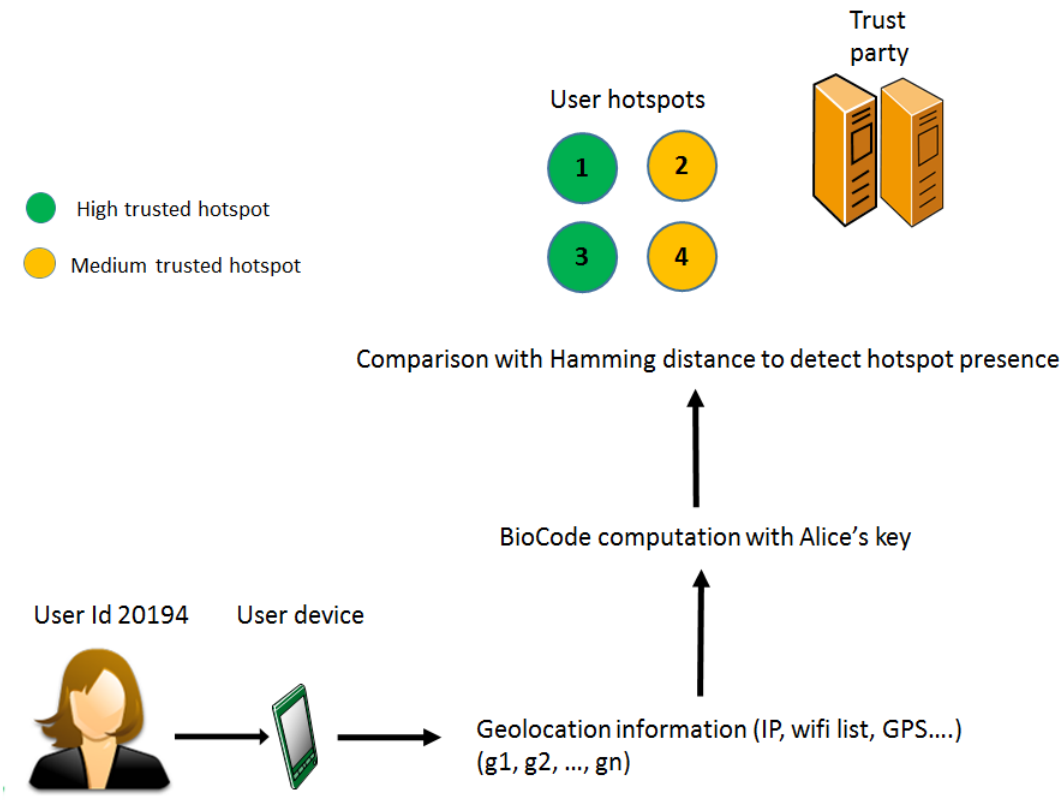

FIGURE 4.14: Illustration du scénario adopté pour déterminer la présence ou non d'un objet connecté dans un hotspot de confiance.

La figure 4.16 montre l'allure de l'évolution du niveau de confiance pour chaque device de l'aura. Maintenant considérons un exemple où le nombre de device d'Alice $n_H = 3$. Avant de rentrer chez elle, Alice passe chez ses parents (déclarée comme une zone de confiance moyenne). Elle veut consulter sa boite mail sur son ordinateur bureautique chez ses parents (voir figure 4.17). L'accès est autorisé et la confiance de l'ordinateur est donnée par :

$$C_H(PC) = min(NAH \times C(S) + NAH \times C(L) + C(PC), 100) \tag{4.3}$$

Comme Alice est placée dans une zone de confiance moyenne, NAH peut être fixé à 0.5. :

$$C_H(O) = min(0.5 \times (C(S) + C(PC)) + C(O), 100) \tag{4.4}$$

La figure 4.18 montre l'allure de l'évolution du niveau de confiance pour chaque device de l'aura. On constate qu'en essayant de s'authentifier auprès d'un nouveau objet connecté, on ne perd pas le niveau de confiance déjà acquis sur les autres objets de l'aura. Donc, initialement, le nouveau appareil gagne le niveau d'authentification de toute l'aura, quelque soit le type de Hotspot.

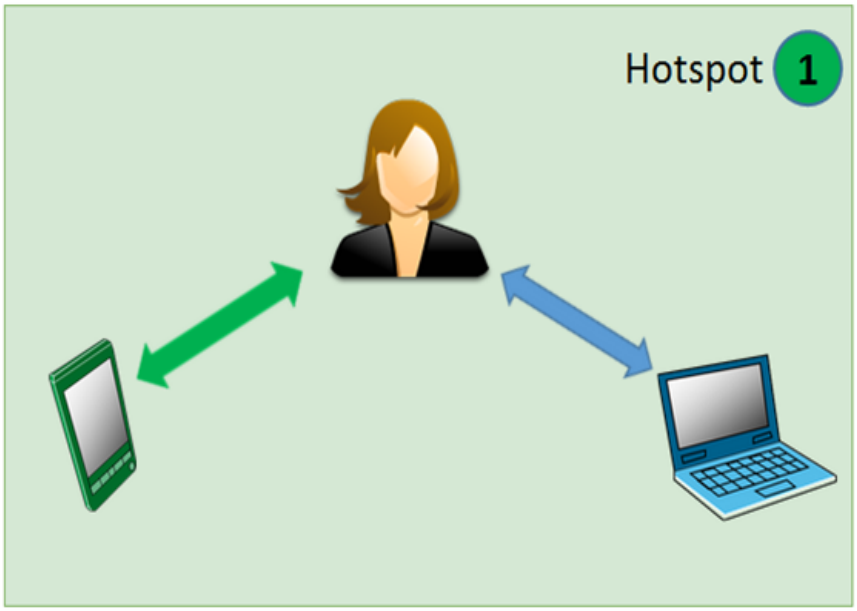

FIGURE 4.15: Aura des devices dans un même hotspot 1

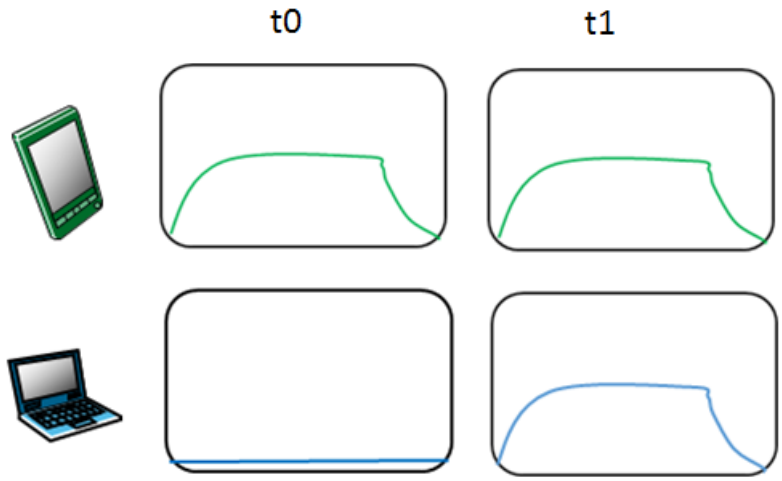

FIGURE 4.16: Évolution du niveau d'authentification pour chaque device de l'aura hotspot 1

Cas 2 : Aura avec une mesure de proximité entre les objets connectés

Dans le cas où Alice souhaite s'authentifier auprès de son objet connecté mais qui n'appartient pas à un hotspot déclaré de confiance dans son application, une Aura A_p peut être définie, basée sur la proximité entre les devices. Cette proximité est donnée par le calcul de la distance de Hamming entre les Biocodes des données

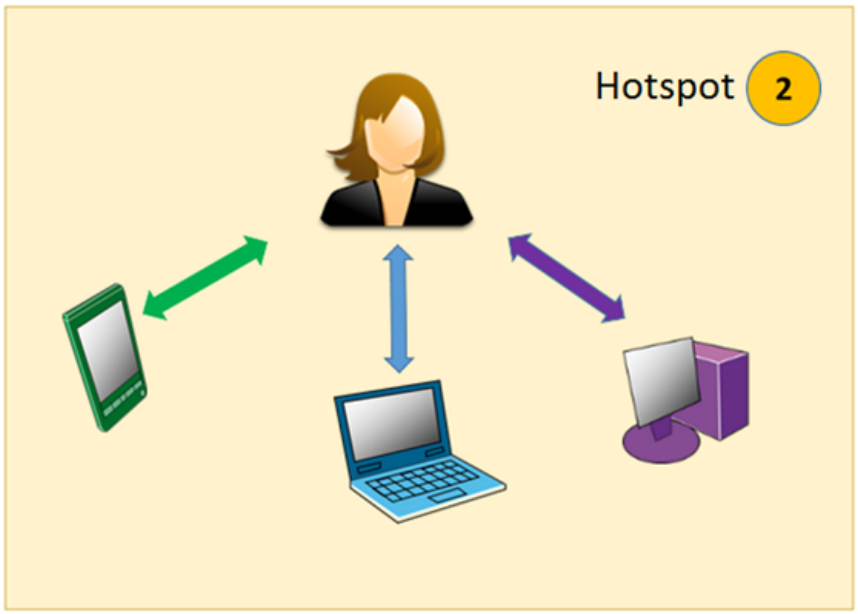

FIGURE 4.17: Aura des devices dans un même hotspot 2

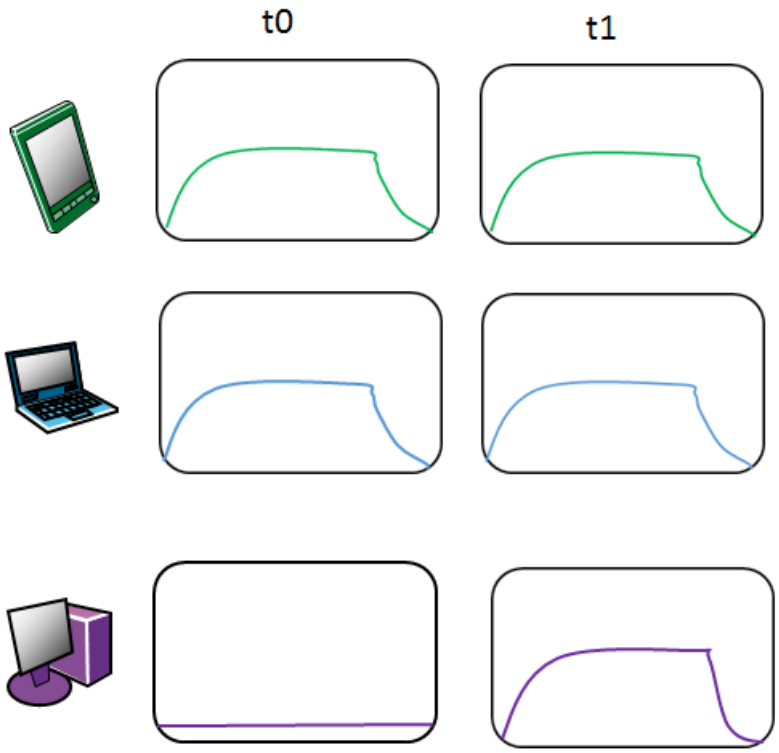

FIGURE 4.18: Évolution du niveau d'authentification pour chaque device de l'aura hotspot 2.

de géolocalisation (désigné par BG) des 2 objets connectés générés par la clé secrète d'Alice comme illustré dans la figure 4.19 .

