le cnam

École doctorale Informatique, Télécommunications et Électronique (Paris)

Centre d'Études et de Recherche en Informatique et Communications

THÈSE DE DOCTORAT

présentée par : **Youcef OULD YAHIA**

soutenue le : **24 juin 2019**

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Spécialité : **Informatique**

Proposition d'un modèle de sécurité pour la protection de données personnelles dans les systèmes basés sur l'internet des objets

THÈSE dirigée par

Mme. BOUZEFRANE Samia	*Maître de conférence au CNAM, Paris*
Mme. BOUCHENEB Hanifa	*Professeur à Polytechnique Montréal, Canada*

RAPPORTEURS

M. BOUABDALLAH Abdelmadjid	*Professeur à l'Université de Technologie de Compiègne*
Mme. LAURENT Maryline	*Professeur à Télécom SudParis*

PRÉSIDENT

M. PARADINAS Pierre	*Professeur au CNAM, Paris*

EXAMINATEURS

M. PARADINAS Pierre	*Professeur au CNAM, Paris*
M. YACOUB Meziane	*Maître de conférence au CNAM, Paris*

Abstract

Ubiquitous computing through the democratization of mobile technologies, the Internet of Things (IoT) and cloud computing (with its different variants) have led to the emergence of new applications, particularly in the field of e-health. However, with the massive data collected via the IoT, it is easy to imagine the extent of the threat to users' privacy, if these technologies are used to spy on us. On the other hand, in the world where users' data have become a marketable resource, we are witnessing an evolution of the threat to private data. Indeed, this threat is no longer only embodied by the traditional "hacker" but can be carried by the service providers themselves, voluntarily or not. This is the case of Facebook-Cambridge Analytica, where personal data of Facebook users were used, without the knowledge of their owners, for political purposes. This example perfectly illustrates the problem of trust in the service provider and the consequences of the loss of control over personal data by their owners. To address these threats, robust data protection solutions and effective mechanisms for trust assessment must be implemented for end-to-end security solution.

However, the implementation of traditional security measures on IoT equipment is a first challenge related to the physical limits (energy, memory, processor) of connected things. This leads us to have to propose schemes adapted to the constraints of this environment, in particular through the offloading of heavy processes to unconstrained nodes. This path is made all the more viable by the emergence of new paradigms for the externalization of computer processing (Computational Offloading), such as fog-computing. On the other hand, the delegation of the processing and storage of our personal data to service providers (e.g. Cloud) raises the problem of trust in these service providers. The user, no longer having control over his data, must ultimately rely on the honesty and the security ability

of these providers. This problem of confidence also arises for the selection of the offloading service in terms of security and quality of service.

In this context, we have adopted an encryption solution that provides data protection centered on their owners.And we have worked to adapt it to the Internet of things environment, while integrating it into a protocol that reduces the privileges of the service providers. Our proposal for a suitable encryption scheme has been validated experimentally, in terms of performance, on a Raspberry PI card and formally, in terms of security. Then, in response to the trust issue and service selection, we explored the possibilities offered by artificial intelligence tools. To do this, we have proposed a collaborative filtering model based on Kohonen maps. This model has been experimentally evaluated on real-world dataset. This evaluation has shown that this approach is viable with promising results.

To conclude, we proposed a model for personal data protection, centered on their owners. This model integrates into the IoT environment with the offloading paradigms (e.g. Cloud computing, Fog/Edge Computing). These paradigms will undeniably be driven by the advent of 5th generation communication standards. That is, notwithstanding the encouraging experimental results obtained with the use of self-organizing mapss (Kohonen maps), an improvement in efficiency is possible by refining the parameters of the algorithm. Thus, an improvement of the proposed model is under study. This improvement aims to optimize the initialization of the maps and to improve the convergence of the model in our context by using a hybridization with the "firefly algorithm". Finally, as a future work, the proposed encryption scheme may evolve to support the decryption on constrained devices.

Keywords : Personal data, Internet of Things, Attributes-based encryption, Trust management, service offloading.

Résumé

L'informatique ubiquitaire à travers la démocratisation des technologies mobiles, de l'Internet des objets (IdO) et de l'informatique en nuage (avec ses différentes variantes) ont conduit à l'émergence de nouvelles applications, notamment dans le domaine de l'e-santé. Cependant, avec la masse de données collectées via l'IdO, il est facile d'imaginer l'ampleur de la menace sur la vie privée des utilisateurs, si ces technologies venaient à être utilisées pour nous espionner. Par ailleurs, dans un monde où les données des utilisateurs sont devenues une ressource monnayable, nous assistons à une évolution de la menace sur les données privées. En effet, cette menace n'est plus incarnée seulement par le traditionnel "pirate informatique" mais peut être portée par les fournisseurs de services eux-mêmes, volontairement ou non. C'est le cas de l'affaire Facebook-Cambridge Analytica, où des données personnelles d'utilisateurs de Facebook ont été utilisées, à l'insu de leurs propriétaires, à des fins politiques. Cet exemple illustre parfaitement le problème de confiance envers le fournisseur de service et les conséquences de la perte de contrôle sur les données personnelles par leurs propriétaires. Pour remédier à ces menaces, des solutions de protection des données robustes et des mécanismes efficaces pour l'évaluation de la confiance doivent être mise en œuvre de bout en bout.

Cependant, l'implémentation des moyens de sécurité traditionnels sur les équipements de l'IdO constitue un premier défi lié aux limites physiques (énergie, mémoire, processeur) des objets connectés. Ce qui nous amène à devoir proposer des schémas adaptés aux contraintes de cet environnement, notamment à travers la déportation des traitements lourds vers des équipements n'ayant pas de contraintes de ressources. Cette voie est rendue d'autant plus viable que de nouveaux paradigmes d'extériorisation des traitements informatiques (Computational Offloading) ont fait leur apparition, à l'image de l'informatique en brouillard

(fog-computing). D'un autre coté, la délégation du traitement et du stockage de nos données personnelles aux fournisseurs de services (e.g. cloud) nous pose le problème de confiance envers ces fournisseurs de service. L'utilisateur, n'ayant plus le contrôle sur ses données, doit, au final, se fier à l'honnêteté et à la capacité de sécurisation de ces fournisseurs. Ce problème de confiance se pose également pour la sélection du service d'extériorisation des traitements informatiques en termes de sécurité et de qualité de service.

Dans ce contexte, nous avons adopté une solution de chiffrement qui assure une protection de données centrée sur leurs propriétaires. Et nous nous sommes attelés à l'adapter à l'environnement contraignant des objets connectés, tout en l'intégrant dans un protocole qui tient compte des exigences de sécurité de l'environnement et de la réduction des privilèges des fournisseurs de services informatiques. Notre proposition de schéma de chiffrement adapté a été validée expérimentalement, en terme de performances, sur une carte de type Raspberry PI et formellement, en terme de sécurité, à l'aide d'un jeu cryptographique. Ensuite, en réponse à la problématique de la confiance et de la sélection du service, nous avons exploré les possibilités offertes par les outils de l'intelligence artificielle. Pour ce-faire, nous avons, proposé un modèle de filtrage collaboratif basé sur les cartes de Kohonen. Ce modèle a été évalué expérimentalement sur des données du monde réel en libre accès. Cette évaluation a démontré que cette approche est viable et donne des résultats satisfaisants.

Au final, nous avons proposé un modèle de protection des données personnelles, centré sur leurs propriétaires. Ce modèle s'intègre dans l'environnement de l'IdO et des paradigmes de déportation des traitements et du stockage de données (e.g. Cloud computing, Fog/Edge Computing). Ces paradigmes seront indéniablement portés par l'avènement des standards de communication de 5ème génération. Ceci étant, nonobstant, les résultats expérimentaux encourageants obtenus avec l'utilisation des cartes auto-organisatrices (Kohonen), une amélioration de l'efficacité est possible en affinant les paramètres de l'algorithme. Par ailleurs, dans la même optique, une amélioration du modèle proposée est à l'étude. Cette amélioration vise à optimiser l'initialisation des cartes et à améliorer la convergence du modèle dans notre contexte et ce, à l'aide d'une hybridation avec l'algorithme "firefly". Enfin, s'agissant du schéma de chiffrement proposé, il reste à le faire évoluer, afin de prendre

en charge le déchiffrement au niveau des dispositifs à contraintes, en vue d'une plus grande généralisation.

Mots clés : Données personnelles, Internet des Objets, Chiffrement basé sur les attributs, Gestion de la confiance, Externalisation des traitements.

Remerciements

Avant tout, je souhaite remercier très respectueusement Monsieur Abdelmadjid BOUABDALLAH, Professeur à l'Université de Technologie de Compiègne et Madame Maryline LAURENT, Professeure à Télécom SudParis, qui ont pris de leur temps pour rapporter mon mémoire, ainsi que Messieurs Pierre PARADINAS et Meziane YACOUB, du Conservatoire National des Arts et Métiers (CNAM), qui ont bien voulu faire partie de mon jury de soutenance.

J'adresse mes remerciements les plus vifs à Madame Samia BOUZEFRANE, ma directrice de thèse et Madame Hanifa BOUCHENEB, ma co-encadrante, qui m'ont accompagné durant cette thèse. Au delà du travail scientifique accompli durant ces quatre années, j'ai pu apprécier leurs qualités humaines et professionnelles. Ce fut un vrai plaisir de travailler avec elles.

Je remercie également toutes les personnes, qui ont contribué à l'accomplissement de ma thèse : Monsieur Soumya BANERJEE, Maître de conférences invité au CNAM et Monsieur Meziane YACOUB, qui n'ont pas hésité à partager avec moi leur savoir.

Par cette même occasion, je tiens à exprimer toute ma gratitude à Monsieur Mohamed KAIDI pour le soutien qu'il a apporté à la concrétisation de ce projet.

Je ne pourrais pas finir ces remerciements sans en adresser les plus chaleureux à toute ma famille, ma femme, mes enfants, mes parents et mes sœurs, pour leur soutien continu et leurs encouragements.

Enfin, une pensée aux miens qui nous ont quittés durant l'accomplissement de cette thèse, Ma grand mère Ghnima, mon oncle Ouaamar et Da Ali.

A mes grands pères Youcef et Arezki.

Table des matières

Liste des tableaux

Table des figures

Première partie

Introduction

Introduction générale

1.1 Contexte et problématique

De nos jours les systèmes d'information basés sur les nouvelles technologies d'information et de communication sont omniprésents dans notre quotidien. On parle alors d'informatique ubiquitaire. Ainsi, ont émergé des techniques qui rendent le déploiement de tels systèmes de plus en plus fonctionnel, à travers notamment le développement de l'Internet des objets (qui facilite la collecte et la restitution des données) et des services d'externalisation des capacités de calcul et de mémoires des équipements à contraintes physiques. Cette extension capacitaire est offerte via le modèle du Cloud computing et plus récemment via les nouveaux paradigmes de Skin, Fog et MEC (pour Mobile Edge Computing), qui améliorent les performances en rapprochant ces services de l'utilisateur final. Ces techniques offrent la possibilité de déployer de nouvelles applications de santé (ECG, mouvements, taux de glycémie, etc.), de villes intelligentes (gestion des transports, de l'énergie,etc.), de l'agriculture de précision, etc.

Néanmoins, la prolifération de telles applications qui manipulent nos données personnelles met en péril la préservation de notre vie privée. De plus, l'utilisation de l'Internet des Objets introduit des défis pour la mise en œuvre d'outils de sécurité traditionnels (chiffrement, PKI, etc.) alors que l'externalisation de nos traitements pose le problème de la confiance. En effet, en faisant appel aux fournisseurs de services pour traiter et stocker ses données, l'utilisateur perd le contrôle effectif sur elles. Au final, nous devons nous fier à la capacité du fournisseur de services d'assurer une protection adéquate de nos données et dans un cas extrême à son honnêteté. Cela, dans un monde où les données personnelles sont une ressource monnayable.

Ce contexte impose la mise en place de solutions de protection adéquates et une gestion de la confiance efficace, en réponse aux menaces et risques encourus par la compromission des fournisseurs de service ou à leur malhonnêteté éventuelle. Malheureusement, du point de vue de la protection de données et de la vie privée, nous constatons que beaucoup de solutions proposées dans la littérature, tendent à idéaliser certains composants et acteurs des systèmes au point de s'éloigner de la réalité du monde. À ce sujet, Gope et Hwang [44] référencent plusieurs exemples d'applications et de projets de recherche sur la e-santé où la sécurité présente des lacunes ou bien est laissée comme un élément à développer ultérieurement.

Ce travail de thèse porte sur le développement d'un modèle pour la protection des données et de la vie privée centré sur le propriétaire des données, le tout dans une architecture composée d'objets connectés et de dispositifs d'externalisation des traitements et de stockage.

Dans ce qui suit, la section 1.2 est consacrée au concept de l'Internet des Objets ainsi qu'aux paradigmes utilisés pour l'externalisation des traitements. Ces concepts et paradigmes seront illustrés par un cas d'utilisation du domaine de la e-santé, à chaque fois que ceci s'avère opportun. La problématique de recherche ainsi que les objectifs visés sont détaillés dans la section 1.3. La section 1.4 est dédiée aux contribution de ce travail et nous finirons par une conclusion.

1.2 Concepts mis en œuvre

1.2.1 Internet des objets (IdO)

Il est admis que Kevin Ashton est le premier à avoir utilisé le terme d'Internet des Objets en 1999. Selon l'International Data Corporation (IDC), l'Internet des Objets est un réseau de réseaux de terminaux identifiables (ou Objets) qui communiquent sans intervention humaine via la connectivité IP. Les objets connectés couvrent une large gamme d'équipements allant du capteur, peu coûteux, accomplissant une tâche spécifique (mono-tâche) aux smartphones embarquant divers capteurs (GPS, accéléromètre, etc.) et sur lesquels nous pouvons installer des applications variées. Les principales caractéristiques des objets connectés sont leur

limitation en matière d'énergie, de capacités de calcul et de mémoire, leur mobilité mais aussi leur capacité de communiquer entre eux et avec l'Internet.

Pour illustrer les fonctionnalités de l'Internet des Objets, nous avons opté pour les systèmes e-santé de surveillance de l'état de santé des individus (fréquence cardiaque, pression artérielle, mouvements, etc.), et ce, dans leur environnement direct. Cette surveillance s'effectue à travers des capteurs connectés via des liaisons de courte portée (Wifi, Bluetooth, ZigBee, 6LoWPAN, etc.), qui s'appuient sur un dispositif d'externalisation de traitement et sur le Cloud pour fournir les services de surveillance de santé. Ce type d'applications, compte tenu des données manipulées et des technologies employées, présente de fortes contraintes de types légal, éthique et techniques sur la protection des données personnelles et de la vie privée [108].