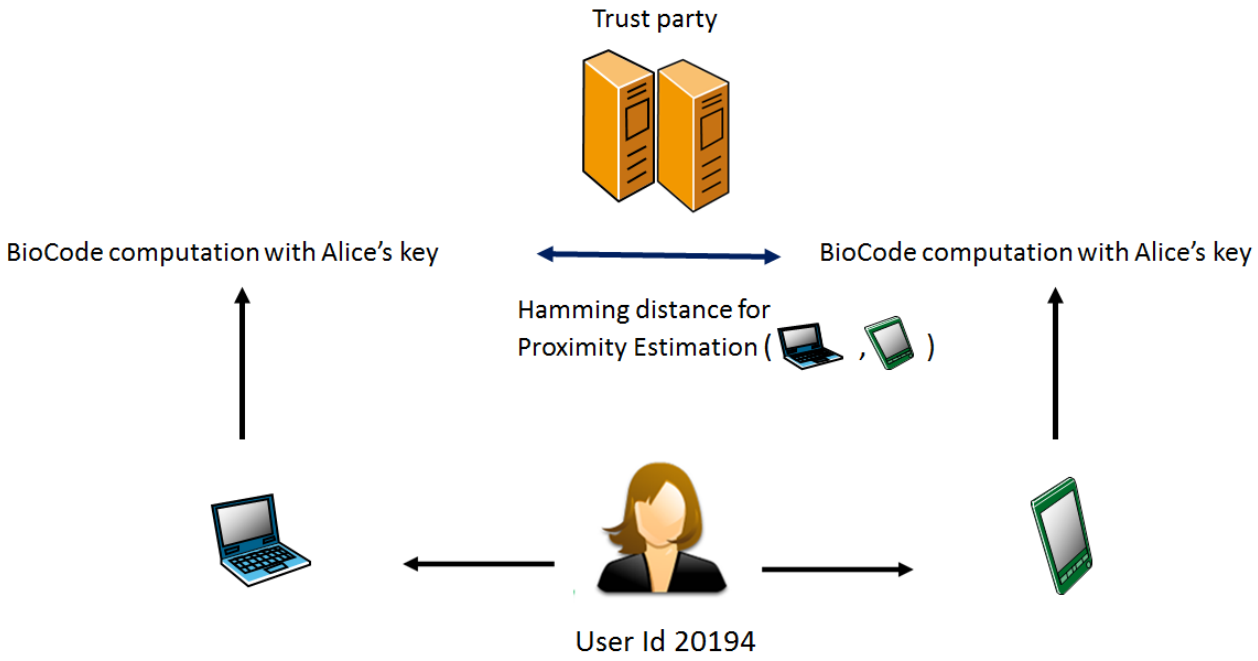

FIGURE 4.19: Illustration du scénario adopté pour mesurer la proximité entre les objets connectés lorsque Alice n'est pas présente dans un hotspot de confiance.

La proximité P entre deux objets connectés O_i et O_j est mesurée par :

$$P(O_i, O_j) = D_H(B_G(O_i), B_G(O_j)) \tag{4.5}$$

En plus, le tiers de confiance est capable de calculer la proximité entre les objets connectés mais ne dispose d'aucune information sur leur géolocalisation car ils sont protégés par l'algorithme BioHashing et la clé secrète d'Alice. La confiance $C_A(O_i)$ associée à l'objet connecté O_i, est définie par la somme des produits de la proximité des autres objets connectés O_j (pour tout i≠ j) par la formule suivante :

$$\mathbf{C}_A(O_i) = min(\sum_{i \neq j}^{nd} P(O_i, O_j) \times C(O_j) + C(O_i), 100) \tag{4.6}$$

avec n_d le nombre de devices d'Alice. Considérons l'exemple où Alice est dans une zone non déclarée comme zone de confiance et elle veut utiliser son ordinateur portable sachant qu'elle utilise aussi son smartphone, donc grâce à la proximité de son smartphone, elle peut s'authentifier auprès de son ordinateur implicitement et la confiance de l'ordinateur est donnée par :

$$C_A(L) = min(P(L, S) \times C(S) + C(L), 100) \tag{4.7}$$

Il faut noter que la confiance associée à un objet connecté est mise à jour à chaque intervalle de temps, avec le transfert depuis d'autres objets connectés (elle pouvait être non égale à zéro au moment précédent). Dans la section suivante, nous validons l'approche proposée dans un contexte expérimental.

4.5 Protocole expérimental

Nous détaillons dans cette partie le protocole expérimental utilisé pour la validation de la méthode proposée. Nous commençons par étudier le comportement de notre méthode sur des données simulées pour voir et analyser l'évolution de la confiance de l'aura par rapport à la confiance d'un unique objet connecté. Ensuite, nous validons notre méthode sur des données réelles en considérant les résultats de l'évolution de confiance trouvés sur un seul objet connecté dans le chapitre 3.

Pour générer des données théoriques, nous avons considéré un scénario exemple qui décrit une journée typique d'Alice, principalement en utilisant son smartphone, son ordinateur portable et son ordinateur bureautique. Alice se réveille tous les matins à 7h, elle utilise son smartphone pendant 20 min pour consulter ses e-mails et ses comptes sur les réseaux sociaux. Pendant le trajet à son travail de 8h à 9h, elle appelle ses parents. Elle utilise son ordinateur portable pour travailler de 9h à midi. Elle utilise également son ordinateur bureautique au travail de 9h50 à 11h50. Avant d'aller déjeuner à la cafétéria de midi à 14H, elle fait un appel depuis son smartphone de 11H à midi. Après avoir fini de manger, elle regarde les news sur son smartphone de 13h25 à 14h. À son retour au bureau, elle continue son travail sur son ordinateur portable de 14h à 16h. Elle utilise également son ordinateur bureautique de 15h à 17h50, et reprend le travail sur son ordinateur portable de 17h à 17h55. Pour des transferts de fichiers, elle utilise son smartphone de 16h à 19h. Le soir à la maison, elle regarde une série sur son ordinateur portable de 21h à minuit, et appelle une amie de 21h50 à 22h.

Nous générons d'abord la confiance de chaque objet connecté tout seul, le smartphone, l'ordinateur p et l'ordinateur portable bureautique. Après, nous calculons la confiance créée par l'Aura de chaque objet connecté en se basant sur l'équation 4.1 décrite dans la partie méthode proposée, quand il s'agit d'appartenance à un hotspot de confiance, et en se basant sur l'équation 4.6 quand il s'agit d'une Aura générée à partir d'une mesure de proximité entre les objets connectés. La validation de la méthode sur données réelles se fait par la suite en utilisant la reconnaissance faciale pour le calcul de confiance sur l'ordinateur portable et sur l'ordinateur bureautique, et les données basées sur les habitudes d'appels sur le smartphone (voir chapitre précédent).

Afin de garder une cohérence entre la partie utilisant la confiance simulée et celle calculée à partir de données réelles, et pour employer correctement les résultats de calcul de la confiance en authentification transparente de chaque objet dans le

chapitre 3, nous avons utilisé le même scénario d'usage d'objets connectés avec un NAH=100 pour tous les hotspots.

4.6 Résultats expérimentaux

Avant de valider notre méthode proposée sur des données réelles, nous souhaitons d'abord faire une étude de l'évolution de la confiance de l'Aura en utilisant des données chimériques. Nous présentons les résultats dans les 2 cas d'étude proposés : Calcul de la confiance de l'Aura avec appartenance à un hotspot de confiance et celle en se basant sur une mesure de proximité entre les objets connectés.

4.6.1 Confiance simulée

Aura avec appartenance à un Hotspot de confiance

Quand un objet connecté d'un utilisateur est détecté par le serveur dans un hotspot déclaré comme un hotspot de confiance, la confiance calculée sur cet objet connecté est ajouté à la confiance d'un autre objet connecté du même utilisateur qui appartient aussi au même hotspot de confiance, pour donner la confiance globale de l'Aura. Les figures 4.20, 4.21, 4.22 montrent respectivement l'évolution du niveau de confiance sur le smartphone d'Alice, son ordinateur portable et son ordinateur bureautique, au cours de la journée, dans un contexte d'authentification transparente via un seul objet connecté (pris séparément). Ces courbes de confiance sont simulées dans un premier temps.

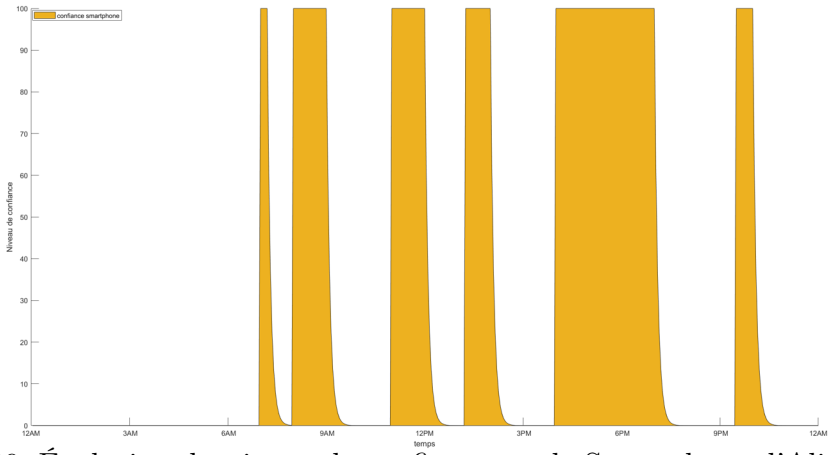

FIGURE 4.20: Évolution du niveau de confiance sur le Smartphone d'Alice au cours de la journée, dans un contexte d'authentification transparente via un unique objet intelligent.

Les figures 4.23, 4.25 et 4.24 montrent l'impact de la solution d'Aura d'authentification, basée sur l'appartenance dans un même hotspot, sur le niveau de confiance

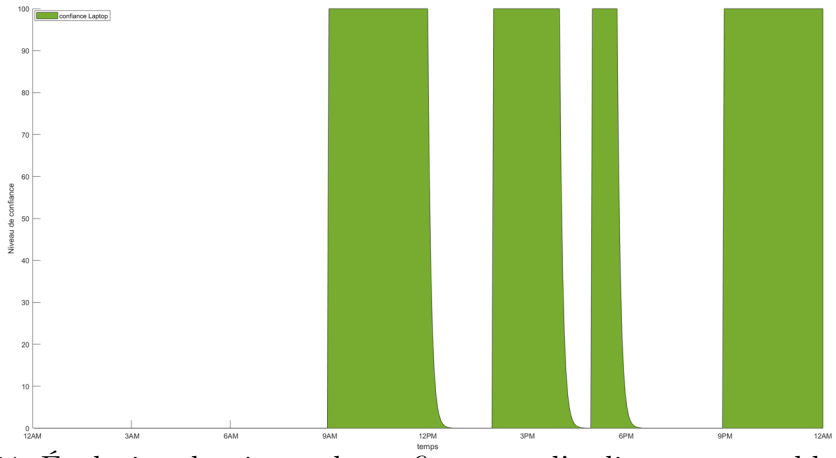

FIGURE 4.21: Évolution du niveau de confiance sur l'ordinateur portable d'Alice au cours de la journée, dans un contexte d'authentification transparente via un unique objet intelligent.

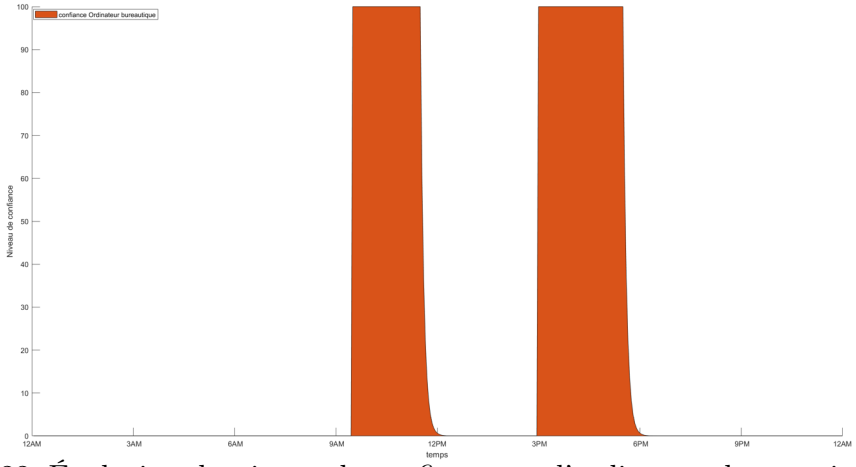

FIGURE 4.22: Évolution du niveau de confiance sur l'ordinateur bureautique d'Alice au cours de la journée, dans un contexte d'authentification transparente via un unique objet intelligent.

des objets connectés d'Alice, respectivement sur son smartphone, son ordinateur bureautique et son ordinateur portable. Nous pouvons clairement voir que la courbe de confiance d'Aura est au moins égale à la courbe de confiance d'un objet connecté sans appliquer la solution d'Aura. Ceci est conforme à la définition de l'équation 4.1. Dans cette partie, nous avons uniquement pris en compte l'impact positif de l'Aura sur le niveau de confiance de l'authentification sur un objet connecté.