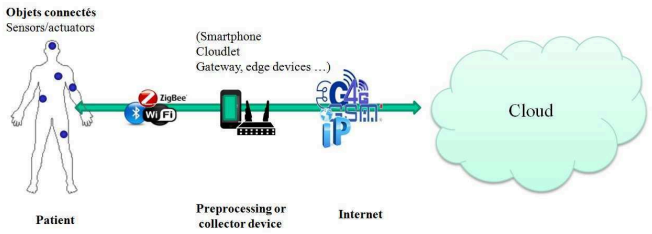


FIGURE 1.1 – Vue général d'un système IoT, exemple d'application e-santé

Une vue générale de cette architecture est donnée dans la figure 1.1. Elle est constituée de capteurs connectés, via une liaison sans fil de courte portée (Wifi, Bluetooth, NFC, ZigBee, etc.), à un équipement mobile (Smartphone, tablette, etc.) ou à une station de base qui joue le rôle d'intermédiaire entre le réseau local et le réseau Internet, et de là aux services du Cloud. Ce type d'architecture permet d'envisager diverses applications e-santé. Nous pouvons citer, à titre d'exemple, les travaux de Doukas *et al.* [37] qui ont démontré la facilité de mise en œuvre d'une telle architecture et proposé un prototype d'implémentation d'un système e-santé basé sur le Cloud Computing avec le service « Amazon S3 Cloud Storage » et un client mobile sous un système Android. Karunarathne *et al.* [65] ont proposé une architecture similaire, pour la surveillance des personnes prédisposées aux accidents cardiaux-vasculaires avec des capteurs de mouvements et d'accélération. Le tout via une application sur Smartphone et un service sur le Cloud pour le traitement de données et le

stockage.

Avec la prolifération des dispositifs de l'Internet des objets (IdO), caractérisés par des limitations en énergie, puissance de calcul et mémoire, conjuguées à des demandes croissantes en ressources des applications (réalité augmenté, surveillance...), il s'est avéré que les traitements liés à ces applications ne peuvent être effectués sur le dispositif lui-même. La solution viable est l'externalisation de certaines tâches vers des équipements n'ayant pas de contraintes de ressources [4]. Dans la suite, nous allons aborder certaines technologies qui permettent l'externalisation des services informatiques.

1.2.2 Externalisation des services informatiques (Computation Offloading)

Les années 2000 ont connu une évolution importante des systèmes informatiques vers la notion de service informatique, bien que les prémices de cette notion étaient déjà mises en œuvre dans les années 1950. Cette notion s'est concrétisée avec la possibilité de louer à la demande des capacités informatiques (capacités de calcul et de stockage aux logiciels) sur des serveurs centralisés et distants, accessibles via le réseau Internet. Alors que jusque-là, les systèmes informatiques étaient essentiellement construits autour de serveurs propriétaires des entités exploitantes. C'est le paradigme du Cloud computing ou l'informatique en nuage.

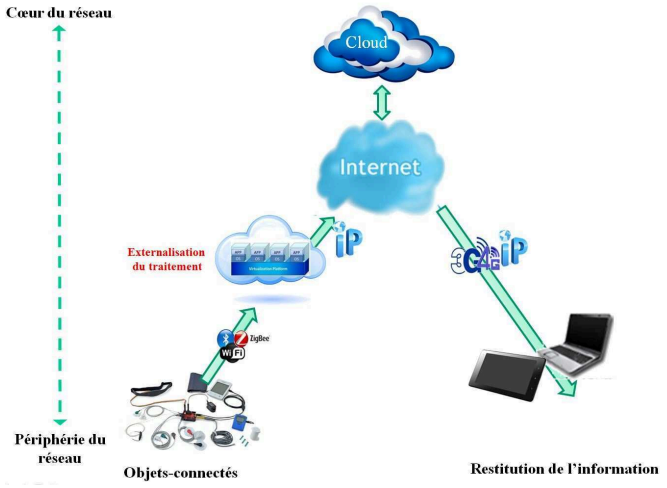


FIGURE 1.2 – Architecture IoT-Edge/fog-Cloud, application e-santé monitoring

Néanmoins, avec le besoin de traiter des données massives (big-data en anglais), de l'informatique ubiquitaire associée à l'Internet des objets, et en particulier avec des applications nécessitant une prise en compte du contexte, de la mobilité et des applications sensibles à la gigue et au temps de réponse du réseau, le modèle du Cloud computing est rendu inefficace. Cette inefficacité résulte principalement de la distance géographique entre les centres de données et les clients finaux. Pour remédier à cette situation, et prendre en compte les contraintes physiques des objets connectés, de nouveaux paradigmes ont fait leur apparition. Ces paradigmes ont en commun l'objectif de rapprocher les fonctionnalités du Cloud du client. Les plus couramment abordés dans la littérature sont le Cloud mobile (ou mobile Cloud computing en anglais) [143], l'informatique en brouillard (fog computing) [138] et l'infrastructure réseau des frontières (ou mobile edge computing en anglais) [136]. Un aperçu sur ces paradigmes qui permettent d'extérioriser les services informatiques (capacité de calcul, de stockage, plates-formes de développement, logiciels applicatifs, etc.) est présenté dans ce qui suit. La figure 1.2 schématise le fonctionnement global de cette architecture.

1.2.2.1 L'informatique en nuage (Cloud Computing)

La commission générale de terminologie et de néologie propose le terme français d'informatique en nuage et lui donne une définition officielle [22] : "Mode de traitement des données d'un client, dont l'exploitation s'effectue par l'Internet, sous la forme de services fournis par un prestataire." et note que c'est une forme de gérance de l'informatique "dans laquelle l'emplacement et le fonctionnement du nuage ne sont pas portés à la connaissance des clients."

Mais nous préférons employer par convenance l'anglicisme Cloud Computing, qui reste le plus usité dans la littérature.

Le Cloud Computing selon le National Institute of Standards and Technology (NIST), est un modèle qui permet un accès via un réseau omniprésent, pratique et à la demande à un ensemble partagé de ressources informatiques configurables (e.g., réseaux, serveurs, stockage, applications et services) qui peuvent être mises à disposition et libérées avec un minimum d'effort de gestion ou d'interaction avec le fournisseur de services.

FIGURE 1.3 – Services offerts par l'informatique en nuage

Le Cloud Computing, apporte une réponse viable aux challenges auxquels est confronté l'Internet des objets, notamment la croissance du volume de données, qui est accentuée par la prolifération de capteurs et objets connectés. A titre d'exemple, General Electric (GE) Healthcare estime qu'en moyenne, le monitoring d'un patient génère 1,5 GB de données par jour. L'utilisation de plus en plus répandue de terminaux mobiles (Smartphones, tablettes, etc.) requiert l'extension de leurs capacités limitées en utilisant celles du Cloud.

Le Cloud offre, comme nous pouvons le voir sur la figure 1.3, une variété de services d'infrastructure (IaaS pour Infrastructure as a Service) tels que des infrastructures réseau virtuelles, de plates-formes (PaaS pour Platform as a Service) tels que des systèmes d'exploitation ou encore de logiciels (SaaS pour Software as a Service) comme les applications d'authentification, de comptabilitité ou autre.

1.2.2.2 Traitement des données à la périphérie du réseau (Edge-computing)

Avec l'augmentation du nombre de dispositifs connectés et mobiles, le modèle centralisé du Cloud computing ne répond plus aux besoins de nombreuses applications en matière de qualité de service délivrée. En effet, la centralisation des traitements, même avec des ressources importantes, engendre des nœuds de congestion dans le réseau. Ce qui se répercute négativement sur la qualité de service délivré, en termes de temps de réponse

notamment. La solution apportée par le paradigme de Edge-computing, ou son extension mobile edge-computing (MEC), est de déployer de plus petites capacités de traitements qui soient le plus proche possible de l'utilisateur final. Ainsi, dans l'architecture MEC poussée par les réseaux mobiles de 5ème génération, les nœuds périphériques du réseau permettent de fournir des services informatiques, optimisant ainsi l'utilisation du réseau tout en améliorant la qualité de l'expérience de l'utilisateur final [4].

De manière générale, les nœuds définis dans les plates-formes MEC sont appelés "serveurs MEC". Mais, certains auteurs utilisent le terme de "Cloudlet" qui est plus générique. Bien que les serveurs MEC ou Cloudlets aient des capacités bien inférieures à celles offertes par le Cloud, ils fournissent une bien meilleure qualité de service [151] grâce à leur proximité de l'utilisateur final. Les équipements pouvant être utilisés comme frontières de réseau sont les routeurs MEC, les stations de base pour mobiles, les points d'accès WIFI et même les *box Internet*[1]. Ces équipements peuvent être dotés de capacités de virtualisation suffisantes pour fournir des services individualisés aux utilisateurs.

1.2.2.3 L'informatique en brouillard (Fog computing)

Le concept de Fog computing a été introduit par Cisco en 2012 pour répondre aux limites du Cloud Computing dans la prise en compte des nouvelles exigences liées aux applications de l'Internet des objets, notamment en termes de temps de réponse et d'efficacité réseau. Il consiste à mettre à disposition des utilisateurs des capacités informatiques situées sur la frontière du réseau entre le dispositif client final et le Cloud traditionnel, de façon complémentaire à ce dernier.

La définition donnée par Bonomi *et al.* [20] stipule que le fog computing est "une plateforme hautement virtualisée qui fournit des services de calcul, de stockage et de réseau, localisés entre les terminaux et les centres de données traditionnels du Cloud Computing".

L'informatique en brouillard permet de disposer d'architectures hiérarchiques où les données collectées par les dispositifs connectés (capteurs et autres) sont traitées localement avant de subir un traitement plus global au niveau du Cloud (voir figure 1.4). Bonomi *et al.* [20] prévoient la création de nouvelles formes de concurrence, entre les fournisseurs

1. Passerelle domestique, servant d'interface entre une FAI et l'abonné.

de services, tout en assurant une certaine forme de coopération entre ces fournisseurs pour fournir un service global.

Certains auteurs confondent les deux concepts de *Fog* et d'*Edge computing*, car très proches l'un de l'autre. Néanmoins, contrairement aux nœuds de l'*Edge-computing*, le déploiement des nœuds du *Fog-computing* ne se limite pas à la périphérie du réseau, mais peut être effectué plus en profondeur du réseau (ex. routeurs du cœur du réseau, switchs WAN , etc.). Aussi, contrairement au *Edge-computing*, le *Fog-computing* peut facilement étendre les modèles de services (IaaS, PaaS, SaaS) offerts par le Cloud vers la périphérie du réseau [86].

Enfin, Mahmud *et al.* [86] considèrent le *Fog-computing* comme ayant plus de potentiel de développement pour les applications de l'Internet des objets par rapport aux autres paradigmes comparables.

La figure 1.4 permet de situer globalement les différents paradigmes utilisés pour déporter les traitements informatiques dans un environnement IdO.

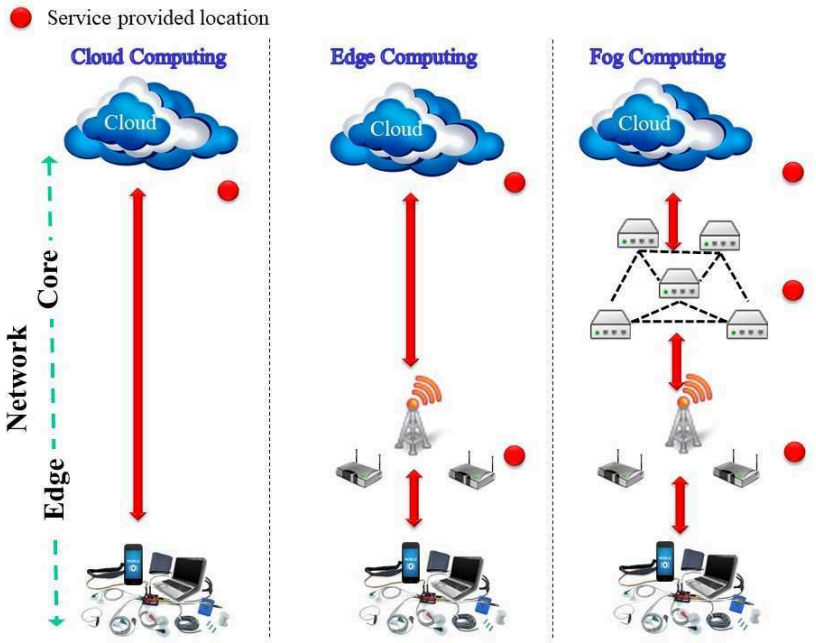


FIGURE 1.4 – Vues globales des paradigmes de déportation des traitements

1.3 Méthodologie et problématiques de recherche

A l'issue d'une recherche scientifique et technologique sur l'état de l'art de la sécurité des données dans un environnement IdO/Cloud, nous sommes arrivés aux conclusions suivantes : Le propriétaire des données doit avoir un contrôle effectif sur le processus de (i) chiffrement et (ii) de partage de ces données, indépendamment de leur localisation. Considérant ces deux objectifs, nous avons opté pour le chiffrement par attributs, qui assure les deux fonctionnalités simultanément. Néanmoins, les contraintes liées aux limites de capacité des objets connectés diminuent le champ d'application de ce type de chiffrement. Ceci nous a amenés à considérer l'externalisation de certaines tâches du chiffrement. Dans cette approche nous devons veiller à réduire au maximum la charge de calcul et d'utilisation de la mémoire des objets connectés tout en répondant aux exigences de protection des données personnelles.

Le choix de l'externalisation est conforté, en particuler, par les avancées techniques notamment dans la virtualisation et les réseaux 5G. Cependant, l'externalisation des tâches en général et des processus de sécurisation en particulier nécessite, au sus des techniques d'authentifications des tiers fournisseurs de service, une garantie sur la qualité de service qui sera délivrée et la confiance dans leur comportement futur. Cet aspect peut être traité à travers les mécanismes de gestion de la confiance.

Ceci étant, nos travaux de recherche se sont articulés autour de deux problématiques complémentaires de la sécurité des systèmes considérés dans notre thèse : (i) la sécurité cryptographique des équipements à fortes contraintes de ressources et (ii) la sécurité collaborative des systèmes à multi-offres de services.