Le fait d'être à proximité d'autres objets connectés et d'appartenir au même hotspot, entraîne un niveau de confiance plus élevé, grâce au transfert de confiance entre les objets connectés. Grâce à cette solution, nous n'avons pas besoin d'une autre demande d'authentification. Les appareils de l'utilisateur profitent de l'appartenance

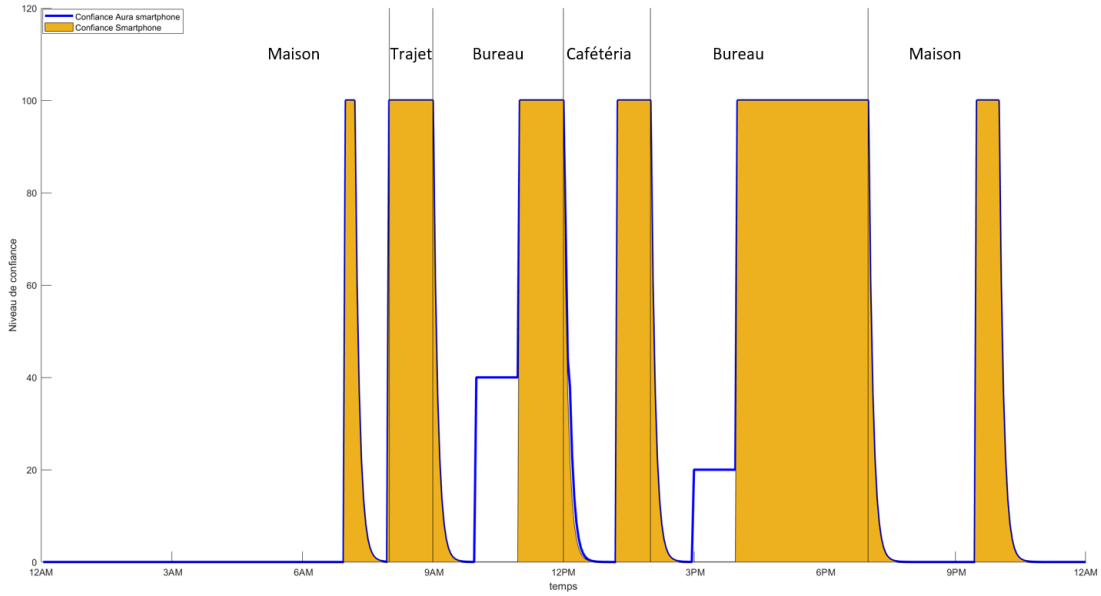

FIGURE 4.23: La courbe d'évolution de la confiance d'Aura d'Alice contre celle du smartphone d'Alice au cours d'une journée.

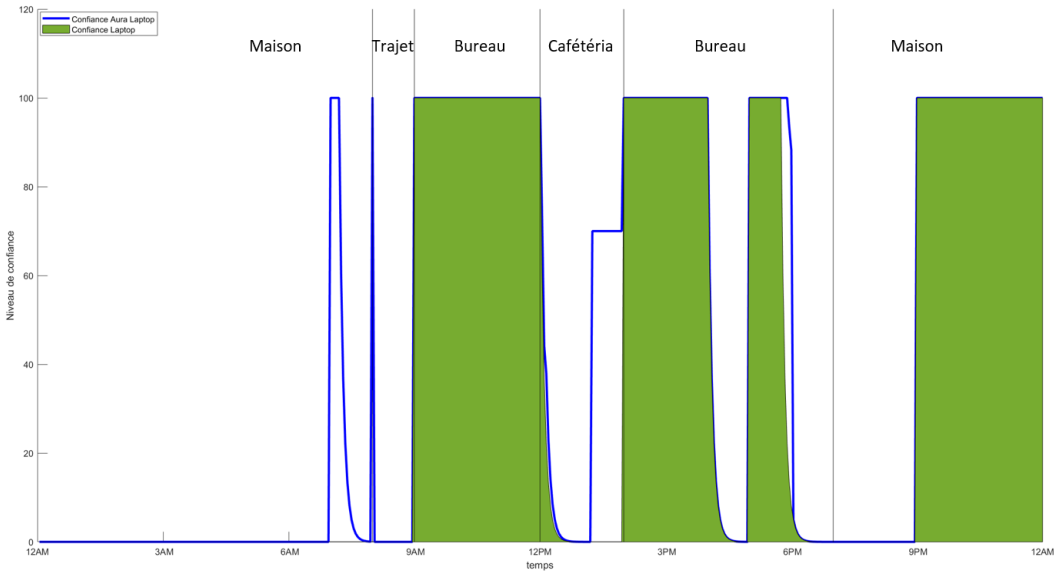

FIGURE 4.24: La courbe d'évolution de la confiance d'Aura d'Alice contre celle de l'ordinateur portable d'Alice au cours d'une journée.

au même hotspot et de l'authentification transparente acquise sur un seul objet connecté. Par exemple, on peut voir sur la figure 4.23, quand Alice est dans son bureau à 10h, le niveau de confiance sans la solution Aura est de 0% ce qui ne lui permet pas de s'authentifier auprès de son smartphone implicitement. Cependant,

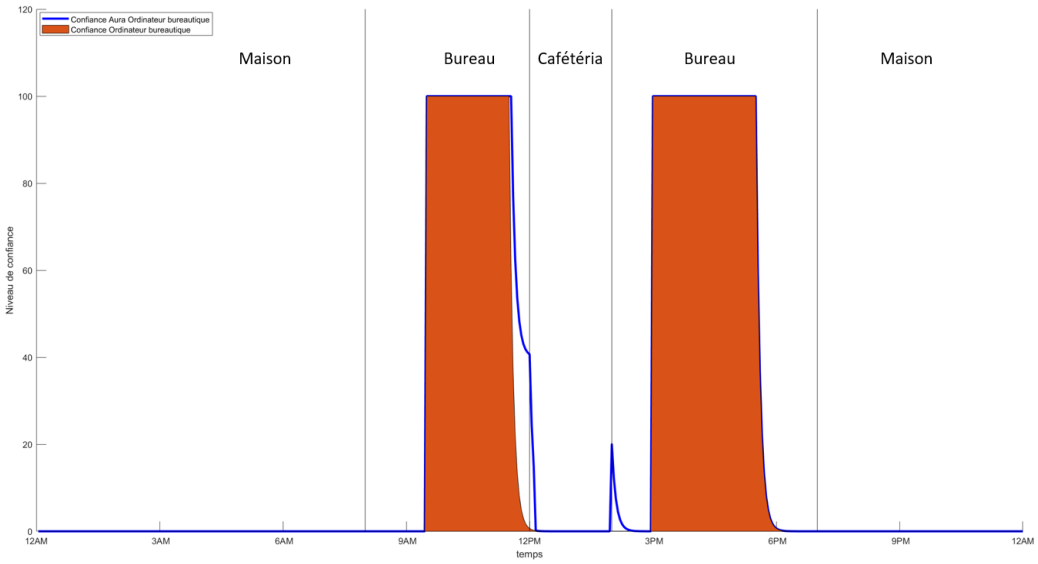

FIGURE 4.25: La courbe d'évolution de la confiance d'Aura d'Alice contre celle de l'ordinateur bureautique d'Alice au cours d'une journée.

en utilisant la solution Aura, nous pouvons voir que le niveau de confiance à 10h est égal à 42%, ce qui permet à Alice d'utiliser son smartphone sans aucune preuve supplémentaire, grâce à la proximité de son ordinateur et au fait d'être dans un hotspot de confiance. Nous constatons aussi sur la figure 4.24, que le niveau de confiance avec la solution Aura passe de 0% à 100% quand Alice est chez elle le matin à 7h, le moment où elle utilise son smartphone. Donc la confiance qu'elle a gagnée sur smartphone est transférée en totalité à son ordinateur portable qui est dans la même pièce avec elle. Alice peut également utiliser son ordinateur portable sans l'obligation de donner une preuve d'authentification supplémentaire en se fiant à la confiance qu'elle a acquise sur son smartphone pendant le temps qu'elle l'a utilisé. La même chose se produit quand elle est à la cafétéria, et qu'elle a suffisamment utilisé son smartphone de 13h20 à 14h, son ordinateur portable gagne en confiance (70%) si jamais elle souhaite l'utiliser à cette période. Une fois Alice est de retour à son bureau à 14h, la confiance augmente de 70% à 100%. Sans la solution proposée, Alice aurait dû présenter une preuve d'identité à son ordinateur portable pour y avoir accès. La solution Aura offre une authentification transparente sur plusieurs objets connectés en même temps.

En ce qui concerne la vie privée, le tiers de confiance, afin d'autoriser l'authentification, recueille des informations pour savoir si un objet connecté se trouve ou non dans un hotspot particulier, mais il n'est pas autorisé à connaître le contenu des données. Dans notre cas, le service de confiance n'a pas le droit d'avoir accès aux données de géolocalisation. Elle ne reçoit que les Biocodes BG car toutes les données collectées sont protégées par l'algorithme de Biohashing. Ainsi, la partie de confiance peut être informée de la présence d'Alice dans un hotspot de confiance

sans savoir exactement où elle se trouve.

Aura avec une mesure de proximité entre les objets connectés

Maintenant, comment calculer la confiance de l'Aura si Alice n'est pas placée dans un hotspot de confiance ? La solution alternative est simple, nous nous basons sur la proximité entre les objets connectés. Quand la distance entre les 2 objets est petite (c'est à dire inférieure à un certain seuil), alors la confiance gagnée sur un seul objet connecté est transférée à un autre. La confiance est calculée selon l'équation 4.6 donnée dans la méthode proposée. Les figures 4.6.1,4.27, 4.28 montrent l'impact de la solution d'Aura d'authentification, basée sur la proximité mesurée entre les objets connectés, sur le niveau de confiance des appareils d'Alice, respectivement sur son smartphone, son ordinateur portable et son ordinateur bureautique.

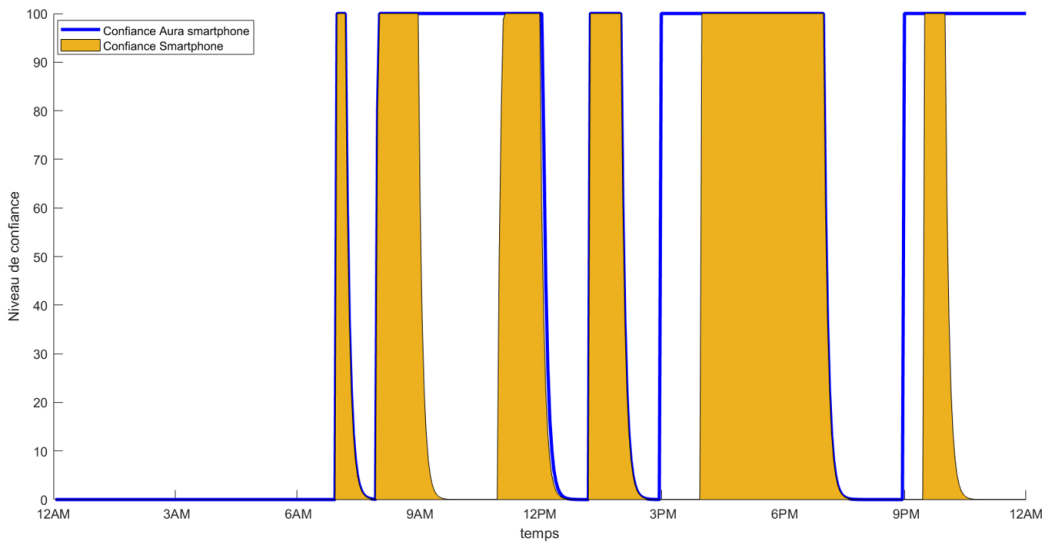

FIGURE 4.26: La courbe d'évolution de la confiance d'Aura d'Alice basée sur la proximité entre les devices contre celle du smartphone d'Alice au cours d'une journée.

Nous constatons sur la figure que la confiance de l'Aura du smartphone est à 100% de 10h à 11h par exemple même si à cette heure Alice n'utilise pas son smartphone, mais grâce à la proximité avec son ordinateur portable et son ordinateur bureautique, le smartphone gagne en confiance. De même à 15h et aussi à 21h, la confiance de l'aura du smartphone est maximale.