La première problématique est liée à la sécurité cryptographique dans les équipements connectés à fortes contraintes de ressources que représente le monde de l'Internet des objets. Cet axe de recherche est primordial, quand la mise en œuvre des solutions de chiffrement classique se heurtent aux limites des ressources des équipements (énergie, bande passante, mémoire, puissance de calcul, etc.). Ce manque de ressources induit la nécessité de développer une solution de chiffrement capable de répondre aux exigences de sécurité (confidentialité, intégrité et disponibilité). A ceci, nous ajoutons une nouvelle exigence qui

est le maintien sous le contrôle du propriétaire de données du processus de sécurisation de ces dernières. Cette exigence découle de l'existence de processus légitime dit honnête mais curieux. Ce type de processus peut avoir des comportements malveillants et/ou malicieux tout en respectant les protocoles en vigueur.

La deuxième problématique de recherche est inhérente aux systèmes évoluant dans un environnement à multi-fournisseurs de ressources et services. Avec la multiplication des offres de Cloud computing et l'émergence des nouveaux concepts de *Edge/fog computing* que les opérateurs de télécommunication et les fournisseurs d'équipements ne manqueront pas d'investir, il est nécessaire de pouvoir évaluer la confiance envers les fournisseurs potentiels. En effet, si les techniques de chiffrement assurent l'authentification et l'authenticité du fournisseur de service, ils n'offrent aucune garantie sur son comportement futur en termes de sécurité et de qualité de services. La sécurité collaborative s'appuie sur la capitalisation des informations issues des expériences des utilisateurs pour donner une évaluation plausible des fournisseurs de services.

1.4 Contribution

Proposer un modèle de sécurité des données et de protection de la vie privée pour l'Internet des objet est en soi un exercice difficile, eu égard aux contraintes de ressources à prendre en compte. Cet exercice est rendu d'autant plus difficile que les applications IdO s'appuieront de plus en plus sur des infrastructures d'externalisation des services informatiques et les problèmes de confiance inhérents. Force est de constater que peux de travaux dans la littérature traitent les deux aspects simultanément.

Dans cette thèse, nous proposons un modèle de sécurisation des données personnelles centré sur le propriétaire des données et sur la confiance collaborative. Notre contribution se résume en trois points principaux :

— **1ère contribution :** Dans cette thèse, nous avons mis en évidence la nécessité que le propriétaire des données contrôle le processus de chiffrement et de distribution de ces données. Pour répondre à cet impératif, nous avons identifié un schéma de chiffrement prometteur, qui est le chiffrement basé sur les attributs. Ce schéma a

nécessité une adaptation mathématique pour son utilisation dans le contexte de l'Internet des objets. Cette adaptation a été suivie d'une preuve mathématique de la sécurité du schéma et de tests d'évaluation des performances effectués sur une plate-forme physique réelle.

— **2ème contribution :** Le nouveau schéma de chiffrement par attribut a été intégré à un protocole de sécurité, qui garantit la confidentialité et l'intégrité des données personnelles de l'utilisateur, indépendamment des fournisseurs de services de calcul et de stockage. Ceci, en considérant les fournisseurs de service externes comme des entités honnêtes mais curieuses et en exploitant les possibilités offertes par la technologie des Blockchains, en qualité d'entité de confiance décentralisée.

— **3 ème contribution :** Utilisant une architecture qui fait appel aux fournisseurs de services externes, à travers les paradigmes d'externalisation, il était nécessaire de compléter le modèle à l'aide de l'évaluation de la confiance. Pour cela, nous avons développé une approche basée sur des outils de l'intelligence artificielle et s'appuyant sur le concept de filtrage collaboratif. Notre approche a été testée sur des données réelles et sous des contraintes fortes, en termes de disponibilité des données et de présence d'utilisateurs malveillants dans le système.

1.5 Organisation du document

Notre thèse est organisée en trois parties ainsi qu'une conclusion :

— Première partie : est une introduction à la thèse avec la présentation du contexte et de la problématique traitée. Les notions qui seront utilisées dans ce manuscrit ainsi que la méthodologie de recherche seront y présentées. Cette partie intègre également un état de l'art de la sécurisation des données personnelles dans un environnement IdO/Cloud qui vise l'externalisation des services informatiques.

— Deuxième partie : est consacrée à notre première et deuxième contributions, à savoir, une solution de sécurisation de données personnelles dans un environnement IdO/Cloud à confiance limitée.

— Troisième partie : présente notre troisième contribution qui étend la précédente en

prenant en compte la gestion de la confiance dans un environnement IdO/Cloud.

— Conclusion et perspectives : ce manuscrit se termine par une conclusion, qui rappelle, dans un premier temps, les objectifs de cette thèse. Elle résume ensuite les principales contributions et leurs intérêts. Elle termine par les perspectives envisagées à ce travail.

Chapitre 2

État de l'art de la sécurité des données dans un environnement Internet des Objets et Cloud

2.1 Introduction

L'Internet des Objets et le Cloud sont des environnements hétérogènes qui rendent le développement de solutions difficilement réalisable, de bout en bout, pour sécuriser les données et protéger la vie privée tout en assurant les fonctionnalités de collecte, de stockage, de traitement et de partage de données.

En effet, du côté de l'IdO, le manque de ressources en termes d'énergie, de mémoire et de capacité de calcul, qui caractérise les objets connectés, limite le déploiement des solutions de sécurité traditionnelles. Alors que du côté du Cloud, les ressources sont relativement illimitées. Néanmoins, l'utilisateur peut être confronté à un problème de confiance, dans le cas où le Cloud est géré par un tiers [113].

2.1.1 Impact de l'hétérogénéité de l'environnement sur la sécurité des données

L'analyse des solutions proposées dans la littérature montre qu'elles sont confrontées à la sécurisation des données dans un environnement IdO/Cloud qui est hétérogène. Par exemple, Pussewalage et Oleshchuk [113] et Li *et al.* [78] donnent une solution intéressante pour le contrôle d'accès dans l'environnement Cloud, mais elle ne prend pas en charge les

contraintes de l'Internet des Objets. Par contre, Gong *et al.* [43] et Khemissa et Tandjaoui [67] proposent de sécuriser les données et de protéger la vie privée dans l'environnement IdO sans pour autant proposer de solutions pour le partage de données et le contrôle d'accès au Cloud. Ces exemples sont une bonne illustration de la difficulté engendrée par l'environnement IdO/Cloud pour développer une solution de sécurité générale et viable. Dans ce contexte, l'adaptation de solutions existantes (notamment celles qui sont issues des réseaux de capteurs sans fil) avec un minimum d'innovation, à des cas d'utilisation précis, est plus rentable que de chercher à concevoir une solution innovante de bout en bout. Néanmoins, nous constatons que dans ces travaux, on impose des hypothèses fortes sur l'environnement au risque de s'éloigner du monde réel. Par exemple, Sharma *et al.* [128] proposent une construction dans l'environnement IdO/Cloud qui peut être implémentée pour le contexte de la santé. Dans cette proposition, pour la sécurité de la transmission, le protocole Datagram Transport Layer Security (DTLS) est utilisé. Néanmoins pour la sécurité du stockage des données et le contrôle d'accès, il est nécessaire d'avoir un Cloud privé totalement sécurisé. Le même type de solutions est proposé par Abawajy et Hassan [2] et Abdmeziem et Tandjaoui [3].

Dans les solutions citées ci-dessus, nous constatons que le problème du contrôle d'accès se pose au niveau du Cloud qui doit être de confiance. Ainsi, il existe des solutions qui proposent de résoudre ce problème. L'approche proposée par Li *et al.* [78] est basée sur le chiffrement par attributs (ABE) pour contrôler l'accès aux données de santé hébergées dans le Cloud. Cette solution a la particularité de renforcer la protection de la vie privée en réduisant les privilèges de l'hébergeur des données. Ce type de contrôle d'accès est centré sur le patient, contrairement aux approches traditionnelles. Cependant, la solution proposée n'élimine pas complètement l'autorité de confiance pour la gestion du contrôle d'accès et ne prend pas en charge les ressources limitées des objets connectés. En effet, l'implémentation d'ABE, en l'état actuel, nécessite de disposer d'un certain niveau de ressources [140], qui n'est pas toujours compatible avec les ressources limitées des objets connectés.

2.1.2 Externalisation des services : solution ou problème ?

A défaut d'augmenter les capacités physiques des objets connectés et/ou d'améliorer l'efficacité des algorithmes et primitives (au sens large), la solution au problème de ressources disponibles est la déportation ou l'externalisation des calculs lourds. L'externalisation des traitements est une solution viable, notamment avec les nouveaux paradigmes d'informatique en nuage, en brouillard, etc. [64] permettant, ainsi, d'enrichir les fonctionnalités offertes par l'IdO. Pour le chiffrement des données, de nombreux auteurs comme Asim *et al.* [9] et Wang *et al.* [142] proposent des solutions basées sur ce principe d'externalisation vers un tiers.

Cependant, le mécanisme d'externalisation pour l'IdO introduit de nouveaux défis en matière de sécurité mais également de protection de la vie privée. Roman *et al.* [116] listent plusieurs menaces pouvant impacter la sécurité de l'écosystème IdO-technologie d'externalisation. Certaines de ces menaces, comme la possibilité de déploiement de services malicieux et la curiosité des acteurs "honnêtes", nécessitent des mécanismes de gestion de la confiance. Il existe très peu de travaux scientifiques dédiés explicitement à la gestion de la confiance dans le contexte de l'informatique en brouillard ou d'infrastructures réseau de frontière. Néanmoins, des travaux similaires utilisés dans d'autres domaines comme ceux menés dans le Cloud mobile, pourront facilement s'adapter aux architectures Fog ou MEC.

Ceci nous amène, comme nous l'avons mentionné dans notre introduction 1.3, à notre seconde problématique de recherche à savoir la gestion de la confiance et le problème de sélection des fournisseurs d'externalisation des services informatiques. La problématique de la sélection des services et de l'évaluation de la confiance de ces services a été largement étudiée. Plusieurs approches ont été proposées dans la littérature comme la maximisation (ou la minimisation) des fonctions d'utilité (approches basées sur le contenu) [18], l'utilisation d'outils issus de l'intelligence artificielle comme la logique floue [96] et les réseaux de neurones artificiels [148], ainsi que les approches qui se basent sur la mise en commun des expériences des utilisateurs (filtrage collaboratif) [154].

En résumé, dans un contexte d'externalisation des services informatiques, en sus des solutions de sécurité traditionnelle (cryptographiques), la protection des données

personnelles nécessite d'avoir une certaine visibilité sur le comportement futur du fournisseur de service. Cette problématique peut trouver des réponses à travers les systèmes d'évaluation de la confiance et de sélection de services.

2.1.3 Exigence de sécurité - cas de la e-santé -

Au regard de la nature ubiquitaire de l'IdO dans une multitude de domaines (ville intelligente, santé, voiture autonome, etc.) et leur corollaire de technologies qui accompagnent la prolifération des objets connectés (Cloud, Fog, etc.), il est difficile de cerner les exigences de sécurité de cet environnement, sans en spécifier le périmètre. A ce titre, dans la revue de littérature sur la sécurité de l'IdO de Kouicem *et al.* [71], les auteurs ont pris la précaution de présenter les exigences de sécurité propres à certains domaines d'application (santé, transports, villes intelligentes, etc.). Pour ce qui nous concerne, nous avons opté pour le cas d'utilisation relatif à la santé. D'après Kouicem *et al.* [71] les exigences de sécurité pour les applications de e-santé sont :

— L'authentification : l'accès aux données ne doit être effectué que par les entités identifiées et autorisées à y accéder.

— La confidentialité : le contenu des données personnelles ne doit pas être révélé à un tiers non autorisé.

— L'intégrité : Le contenu des données ne doit pas être altéré.

— Protection de la vie privé : les données ne doivent révéler aucune information sur le patient (pathologie, localisation, etc.).

Le choix du domaine de la santé, sans être exclusif par rapport à d'autres cas d'utilisation, est motivé par sa représentativité en termes de contraintes fortes sur la protection des données personnelles. De plus, comme nous l'avons démontré dans [108], la protection de ces données est un enjeu majeur pour la protection de la vie privée.

Dans la suite, nous allons présenter les deux dimensions qui doivent être prises en compte, à notre sens, pour assurer la protection des données privées dans l'écosystème IdO et déportation des services informatiques (e.g. Cloud-computing et Fog-computing). La première dimension est la dimension cryptographique et la deuxième est la gestion de la confiance.

2.2 Le chiffrement

Le paradigme de chiffrement a pour vocation d'assurer la confidentialité et l'authenticité des données. Jusqu'à récemment, le schéma traditionnel des primitives de chiffrement consistait à assurer la sécurité des échanges point à point, entre deux entités, par exemple, Alice et Bob. Actuellement, une évolution du paradigme s'est effectuée pour répondre aux nouveaux besoins, notamment avec le développement des services d'externalisation de l'informatique sur le *Cloud*. En effet, le *Cloud* permet de traiter et de stocker des données, dans des fermes de serveurs sur lesquelles les utilisateurs n'ont aucun contrôle ni de réelle garantie de sécurité. Une situation similaire peut être observée dans l'environnement *Fog* ou *Edge computing*. Parmi les nouveaux besoins induits par ce type d'environnement, nous pouvons citer le contrôle total de la diffusion des données personnelles par le propriétaire de ces dernières et les traitements sécurisés sur celles-ci. Par contrôle total et traitement sécurisé, nous entendons notamment, l'interdiction aux fournisseurs de service informatique d'avoir connaissance des données en clair. De nouvelles solutions cryptographiques apportent des réponses viables à ces besoins comme l'utilisation du chiffrement par attributs pour répondre au besoin du contrôle d'accès aux données et le chiffrement homomorphique pour le calcul sécurisé sur celles-ci.

Dans la suite, nous allons donner un aperçu des principaux schémas de chiffrement traditionnel et de ceux qui sont émergents. Comme nous l'avions noté dans la section 1.3, dans notre travail, nous avons opté pour le développement d'une variante du chiffrement par attributs adaptée à l'Internet des objets. C'est la raison pour laquelle nous focaliserons sur l'utilisation de ce schéma dans la littérature en vue de sécuriser les données personnelles dans un environnement IdO. Nous nous limiterons à ce périmètre car il présente une forte contrainte en termes de disponibilité de ressources nécessaires à l'implémentation de solutions robustes de sécurité.