La confiance de l'Aura de l'ordinateur portable est à 100% pendant le trajet en voiture pour aller au travail le matin à 8h comme on peut le voir sur la figure 4.27 parce qu'il se trouve à proximité du smartphone pendant le trajet. Cette confiance reste maximale jusqu'à l'arrivée au bureau, où se trouve à proximité en plus l'ordinateur bureautique, et Alice n'a pas besoin de s'authentifier auprès de son ordinateur por-

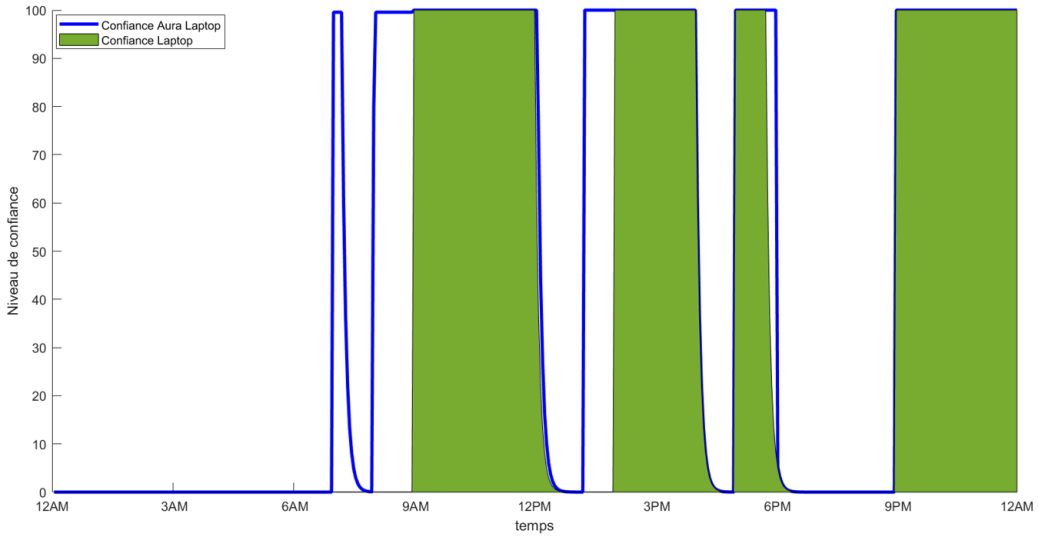

FIGURE 4.27: La courbe d'évolution de la confiance d'Aura d'Alice basée sur la proximité entre les devices contre celle de l'ordinateur portable d'Alice au cours d'une journée.

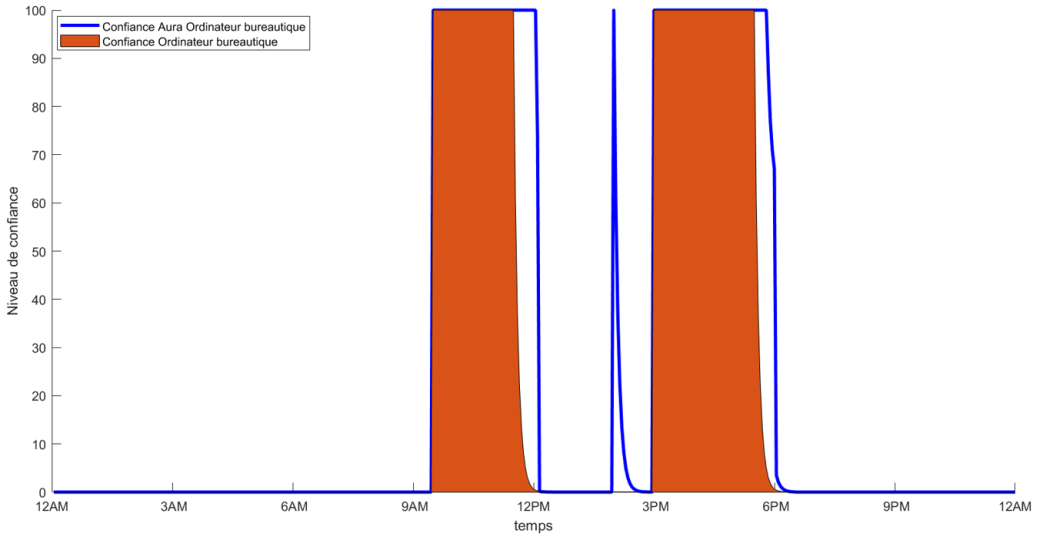

FIGURE 4.28: La courbe d'évolution de la confiance d'Aura d'Alice basée sur la proximité entre les devices contre celle de l'ordinateur d'Alice au cours d'une journée.

table, non plus de son ordinateur bureautique comme on peut le voir sur la figure 4.28.

4.6.2 Confiance réelle

Après avoir étudié et analyser la solution de l'Aura avec des données chimériques (avec une confiance simulée et sans attaque), nous souhaitons valider l'approche pro-

posée sur des donnée réelles. Nous utilisons les données de reconnaissance faciale sur l'ordinateur portable et l'ordinateur bureautique et les données d'habitudes d'appels sur le smartphone d'Alice.

Aura avec appartenance à un Hotspot de confiance

Nous rappelons d'abord les courbes de confiance générées dans le chapitre 3 lors de l'usage du smartphone, de l'ordinateur portable et de l'ordinateur bureautique, avec protection de données à clé secrète. Nous considérons différentes contextes lors de la reconnaissance faciale (expressions, luminosité, occultations partielles). Une tentative d'intrusion est réalisée sur l'ordinateur bureautique de 16h à 17h15. Les figures 4.29, 4.30, 4.31 montrent respectivement l'évolution du niveau de confiance sur le smartphone d'Alice, son ordinateur portable et son ordinateur bureautique, au cours de la journée, dans un contexte d'authentification transparente via un unique objet connecté.

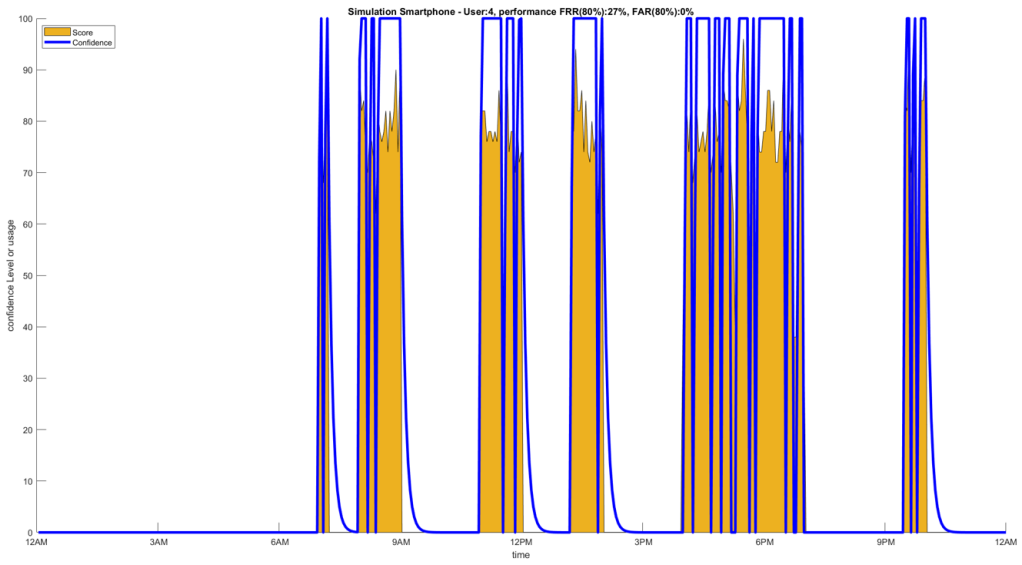

FIGURE 4.29: Évolution du niveau de confiance sur le Smartphone d'Alice au cours de la journée, dans un contexte d'authentification transparente via un unique objet intelligent (données réelles).

Les figures 4.32, 4.33 et 4.34 montrent l'impact de la solution d'Aura d'authentification avec des données réelles (basée sur l'appartenance dans un même hotspot) sur le niveau de confiance des appareils d'Alice, respectivement sur son smartphone, son ordinateur bureautique et son ordinateur portable.

Nous simulons le cas où un collègue d'Alice tente d'accéder à son ordinateur bureautique quand elle n'est pas présente à son bureau, comme on le présente par l'attaque en rouge de 16h à 17h15 sur la figure 4.31. L'Aura protège l'ordinateur et met la confiance à 0% comme on peut le voir sur la figure 4.34. En revanche, ceci n'affecte

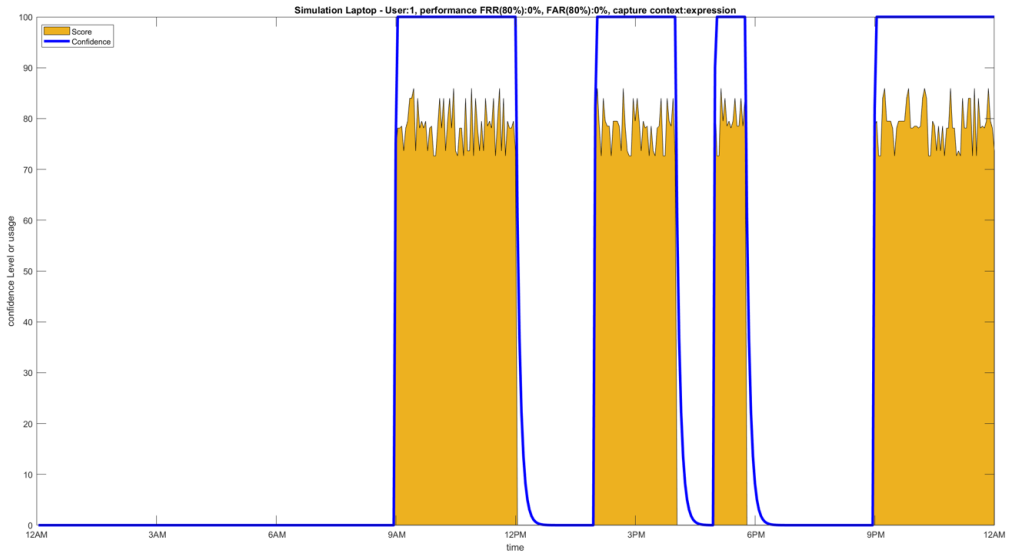

FIGURE 4.30: Évolution du niveau de confiance sur l'ordinateur portable d'Alice au cours de la journée, dans un contexte d'authentification transparente via un unique objet intelligent (données réelles).

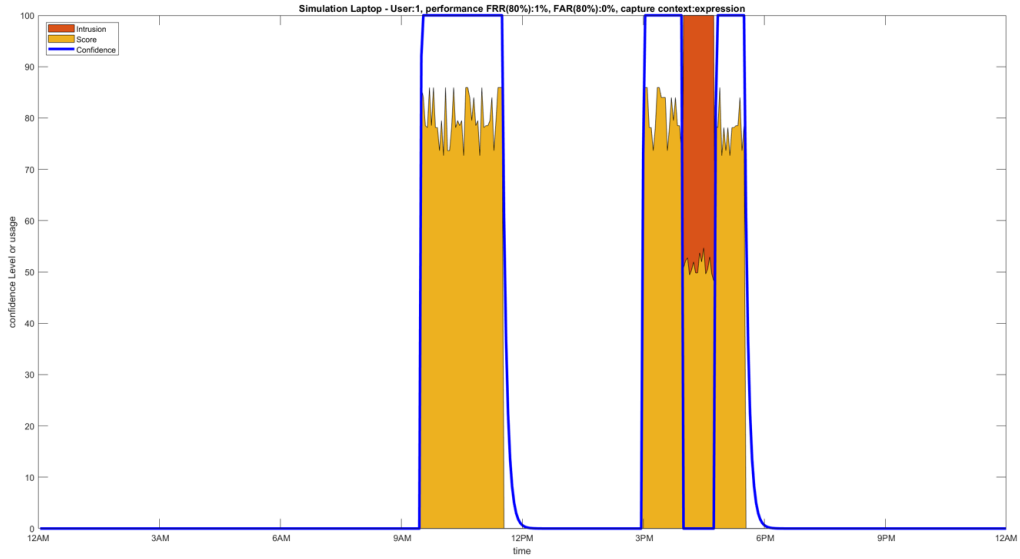

FIGURE 4.31: Évolution du niveau de confiance sur l'ordinateur bureautique d'Alice au cours de la journée, dans un contexte d'authentification transparente via un unique objet intelligent (données réelles).

pas la confiance sur les autres objets connectés d'Alice. Par exemple, si Alice est en train d'utiliser son smartphone le moment de l'attaque détectée sur son ordinateur, elle reste authentifiée sur son smartphone mais elle reçoit en plus une notification qui signale l'intrusion.

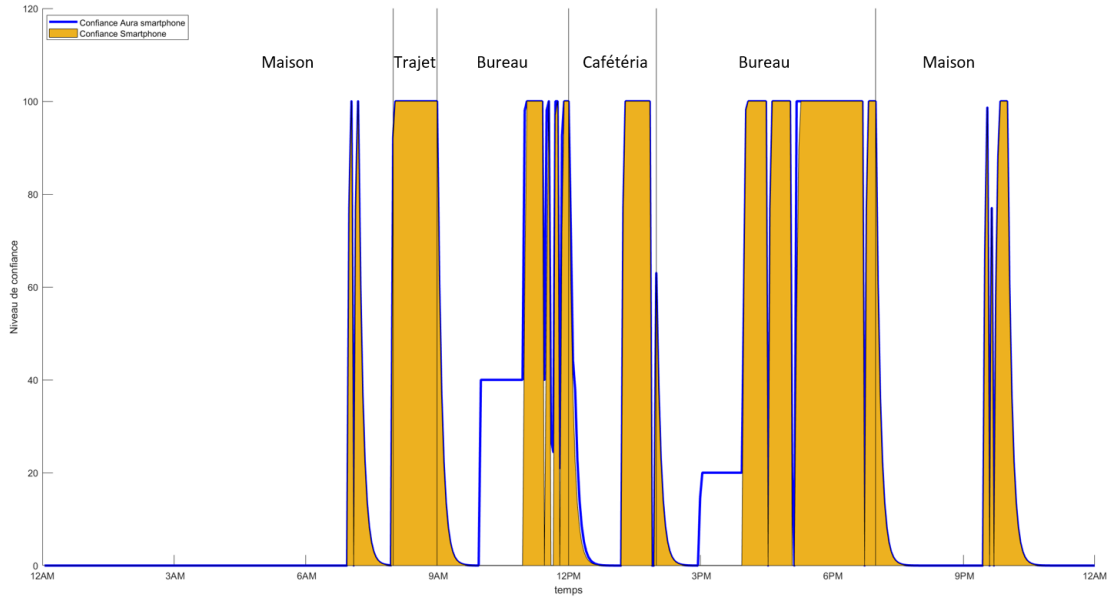

FIGURE 4.32: La courbe d'évolution de la confiance d'Aura d'Alice basée sur l'appartenance au même hotspot contre celle du smartphone d'Alice au cours d'une journée (données réelles).

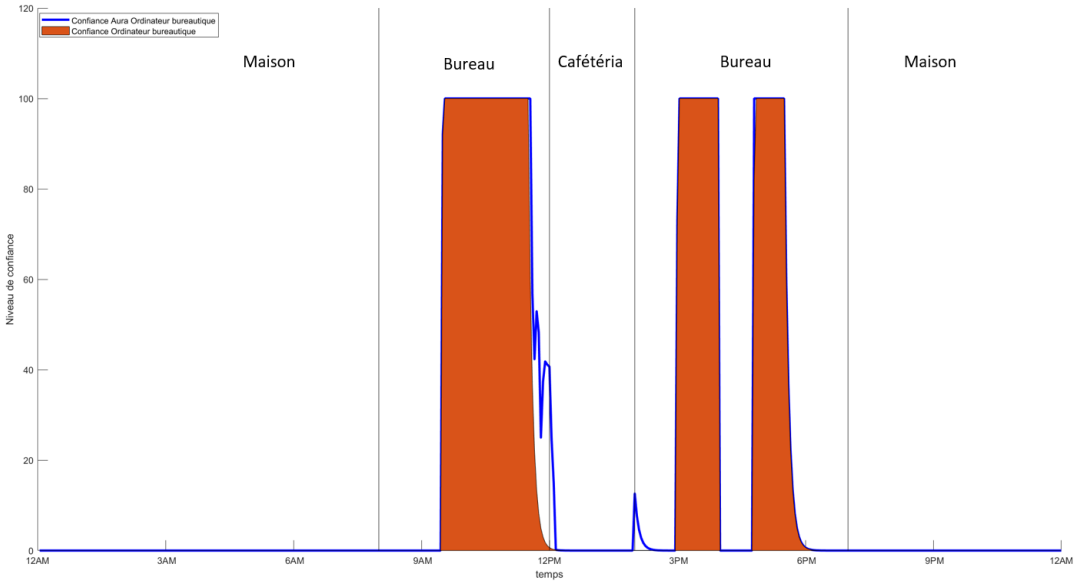

FIGURE 4.33: La courbe d'évolution de la confiance d'Aura d'Alice basée sur l'appartenance au même hotspot contre celle de l'ordinateur bureautique d'Alice au cours d'une journée (données réelles).

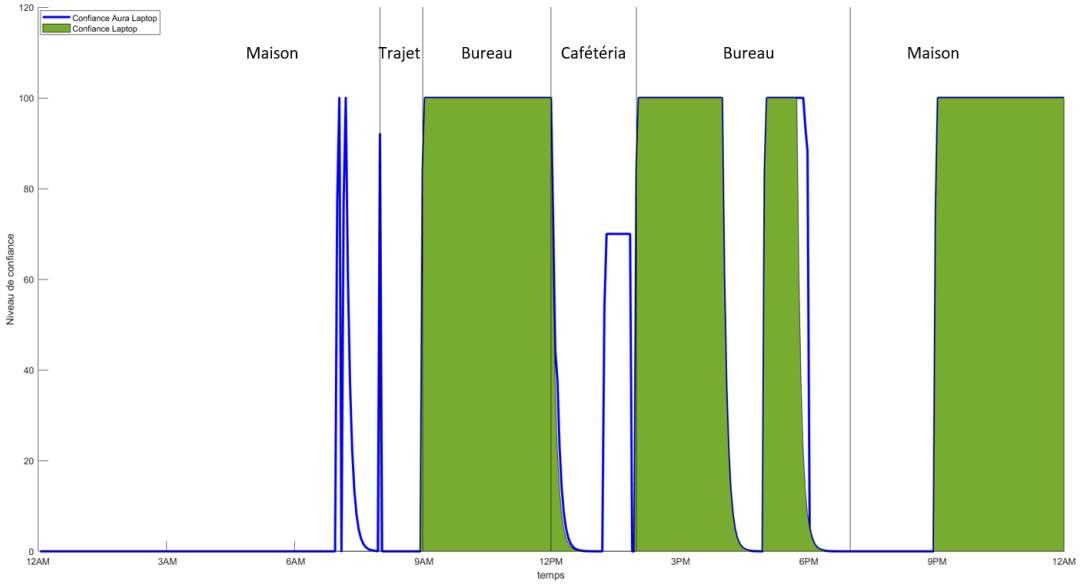

FIGURE 4.34: La courbe d'évolution de la confiance d'Aura d'Alice basée sur l'appartenance au même hotspot contre celle de l'ordinateur portable d'Alice au cours d'une journée (données réelles).

Aura avec une mesure de proximité entre les objets connectés

Nous souhaitons valider et analyser ces résultats sur des données réelles dans un contexte où nous utilisons cette fois la géolocalisation (protégée) des objets intelligents d'Alice. Nous utilisons les mêmes données, les objets connectés et le même scénario employés dans l'étude de l'Aura avec hotspots. Nous montrons sur la figure 4.35, la localisation des objets connectés d'Alice le long de la journée. La géolocalisation du smartphone en bleu, celle de l'ordinateur portable en vert et l'ordinateur bureautique au travail en rouge.

Les figures 4.36,4.37, 4.38 montrent l'impact de la solution d'Aura d'authentification, basée sur la proximité mesurée entre les objets connectés, sur le niveau de confiance des appareils d'Alice, respectivement sur son smartphone, son ordinateur portable et son ordinateur bureautique. Il est clair sur les 3 figures, que la confiance de l'Aura est plus importante que la confiance d'un seul objet connecté. Grâce à la proximité détectée entre les objets connectés d'Alice, la confiance est transférée d'un objet à un autre d'une manière transparente et continue.

Impact de l'Aura sur la confiance en Authentification transparente

Nous voudrons vérifier l'impact de l'Aura sur l'évolution de la confiance dans le cas où on considère une confiance inférieure à 70% sur un seul objet connecté. La figure 4.39 montre respectivement les 3 courbes de l'évolution de la confiance d'Aura d'Alice contre celle de son smartphone, son ordinateur portable et son ordinateur

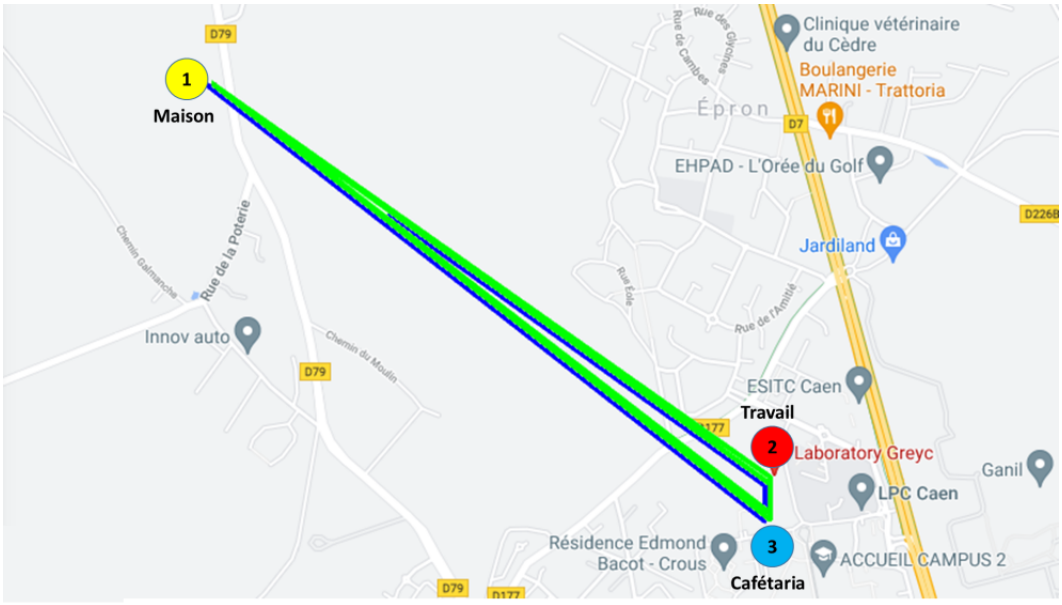

FIGURE 4.35: La localisation des objets connectés d'Alice au cours de la journée.

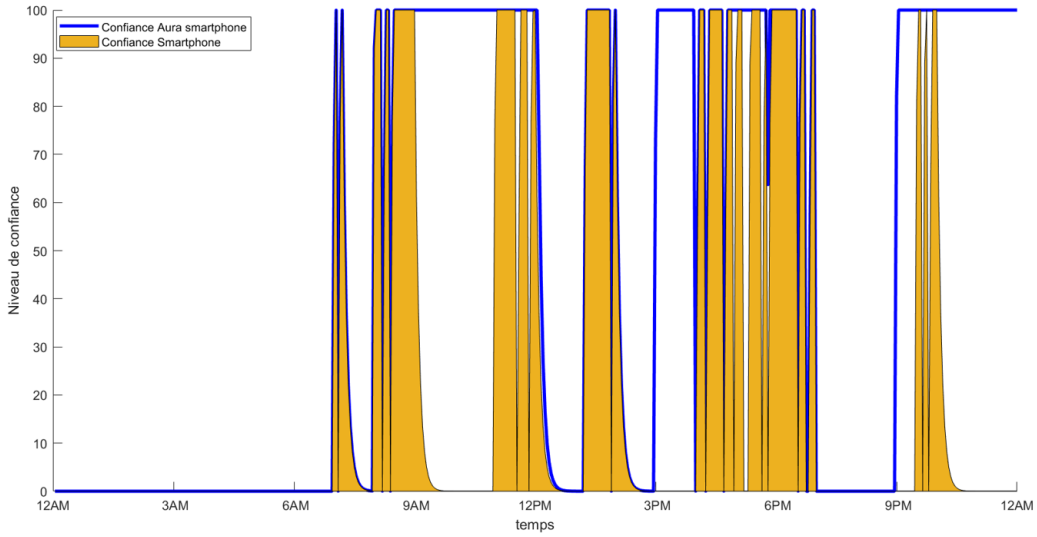

FIGURE 4.36: La courbe d'évolution de la confiance d'Aura d'Alice basée sur la proximité entre les devices contre celle du smartphone d'Alice au cours d'une journée (données réelles).

bureautique. Nous pouvons voir que la confiance de l'Aura est nettement plus importante et ceci est dû à l'appartenance à un hotspot de confiance qui fait que les confiances acquises sur les objets connectés dans le hotspot s'ajoutent entre elles.