2.2.1 Chiffrement symétrique (chiffrement à clé privée)

Un schéma de chiffrement à clé secrète ou chiffrement symétrique ε est un modèle de chiffrement où une clé secrète est partagée entre Alice et Bob (d'où le nom de chiffrement

à clé secrète) et est utilisée pour le chiffrement et pour le déchiffrement (d'où le terme de symétrique). La définition la plus commune dans la littérature est la suivante :

Définition 1

Le chiffrement symétrique avec le paramètre de sécurité k est défini par un doublé (Enc, Dec) de deux algorithmes déterministes : chiffrement, $Enc(K, M) \longrightarrow C$ et déchiffrement, $Dec(K, C) \longrightarrow M$. Où M est un message pris de l'espace des messages $\mathcal{M}$, C est un chiffré de l'espace des chiffrés $\mathcal{C}$ et K est la clé secrète de l'ensemble des clés $\mathcal{K}$. Cet ensemble est généré à partir du paramètre de sécurité k : $\mathcal{K} = \{0,1\}^l$, avec l un entier généré à partir de k.

L'avantage des primitives de chiffrement symétrique, e.g. l'AES (Advenced Encryption Standard) ou le 3DES, est leur efficacité (temps de calcul). Néanmoins, leur principal handicap est la nécessité de mettre en place un mécanisme de gestion des clés secrètes.

Pour l'Id0, des travaux récents proposent d'implémenter le chiffrement symétrique. Seulement, ces travaux ne prennent pas en considération la contrainte de passage à l'échelle de l'écosystème IdO. Une solution traditionnelle est décrite par Henriques et Vernekar [53] qui proposent d'utiliser le chiffrement symétrique et asymétrique (que nous allons décrire ci-après) pour sécuriser l'échange de données dans l'IdO. La combinaison du chiffrement symétrique et asymétrique permet de réduire le temps de chiffrement/déchiffrement, en chiffrant le message avec un chiffrement symétrique. La clé du chiffrement symétrique est transmise en même temps que le texte chiffré, elle même chiffrée avec RSA. Aussi, ils proposent d'utiliser des clés aléatoires différentes pour chaque message, réduisant ainsi la surface de vulnérabilité et par la même occasion permettra de résoudre le problème de la distribution des clés de session. Le même schéma est donné par Chandu *et al.* [24] qui proposent d'utiliser AES pour le chiffrement des données, et RSA pour l'échange des clés via le courrier électronique. L'implémentation d'AES est réalisée sur FPGA pour plus d'efficacité. Cette solution est applicable dans des scénarios bien spécifiques, bien que l'utilisation du RSA ne résout pas certains problèmes organisationnels de partage de données comme le ferait le chiffrement par attributs. S'agissant de l'implémentation matérielle, la proposition de Sung *et al.* [134] reprend cette idée pour l'AES-GCM dans un environnement IdO.

L'utilisation du chiffrement symétrique reste avantageux pour les objets connectés à ressources limitées, sous réserve d'un schéma de déploiement des clés maîtrisable en terme d'approvisionnement des équipements (problème d'échelle). Par ailleurs, il faudra veiller à ce que la compromission d'un nœud ne risque pas de compromettre tout le réseau.

2.2.2 Chiffrement asymétrique (chiffrement à clé publique)

Malgré les performances appréciables du chiffrement symétrique, sa principale limite réside dans la nécessité de sécuriser le partage de la clé secrète. Cette limite est accentuée dans les applications où on a besoin de faire communiquer un grand nombre d'acteurs. Dans cette optique est apparu le modèle à clé publique ou le chiffrement asymétrique. Whitfield Diffie et Martin Hellman sont considérés comme les premiers à proposer le principe de la cryptographie à clé publique dans les années 1970. Parmi les schémas les plus connues, nous trouvons l'algorithme RSA, dont la sécurité est basée sur la difficulté de factoriser les grands nombres entiers et l'algorithme El Gamal, dont la sécurité est basée sur le problème du logarithme discret.

Dans un chiffrement asymétrique, chaque entité possède une paire de clés (publique et privée). Nous pouvons définir le chiffrement asymétrique comme suit.

Définition 2

Un schéma de chiffrement asymétrique ε est défini pour un paramètre de sécurité k par un triplé $(Enc, Dec, KeyGen)$ comme suit :

- *$KeyGen(k) \rightarrow (PK, SK)$. C'est un algorithme aléatoire, qui prend en entrée un paramètre de sécurité k et génère aléatoirement en sortie une clé publique PK et une clé secrète SK.*
- *$Enc(PK, M) \rightarrow C$. C'est un algorithme deterministe, qui à partir d'un message $\mathcal{M}$ et d'une clé de chiffrement publique PK, génère un chiffré C. Avec M un message de l'espace des messages $\mathcal{M}$ et C un chiffré de l'espace des chiffrés $\mathcal{C}$.*
- *$Dec(SK, C) \rightarrow M$. L'algorithme de déchiffrement prend en entrée une clé secrète de déchiffrement SK et un chiffré C et retourne le message claire correspondant M, si la clé SK est correcte.*

Le principal handicap de ce type de chiffrement est le coût calculatoire qui le rend difficilement applicable sur des objets connectés, à capacité limité en énergie, mémoire et de calcul. Ceci étant, malgré cet handicap, il existe dans la littérature des propositions utilisant l'approche asymétrique pour l'Internet des objets. Ces approches utilisent les techniques de chiffrement basées sur les courbes elliptiques (ECC), dont la sécurité repose sur le problème du logarithme discret elliptique ou sur le chiffrement NTRU, proposé par Hoffstein *et al.* [56] et dont la sécurité repose sur le problème du plus court vecteur non nul d'un réseau euclidien.

Le chiffrement NTRU est une alternative à RSA et au chiffrement par courbes elliptiques (comparés à des niveaux de sécurité similaires) [100], vu sa complexité calculatoire qui est de l'ordre de $O(n^2)$, pour un message de taille n (de l'ordre de $O(n^3)$ pour les courbes elliptiques). Les auteurs de NTRU le décrivent comme une primitive "qui présente des clés raisonnablement courtes, faciles à générer, rapides et nécessitant peu de mémoire". Concernant l'écosystème IdO, Guillen *et al.* [46] analysent la faisabilité du chiffrement NTRU sur les équipements contraints et concluent à la viabilité de ce schéma pour les équipements présentant des contraintes de ressources.

Néanmoins, le schéma à clé publique nécessite de sauvegarder en mémoire toutes les clés publiques, pré-distribuées, des entités communicantes. Ceci pose alors le problème de passage à l'échelle. Les auteurs de [103] proposent une infrastructure à clé publique avec la technologie TLS pour l'IdO. Ceci permet de s'affranchir du besoin de pré-distribuer les clés publiques. Par contre, la principale critique de ce travail est qu'il néglige l'aspect ressources disponibles sur les objets connectés.

L'utilisation du chiffrement à clé publique peut être envisagée dans des cas d'applications particuliers. Ce sont les cas où la pré-distribution des clés publiques ne pose pas de problème de mise en œuvre (e.g. nombre limité d'objets à administrer). Une autre approche qui permet l'emploi du chiffrement à clé publique est l'utilisation de passerelles sur lesquelles on peut déporter la charge de travail, comme présenté dans [107]. Dans le même esprit de Sung *et al.* [134], Johnston *et al.* [61] proposent l'utilisation d'un co-processeur de chiffrement (TPM pour Trusted Platform Module) pour les objets connectés. Le TPM permet d'assurer des fonctionnalités de chiffrement, de sécurisation des clés secrètes et d'attestation de l'état

initiale du matériel. Néanmoins, à l'instar des travaux qui proposent d'implémenter des solutions, initialement destinées à l'informatique traditionnelle, ces travaux ne semblent pas prendre en considération suffisamment les limitations physiques des objets connectés.