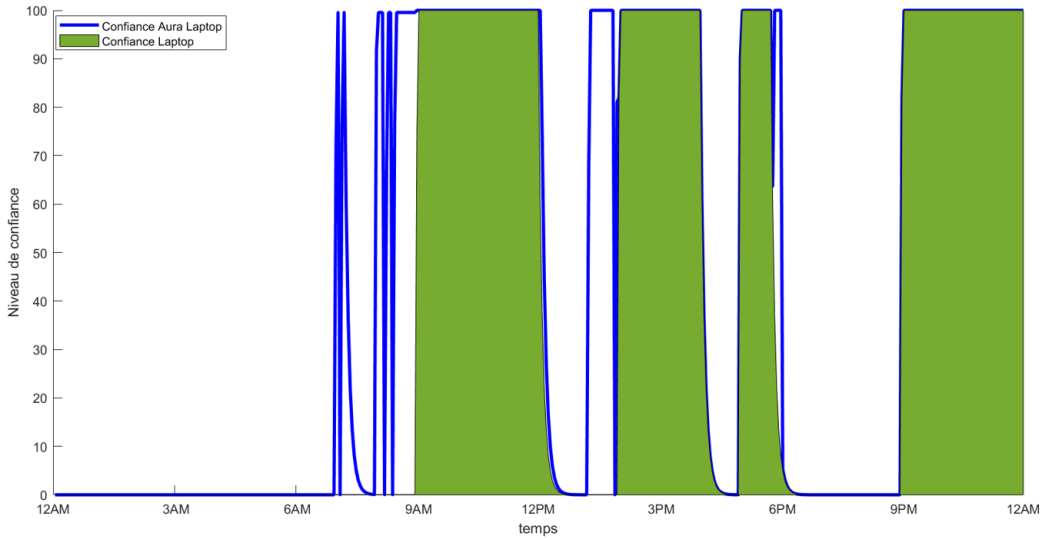

FIGURE 4.37: La courbe d'évolution de la confiance d'Aura d'Alice basée sur la proximité entre les devices contre celle de l'ordinateur portable d'Alice au cours d'une journée (données réelles).

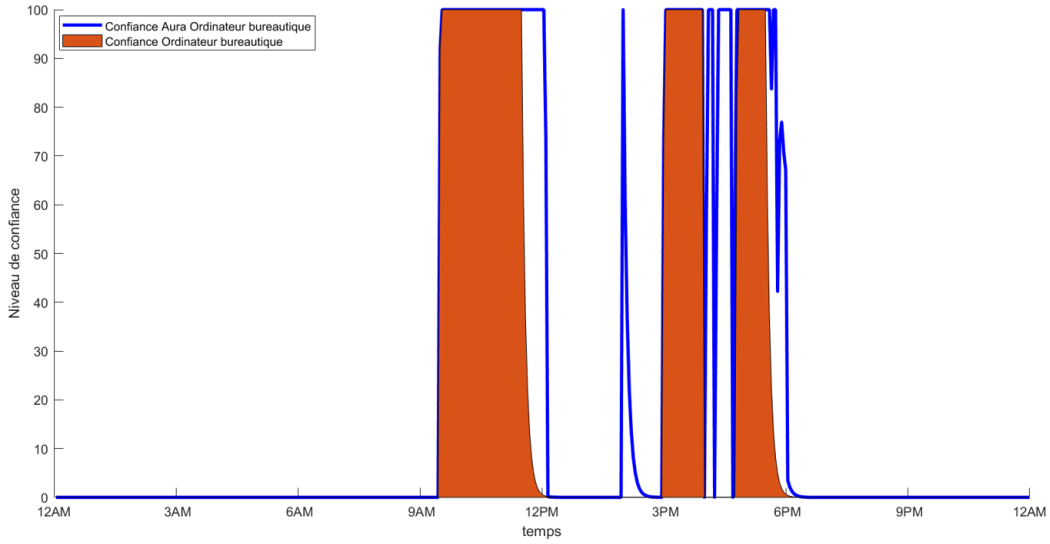

FIGURE 4.38: La courbe d'évolution de la confiance d'Aura d'Alice basée sur la proximité entre les devices contre celle de l'ordinateur bureautique d'Alice au cours d'une journée (données réelles).

4.7 Conclusion

Nous avons présenté et étudié dans ce chapitre une nouvelle approche d'authentification transparente via plusieurs objets connectés qu'on a appelé Aura d'authentification. Cette solution est basée sur le transfert de confiance entre les objets

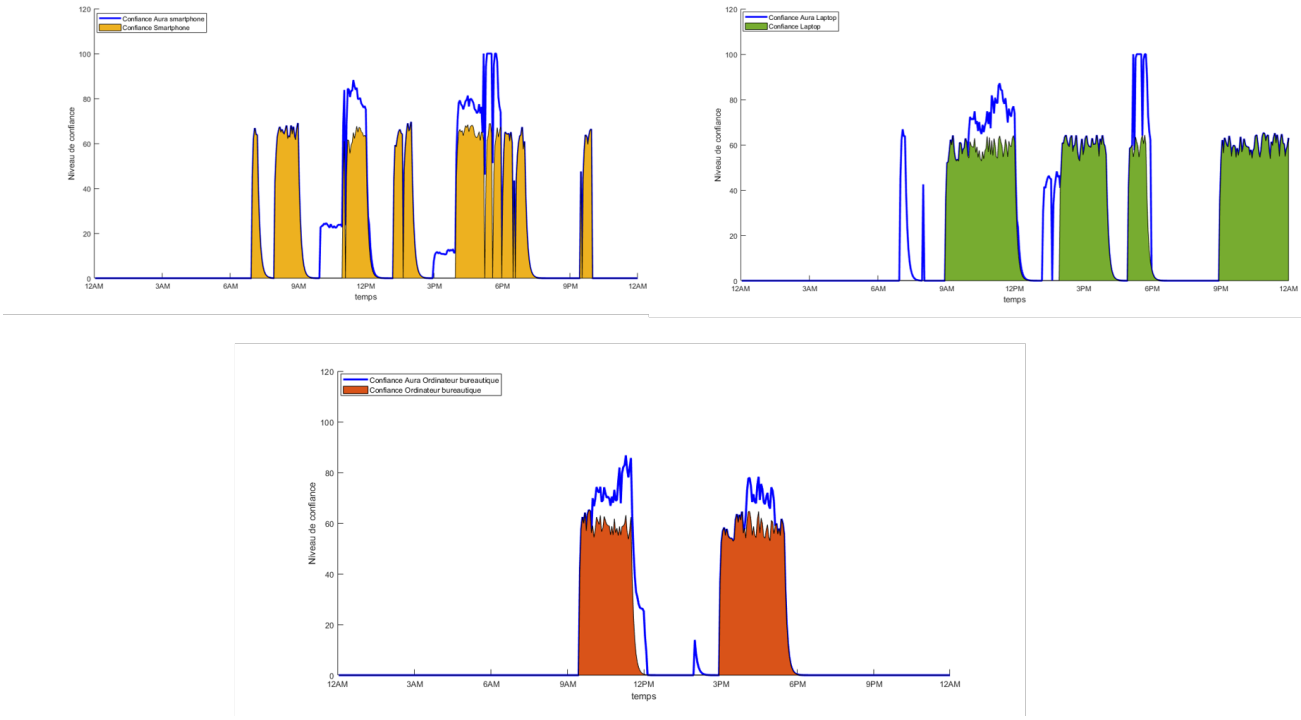

FIGURE 4.39: Impact de l'Aura sur la confiance de l'authentification transparente.

connectés quand ils se trouvent dans un hotspot de confiance ou quand ils sont à proximité. Nous avons commencé par définir quelques notions pour comprendre le concept de l'Aura et nous avons élaboré un état de l'art des travaux existants sur l'authentification via plusieurs objets connectés. Nous avons décrit par la suite la méthode utilisée en donnant des formulations précises du concept et en détaillant le procédé du transfert de confiance dans l'Aura. Deux cas d'usage ont été adopté pour estimer la confiance, le premier se base sur l'appartenance des objets connectés au même hotspot de confiance, et le deuxième sur la proximité entre les objets connectés de l'utilisateur.

Le protocole expérimental a été également détaillé afin de valider notre méthode proposée. Nous avons présenté les résultats expérimentaux de l'évolution de la confiance de l'Aura avec des données chimériques d'abord, et ensuite nous avons validé notre méthode avec des données réelles qui sont les habitudes d'appels sur smartphone et la reconnaissance faciale sur l'ordinateur portable et l'ordinateur, protégées par l'algorithme BioHashing à clé secrète. Les résultats trouvés sont intéressants en terme d'usage, de sécurité et de vie privée. La confiance de l'Aura est plus importante que la confiance d'un seul objet connecté permettant à l'utilisateur de s'authentifier à ses appareils sans la demande explicite d'une preuve d'identité. Le fait d'appartenir à un même hotspot de confiance ou d'être à proximité d'un objet connecté de l'Aura de l'utilisateur permet d'augmenter la confiance transparente et continue sur les objets connectés. Lors d'une tentative d'intrusion détectée, l'aura protège les objets connectés et mets la confiance à 0 sur l'objet attaqué et signale l'intrusion à l'utili-

sateur légitime. En plus, nous proposons un système de protection de la vie privée qui permet à la fois de protéger les informations personnelles de l'utilisateur et de permettre au service de tiers de confiance de calculer le niveau de confiance sans savoir quel type d'information a été utilisé pour une authentification transparente.

Chapitre 5

Conclusion et Perspectives

Nous nous sommes intéressés au cours de cette thèse à l'étude et à la conception des solutions d'authentification transparente, d'abord via un seul objet intelligent et via plusieurs objets connectés par la suite.

Nous avons commencé par une introduction générale avant de positionner la problématique de la thèse dans le chapitre 2. Dans ce chapitre, une définition du principe d'authentification et son rôle ont été présentés, ainsi qu'un aperçu technique des approches d'authentification existantes. Ensuite, un état de l'art sur les algorithmes de protection des données biométriques et sur l'Internet des Objets IoT est également abordé avant de définir les objectifs de cette thèse.

Par la suite, nous avons étudié une nouvelle méthode d'authentification transparente via un unique objet connecté. Nous avons évalué la performance de cette approche pour deux modalités biométriques. La première est basée sur les données d'habitudes d'appels issues d'un smartphone et la seconde utilise des images du visage de l'individu potentiellement acquise par une webcam d'un ordinateur. Nous avons appliqué des algorithmes d'apprentissage machine qui sont, la minimisation de distance, K-means et SVM à classe unique, pour évaluer la performance de la chaîne de traitement dans un contexte d'authentification statique (classique). La protection des données par l'algorithme BioHashing est ensuite prise en compte pour évaluer les performances du système en matière de protection des données. Les résultats expérimentaux sont convaincants avec des faibles valeurs d'EER surtout avec l'utilisation de l'algorithme BioHashing à clé secrète avec les données d'habitudes d'appels sur smartphone et les données faciales sur ordinateur. Ceci s'avère une approche intéressante pour une authentification robuste des utilisateurs qui respecte leur vie privée. Ensuite, nous avons évalué la confiance de l'usage du smartphone et de l'ordinateur au cours de la journée dans un contexte d'authentification transparente. Nous avons défini la fonction de calcul de confiance en fonction d'un score d'authentification généré à partir des données collectées sur l'individu, et à l'aide d'une loi normale centrée et paramétrée avec l'écart-type σ. Une étude sue l'impact de changement d'expression, de lumière et d'occultation a été fait sur les données de visage, l'impact

de la valeur de σ et l'impact du nombre d'échantillons NB sur la base de données du smartphone. Les résultats obtenus montrent la robustesse de l'approche proposée face aux différents changements et son efficacité en assurant une authentification transparente de l'utilisateur.

Nous avons également étudié une nouvelle méthode d'authentification transparente via plusieurs objets connectés basée sur le transfert de confiance d'un objet vers d'autres objets connectés de l'utilisateur appartenant à un même cercle de confiance qu'on a appelé Aura d'authentification. Nous avons validé notre méthode sur des données réelles basées principalement sur les habitudes d'appels sur le smartphone et la reconnaissance faciale sur l'ordinateur portable et l'ordinateur bureautique, protégées par l'algorithme BioHashing à clé secrète. Les résultats trouvés sont intéressants en terme d'usage, de sécurité et de vie privée. La confiance de l'Aura est plus importante que la confiance d'un seul objet connecté permettant à l'utilisateur de s'authentifier à ses appareils sans la demande explicite d'une preuve d'identité. Le fait d'appartenir à un même hotspot de confiance ou d'être à proximité d'un objet connecté de l'Aura de l'utilisateur permet d'augmenter la confiance transparente et continue sur les objets connectés. Lors d'une tentative d'intrusion détectée, l'Aura protège les objets connectés et met la confiance à 0 sur l'objet attaqué et signale l'intrusion à l'utilisateur légitime. En plus, nous proposons un système de protection de la vie privée qui permet à la fois de protéger les informations personnelles de l'utilisateur et de permettre au service de tiers de confiance de calculer le niveau de confiance sans savoir quel type d'information a été utilisé pour une authentification transparente. Ces résultats nous permettent de développer notre solution d'authentification transparente via plusieurs objets connectés dans un but d'industrialisation.

Nous souhaitons proposer différentes perspectives. Tout d'abord, il serait intéressant de tester la solution d'Aura d'authentification avec d'autres données biométriques comportementales (la frappe au clavier, la reconnaissance vocale, la reconnaissance de la démarche...) et étudier l'impact de la fusion de ces modalités sur l'évolution de la confiance de l'Aura. La formulation mathématique proposée pour le calcul de la confiance globale pourrait également être améliorée, en tenant compte de la modalité utilisée.

De plus, dans nos travaux, nous nous sommes limités à l'application de l'algorithme de BioHashing pour la protection des données. Nous visons à utiliser d'autres algorithmes de protection de données biométriques comme le BioPhasor et le GREYCHashing.

Dans le calcul de la confiance sur un objet connecté en authentification transparente, nous avons opté pour un choix empirique de la valeur de σ. Comme cette valeur affecte l'évolution de la confiance, il serait judicieux d'adopter un scénario plus précis pour générer les distributions de tous les scores légitimes et d'impostures de la base de données, pour pouvoir identifier la valeur exacte et précise de σ donnant les meilleures performances.

Nous tenons également à employer d'autres types d'objets connectés utilisés dans différents contextes (maison connectés, voiture intelligente...). Nous envisageons collaborer avec d'autres travaux de recherche de l'équipe de GREYC afin d'appliquer notre solution d'Aura pour aider à la vérification d'identité et la détection de fraude durant un examen à distance [118].

Publications

Conférences internationales

GUIGA, Takoua, ROSENBERGER, Christophe, et SCHWARTZMANN, Jean-Jacques. When my Behavior Enhances my Smartphone Security. In : 2020 International Conference on Cyberworlds (CW). IEEE, 2020. p. 280-284.

GUIGA, Takoua, ROSENBERGER, Christophe, et SCHWARTZMANN, Jean-Jacques. Privacy Aura for Transparent Authentication on Multiple Smart Devices. In : The 18th International Conference on Security and Cryptography, SECRYPT 2021

GUIGA, Takoua, ROSENBERGER, Christophe, et SCHWARTZMANN,Jean-Jacques. Face Transparent User Authentication Respecting Privacy. Soumission International Conference on Cyberworlds (CW). IEEE, 2021

Conférence nationale

GUIGA, Takoua, ROSENBERGER, Christophe, SCHWARTZMANN, Jean-Jacques, FREY, Vincent, et LACHARME, Patrick. Enjeux sécuritaires et de protection de la vie privée de l'authentification dans un environnement numérique ubiquitaire. l'Atelier sur la Protection de la Vie Privée, APVP 2018.

Ecole d'été

GUIGA, Takoua, ROSENBERGER, Christophe, et SCHWARTZMANN, Jean-Jacques. Behavioural Authentication Based on Smartphone Protected Personal Communication Data. In : Summer School on Biometrics and Forensics, Alghero, Italy, May 2019.

Bibliographie

[1] https://nymi.com/, [Online ; accessed 29-April-2019].

[2] S. Schneegass, Y. Oualil, and A. Bulling, "Skullconduct : Biometric user identification on eyewear computers using bone conduction through the skull," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 1379–1384.

[3] https://uidai.gov.in/, [Online ; accessed 30-April-2019].

[4] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 113, 2008.

[5] "Chaos Computer Club breaks Apple TouchID," https://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid, [Online ; accessed 30-April-2019].

[6] Google, "Google Abacus project," http://www.androidcentral.com/project-abacus-atap-project-aimed-killing-password, [Online ; accessed 29-April-2019].

[7] https://techcrunch.com/2016/05/23/google-plans-to-bring-password-free-logins-to-android-ap
[Online ; accessed 30-April-2019].

[8] J. Hatin, "Evaluation de la confiance dans un processus d'authentification," Ph.D. dissertation, 2017.

[9] https://unify.id, Note = [Online ; accessed 30-April-2019].

[10] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2021–2040, 2003.

[11] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore i am : Usability and security of authentication using brainwaves," in *International Conference on Financial Cryptography and Data Security*. Springer, 2013, pp. 1–16.

[12] https://www.teamsid.com/splashdatas-top-100-worst-passwords-of-2018/, Note = [Online ; accessed 07-01-2019].

[13] https://blog.avast.com/fr/41-des-francais-ne-modifient-pas-leur-mot-de-passe-apres-avoir-sub Note = [Online ; accessed 14-03-2018].

[14] https://www.inserm.fr/information-en-sante/dossiers-information/memoire, Note = [Online ; accessed 07-01-2020].

[15] A. Paivio, "Mental imagery in associative learning and memory." *Psychological review*, vol. 76, no. 3, p. 241, 1969.

[16] X. Suo, Y. Zhu, and G. S. Owen, "Graphical passwords : A survey," in *21st Annual Computer Security Applications Conference (ACSAC'05)*. IEEE, 2005, pp. 10–pp.

[17] http://www.realuser.com/, Note = [Online ; accessed 14-03-2019].

[18] N. Clarke, *Transparent user authentication : biometrics, RFID and behavioural profiling*. Springer Science & Business Media, 2011.

[19] I. Traore, M. Alshahrani, and M. S. Obaidat, "State of the art and perspectives on traditional and emerging biometrics : A survey," *Security and Privacy*, vol. 1, no. 6, p. e44, 2018.

[20] L. Wang and X. Geng, *Behavioral Biometrics for Human Identification : Intelligent Applications : Intelligent Applications*. IGI Global, 2009.

[21] B. Vibert, j.-m. Bars, C. Charrier, and C. Rosenberger, "Analyse d'empreintes digitales à partir de paramètres structurels calculés sur une référence réduite de l'image," 05 2016.

[22] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer vision and image understanding*, vol. 189, p. 102805, 2019.

[23] J. J. Winston and D. J. Hemanth, "A comprehensive review on iris image-based biometric system," *Soft Computing*, vol. 23, no. 19, pp. 9361–9384, 2019.

[24] A. de Santos-Sierra, C. Sanchez-Avila, J. Guerra-Casanova, and A. Mendaza-Ormaza, "Hand biometrics in mobile devices," in *Advanced biometric technologies*. InTech, 2011.

[25] Q. D. Smedt, H. Wannous, and J.-P. Vandeborre, "3d hand gesture recognition by analysing set-of-joints trajectories," in *UHA3DS@ICPR*, 2016.

[26] D. Migdal and C. Rosenberger, "Towards a Personal Identity Code Respecting Privacy," in *International Conference on Information Systems Security and Privacy (ICISSP)*, Madeira, Portugal, Jan. 2018. [Online]. Available : https://hal.archives-ouvertes.fr/hal-01620972

[27] T. Feng, Z. Liu, K. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen, "Continuous mobile authentication using touchscreen gestures," in *2012 IEEE Conference on Technologies for Homeland Security (HST)*, Nov 2012, pp. 451–456.

[28] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics : On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE transactions on information forensics and security*, vol. 8, no. 1, pp. 136–148, 2012.

[29] M. Derawi and P. Bours, "Gait and activity recognition using commercial phones," *Computers & Security*, 2013.

[30] I. Bouchrika, J. N. Carter, and M. S. Nixon, "Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 1201–1221, Jan 2016. [Online]. Available : https://doi.org/10.1007/s11042-014-2364-9

[31] M. Swain, A. Routray, and P. Kabisatpathy, "Databases, features and classifiers for speech emotion recognition : a review," *International Journal of Speech Technology*, vol. 21, no. 1, pp. 93–120, 2018.

[32] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, "The quest to replace passwords : A framework for comparative evaluation of web authentication schemes," in *2012 IEEE Symposium on Security and Privacy*. IEEE, 2012, pp. 553–567.

[33] https://www.cnil.fr/fr/reglement-europeen-protection-donnees/, [Online ; accessed 21-March-2021].

[34] A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization : Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management," 2010.

[35] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 2018.

[36] R. Kainda, I. Flechais, and A. Roscoe, "Security and usability : Analysis and evaluation," in *2010 International Conference on Availability, Reliability and Security*. IEEE, 2010, pp. 275–282.

[37] J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.

[38] H. Khan, A. Atwater, and U. Hengartner, "A comparative evaluation of implicit authentication schemes," in *International Workshop on Recent Advances in Intrusion Detection*. Springer, 2014, pp. 255–275.

[39] A. Plateaux, P. Lacharme, A. Jøsang, and C. Rosenberger, "One-time biometrics for online banking and electronic payment authentication," in *International Conference on Availability, Reliability, and Security*. Springer, 2014, pp. 179–193.

[40] N. Ratha, J. Connelle, and R. Bolle, "Enhancing security and privacy in biometrics-based authentication system," *IBM Systems J.*, vol. 37, no. 11, pp. 2245–2255, 2001.

[41] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. Kumar, "Biometric encryption," in *ICSA guide to Cryptography*. McGraw-Hill, 1999, vol. 22.

[42] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proceedings of the 6th ACM conference on Computer and communications security*, 1999, pp. 28–36.

[43] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[44] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237–257, 2006.

[45] J. Feng and A. Jain, "Fm model based fingerprint reconstruction from minutiae template," in *International conference on Biometrics (ICB)*, 2009.

[46] A. Goh and C. Ngo, *Computation of Cryptographic Keys from Face Biometrics*, ser. Lecture Notes in Computer Science. Springer, Berlin, 2003, vol. 2828.

[47] A. Teoh, D. Ngo, and A. Goh, "Biohashing : two factor authentication featuring fingerprint data and tokenised random number," *Pattern recognition*, vol. 40, 2004.

[48] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors : How to generate strong keys from biometrics and other noisy data," in *International conference on the theory and applications of cryptographic techniques*. Springer, 2004, pp. 523–540.

[49] Q. Li, Y. Sutcu, and N. Memon, "Secure sketch for biometric templates," in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2006, pp. 99–113.

[50] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *2007 Biometrics Symposium*. IEEE, 2007, pp. 1–6.

[51] Y. Luo, S. C. Sen-ching, and S. Ye, "Anonymous biometric access control based on homomorphic encryption," in *2009 IEEE International Conference on Multimedia and Expo*. IEEE, 2009, pp. 1046–1049.

[52] D. Boneh, A. Sahai, and B. Waters, "Functional encryption : Definitions and challenges," in *Theory of Cryptography Conference*. Springer, 2011, pp. 253–273.

[53] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, "Sectored random projections for cancelable iris biometrics," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1838–1841.

[54] C. Rathgeb, F. Breitinger, C. Busch, and H. Baier, "On application of bloom filters to iris biometrics," *IET Biometrics*, vol. 3, no. 4, pp. 207–218, 2014.

[55] A. Teoh and D. Ngo, "Cancellable biometrics realization through biophasoring," in *Proceedings of 9th IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV'06)*, 2006.

[56] A. Kong, K.-H. Cheung, D. Zhang, M. Kamel, and J. You, "An analysis of biohashing and its variants," *Pattern recognition*, vol. 39, no. 7, pp. 1359–1368, 2006.

[57] L. Nanni and A. Lumini, "Empirical tests on biohashing," *Neurocomputing*, vol. 69, no. 16-18, pp. 2390–2395, 2006.

[58] K. Simoens, P. Tuyls, and B. Preneel, "Privacy weaknesses in biometric sketches," in *2009 30th IEEE Symposium on Security and Privacy*. IEEE, 2009, pp. 188–203.

[59] K. Atighehchi, L. Ghammam, M. Barbier, and C. Rosenberger, "Greychashing : Combining biometrics and secret for enhancing the security of protected templates," *Future Generation Computer Systems*, vol. 101, pp. 819–830, 2019.

[60] P.-J. Benghozi, S. Bureau, and F. Massit-Folléa, *L'Internet des objets/The Internet of Things : Quels Enjeux Pour L'Europe ?/What Challenges for Europe ?* Les Editions de la MSH, 2009.

[61]

[62] https://www.windriver.com/whitepapers/gigaom-research-enabling-iot/White-Paper-Gigaom-Research-Enabling-iot.pdf, [Online ; accessed 14-April-2018].

[63] T. Godquin, M. Barbier, C. Gaber, J.-L. Grimault, and J.-M. Le Bars, "Iot : Vers un contrôle des fonctionnalités au vu des menaces liées," 2018.

[64]

[65] S. Babar, P. Mahalle, A. Stango, N. Prasad, and R. Prasad, "Proposed security model and threat taxonomy for the internet of things (iot)," in *International Conference on Network Security and Applications*. Springer, 2010, pp. 420–429.

[66]

[67] N. Vidgren, K. Haataja, J. L. Patino-Andres, J. J. Ramirez-Sanchis, and P. Toivanen, "Security threats in zigbee-enabled systems : vulnerability evaluation, practical experiments, countermeasures, and lessons learned," in *2013 46th Hawaii International Conference on System Sciences*. IEEE, 2013, pp. 5132–5138.

[68] M. Vanhoef and F. Piessens, "Key reinstallation attacks : Forcing nonce reuse in wpa2," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 1313–1328.

[69] M. B. Barcena and C. Wueest, "Insecurity in the internet of things," *Security Response, Symantec*, 2015.

[70]

[71] Z. Akhtar, C. Micheloni, and G. Foresti, "Biometric liveness detection : Challenges and research opportunities," *IEEE Security Privacy*, vol. 13, pp. 63–72, 09 2015.

[72] R. Belguechi, E. Cherrier, V. Alimi, P. Lacharme, and C. Rosenberger, "An overview on privacy preserving biometrics," in *Recent Application in Biometrics*. IntechOpen, 2011.

[73] R. Guidorizzi, "Security : Active authentication," *IT Professional*, vol. 15, pp. 4–7, 07 2013.

[74] N. L. Clarke and S. M. Furnell, "Authentication of users on mobile telephones–a survey of attitudes and practices," *Computers & Security*, vol. 24, no. 7, pp. 519–527, 2005.

[75] N. Clarke and A. Mekala, "Transparent handwriting verification for mobile devices," in *Proceedings of the Sixth International Network Conference (INC 2006), Plymouth, UK*. Citeseer, 2006, pp. 11–14.

[76] N. L. Clarke and S. M. Furnell, "Authenticating mobile phone users using keystroke analysis," *International journal of information security*, vol. 6, no. 1, pp. 1–14, 2007.

[77] N. Clarke, S. Karatzouni, and S. Furnell, "Transparent facial recognition for mobile devices," in *Proceedings of the 7th Security Conference, Las Vegas, USA, 2nd-3rd June*. Citeseer, 2008.

[78] ——, "Flexible and transparent user authentication for mobile devices," in *IFIP International Information Security Conference*. Springer, 2009, pp. 1–12.

[79] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley, "Casa : Context-aware scalable authentication," in *SOUPS '13 Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013.

[80] F. Li, N. Clarke, M. Papadaki, and P. Dowland, "Active authentication for mobile devices utilising behaviour profiling," *International Journal of Information Security*, 2013.

[81] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, "Text-based active authentication for mobile devices," in *ICT Systems Security and Privacy Protection*. Springer, 2014, pp. 99–112.

[82] M. Derawi and P. Bours, "Gait and activity recognition using commercial phones," *computers & security*, vol. 39, pp. 137–144, 2013.

[83] M. Tanviruzzaman and S. I. Ahamed, "Your phone knows you : Almost transparent authentication for smartphones," in *Computer Software and Applications Conference (COMPSAC), 2014 IEEE 38th Annual.* IEEE, 2014, pp. 374–383.

[84] G. Kambourakis, D. Damopoulos, D. Papamartzivanos, and E. Pavlidakis, "Introducing touchstroke : keystroke-based authentication system for smartphones," *Security and Communication Networks*, vol. 9, no. 6, pp. 542–554. [Online]. Available : https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.1061

[85] N. A. Safa, R. Safavi-Naini, and S. F. Shahandashti, "Privacy-preserving implicit authentication," in *IFIP International Information Security Conference*. Springer, 2014, pp. 471–484.

[86] M. Nauman, T. Ali, and A. Rauf, "Using trusted computing for privacy preserving keystroke-based authentication in smartphones," *Telecommunication Systems*, vol. 52, no. 4, pp. 2149–2161, 2013. [Online]. Available : http://dx.doi.org/10.1007/s11235-011-9538-9

[87] J. Hatin, E. Cherrier, J.-J. Schwartzmann, and C. Rosenberger, "Privacy Preserving Transparent Mobile Authentication," in *International Conference on Information Systems Security and Privacy (ICISSP)*, Porto, Portugal, Feb. 2017. [Online]. Available : https://hal.archives-ouvertes.fr/hal-01659952

[88] S. Y. Ooi and A. B. Teoh, "Touch-stroke dynamics authentication using temporal regression forest," *IEEE Signal Processing Letters*, vol. 26, no. 7, pp. 1001–1005, July 2019.

[89] D. K. D. Soni, M. Hanmandlu, and H. C. Saini, "A machine learning approach for user authentication using touchstroke dynamics," 2018.

[90] B. Attaullah, B. Crispo, F. Del Frari, and K. Wrona, "Touchstroke : Smartphone user authentication based on touch-typing biometrics," vol. 9281, 08 2015.

[91] M. P. Mufandaidza, T. D. Ramotsoela, and G. P. Hancke, "Continuous user authentication in smartphones using gait analysis," in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, Oct 2018, pp. 4656–4661.

[92] F. Sun, C. Mao, X. Fan, and Y. Li, "Accelerometer-based speed-adaptive gait authentication method for wearable iot devices," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 820–830, Feb 2019.

[93] P. Bours and T. Denzer, "Cross-pocket gait recognition," in *2018 International Conference on Cyberworlds (CW)*, Oct 2018, pp. 331–338.

[94] S. Nejr, A. Alruban, S. Furnell, and N. Clarke, "A novel behaviour profiling approach to continuous authentication for mobile applications," 02 2019.

[95] Y. Ashibani and Q. Mahmoud, "A behavior profiling model for user authentication in iot networks based on app usage patterns," 10 2018, pp. 2841–2846.

[96] H. Gomi, S. Yamaguchi, K. Tsubouchi, and N. Sasaya, "Continuous authentication system using online activities," 08 2018, pp. 522–532.

[97] G. Li and P. Bours, "A novel mobilephone application authentication approach based on accelerometer and gyroscope data," 09 2018, pp. 1–4.

[98] G. Ryu, S.-H. Kim, and D. Choi, "Implicit secondary authentication for sustainable sms authentication," *Sustainability*, vol. 11, no. 1, 2019. [Online]. Available : https://www.mdpi.com/2071-1050/11/1/279

[99] L. Lu, J. Yu, Y. Chen, H. Liu, Y. Zhu, Y. Liu, and M. Li, "Lippass : Lip reading-based user authentication on smartphones leveraging acoustic signals," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, April 2018, pp. 1466–1474.

[100] Q. Wang, X. Lin, M. Zhou, Y. Chen, C. Wang, Q. Li, and X. Luo, "Voicepop : A pop noise based anti-spoofing system for voice authentication on smartphones," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, April 2019, pp. 2062–2070.

[101] G. C. Banks, H. M. Woznyj, R. S. Wesslen, and R. L. Ross, "A review of best practice recommendations for text analysis in r (and a user-friendly app)," *Journal of Business and Psychology*, vol. 33, no. 4, pp. 445–459, 2018.

[102] C.Piech, "Kmeans," 2013. [Online] Available :http ://stanford.edu/cpiech/cs221/handouts/kmeans.html/. [Accessed : 22- Apr- 2019].

[103] A. Ng, "Cs229 lecture notes," *Intelligent Systems and their Applications IEEE*, 01 2000.

[104] A. M. Martinez, "The ar face database," *CVC Technical Report24*, 1998.

[105] P. Viola and M. Jones, "Face detection," *IJCV*, vol. 57, p. 2, 2004.

[106] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct 2016.

[107] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[108] C. G. Hocking, S. M. Furnell, N. L. Clarke, and P. L. Reynolds, "Authentication aura-a distributed approach to user authentication," *Journal of Information Assurance and Security*, vol. 6, no. 2, pp. 149–156, 2011.

[109] https://embeddedams.nl/different-ways-to-connect-iot-devices-to-transmit-and-receive-data/, [Online ; accessed 10-November-2020.

[110] Juniper, "Juniper research press releases, "iot connections to reach 83 billion by 2024, driven by maturing industrial use cases"," https://www.juniperresearch.com/press/press-releases/iot-connections-to-reach-83-billion-by-2024-driven, 2020,, [Online ; accessed 27-January-2021].

[111] O. Riva, C. Qin, K. Strauss, and D. Lymberopoulos, "Progressive authentication : deciding when to authenticate on mobile phones," in *21st {USENIX} Security Symposium ({USENIX} Security 12)*, 2012, pp. 301–316.

[112] B.-R. Cha, S.-H. Lee, S.-B. Park, G.-K. L. Y.-K. Ji *et al.*, "Design of micropayment to strengthen security by 2 factor authentication with mobile & wearable devices," *Advanced Science and Technology Letters*, vol. 109, no. 7, pp. 28–32, 2015.

[113] W. Xu, "Mobile applications based on smart wearable devices," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, 2015, pp. 505–506.

[114] L. Gonzalez-Manzano, J. M. de Fuentes, and A. Orfila, "Access control for the cloud based on multi-device authentication," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1. IEEE, 2015, pp. 856–863.

[115] J. Hajny, P. Dzurenda, and L. Malina, "Multi-device authentication using wearables and iot." in *SECRYPT*, 2016, pp. 483–488.

[116] J. Zhang, Z. Wang, Z. Yang, and Q. Zhang, "Proximity based iot device authentication," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[117] G. Takoua, R. Christophe, and S. Jean-Jacques, "When my behavior enhances my smartphone security."

[118] A. Haytom, C. Charrier, C. Rosenberger, C. Zhu, and C. Régnier, "Identity verification and fraud detection during online exams with aprivacy compliant biometric system," in *17th International Conference on Security and Cryptography (SECRYPT2020)*, 2020.

Table des figures

Liste des tableaux

Authentification Transparente dans un Environnement Numérique Ubiquitaire

L'authentification des individus est une tâche indispensable pour contribuer à la sécurité efficace des systèmes informatiques. Les solutions innovantes foisonnent et la recherche est très active, mais peu d'acteurs s'intéressent à l'expérience utilisateur pris dans la globalité de ses interactions numériques. Dans le cadre de cette thèse, nous proposons l'utilisation de la biométrie pour lier l'utilisateur avec ses objets connectés implicitement avec une solution multidevice, basée sur un cercle de confiance partagé entre les différents objets connectés permettant une authentification sécurisée de l'utilisateur et ses devices, qu'on appelle Aura d'authentification. Nous avons réalisé une étude de l'état de l'art sur les systèmes d'authentification et leurs exigences sécuritaires, les algorithmes de protection des données biométriques et sur les Objets connectés et l'Internet des objets IoT. Nous avons défini une méthode d'authentification transparente via un unique objet connecté, limitant les actions de l'utilisateur tout en protégeant sa vie privée. Nous avons validé cette approche sur des bases de données conséquentes en prenant en particulier le smartphone et l'ordinateur portable comme exemple d'objets intelligents. Nous proposons une approche originale d'authentification transparente via plusieurs objets connectés dans un environnement numérique ubiquitaire, qu'on appelle Aura d'authentification.

User authentication is an essential task to contribute to efficient security of IT systems. Innovative solutions abound and research is very active, but few actors are interested in the user experience considering all his/her digital interactions. In this thesis, we propose to use biometrics in order to connect the user with his/her connected objects implicitly with a multidevice solution, based on a circle of confidence shared between the different connected objects allowing a secure authentication of the user and his devices, which we call Authentication Aura. We have presented a state of the art on authentication systems and their security requirements, on biometric data protection algorithms and on connected objects and the Internet of Things IoT. We have defined a transparent authentication method via a single connected object, limiting the user's actions while protecting his privacy. We have validated this approach on large databases, taking in particular the smartphone and the laptop as examples of connected objects. We have proposed an original approach for transparent authentication via multiple connected objects in a ubiquitous digital environment, which we call Authentication Aura.

Mots clés: AUTHENTIFICATION, AURA, OBJETS CONNECTÉS, SÉCURITÉ, VIE PRIVÉE

ENSICAEN - UNICAEN - CNRS - GREYC UMR 6072, F-14050 Caen, France