2.2.3 Chiffrement par attributs

Traditionnellement, un schéma de chiffrement comme RSA et AES fournit une transmission et un stockage de données sécurisés dans un environnement Cloud mais l'inconvénient de ces schémas est la difficulté de mettre en place un contrôle d'accès avec une granularité fine pour le partage des données, en particulier dans le cas où nous ne connaissons pas l'identité des utilisateurs au préalable. Une solution viable à ces problématiques est donnée par le chiffrement par attributs ou « Attribute Based Encryption » (ABE) [68], qui incorpore un processus de génération des clés de chiffrement et de déchiffrement et la notion de politique d'accès basée sur des attributs. Ce schéma offre simultanément des fonctionnalités de chiffrement et de contrôle d'accès [106].

Le chiffrement par attributs ABE a été introduit en 2005 par Sahai et Waters [120] comme une évolution du chiffrement basé sur les identités floues, lui même étant une amélioration du chiffrement basé sur l'identité. C'est un schéma de chiffrement à clé publique du type un-à-plusieurs, c'est-à-dire qu'on chiffre avec une seule clé et on a la possibilité de générer plusieurs clés pour déchiffrer. Un avantage évident de cette technique est que chaque utilisateur a une clé dédiée, en cas de révocation d'une clé, il n'est pas nécessaire de refaire le chiffrement de données. Les données peuvent être chiffrées à la source et entreposées telles quelles, et à aucun moment le fournisseur de service n'accède au clair. En plus de sécuriser la transmission et le stockage des données, ABE fournit un contrôle d'accès à forte granularité, une gestion évolutive des clés et une distribution flexible des données [52; 76; 79]. Il permet de chiffrer les données et d'assurer le partage sur la base d'attributs descriptifs, sans aucune connaissance préalable de l'identité des destinataires. Seules les entités avec des attributs qui satisfont une politique d'accès aux données peuvent déchiffrer un texte. Les deux principales variantes sont :

— *Key-Policy Attribute Based Encryption » (KP-ABE)* qui a été développée par Goyal *et al.* [45] en 2006. Pour KP-ABE, la politique d'accès est intégrée dans la clé secrète,

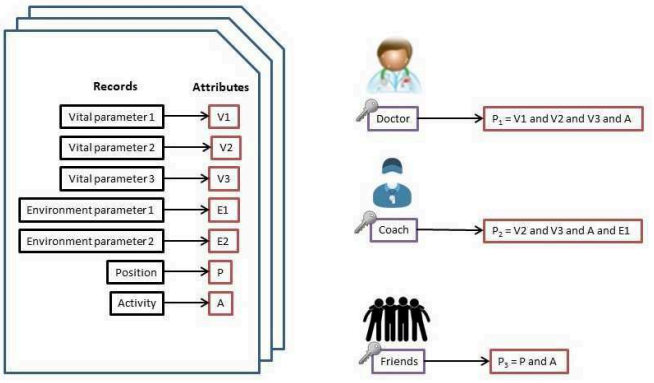


FIGURE 2.1 – Partage de données avec KP-ABE

en d'autres termes, on décide pour chaque utilisateur quels sont les objets auxquels il aura accès. On attache à chaque texte chiffré un ensemble d'attributs. Une clé secrète donnée, avec une politique d'accès donnée, ne peut déchiffrer que le texte chiffré ayant les attributs qui satisfont sa politique d'accès.

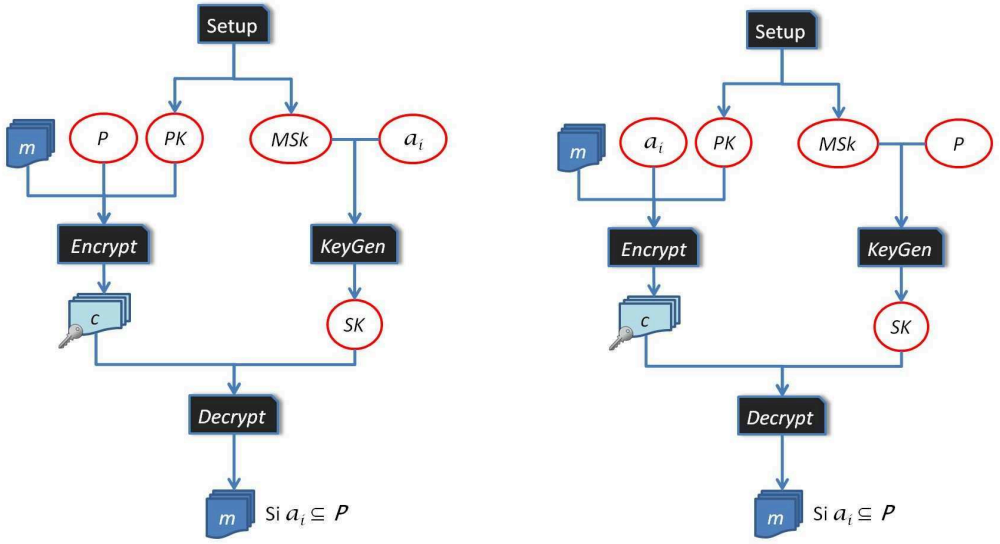
— *Ciphertext-Policy Attribute Based Encryption» (CP-ABE)* proposée pour la première fois par Bethencourt *et al.* [19] en 2007, dans laquelle la politique d'accès est intégrée dans le texte chiffré et les clés secrètes sont générées avec un ensemble d'attributs décrivant l'utilisateur légitime qui pourra déchiffrer ce texte. Seul les clés secrètes avec un ensemble d'attributs qui satisfait la politique d'accès peuvent récupérer le texte en clair.

Concernant les cas d'utilisation typiques, CP-ABE peut être implémenté pour permettre au propriétaire des données de définir la politique d'accès sur ses données, sans avoir à spécifier explicitement l'identité des utilisateurs au préalable. Alors que KP-ABE est utilisé par exemple pour partager les informations du journal d'audit. La Figure 2.1 illustre un exemple de partage de données avec la variante KP-ABE. Chaque enregistrement de données peut être lié à un attribut et une clé de déchiffrement sera liée quant à elle à une politique d'accès P_i qui, pour chaque utilisateur, détermine quels enregistrements peuvent être déchiffrés. Une revue d'applications possibles pour ABE est donnée par Balamurugan et Krishna [12].

Comme souligné par Lee *et al.* [76], un système basé sur ABE permet en théorie de

fournir les fonctionnalités suivantes : (i) la confidentialité des données, (ii) un contrôle d'accès avec une fine granularité, (iii) le passage à l'échelle, (iv) la révocation d'utilisateurs et (v) la résistance à la collusion entre utilisateurs.

Un chiffrement ABE est défini comme un quadruple de quatre algorithme (*Setup*, *Enc*, *Dec*, *KeyGen*), décrits dans le tableau 2.1, dont le schéma fonctionnel est donné dans la figure 2.2a pour CP-ABE et 2.2b pour KP-ABE. La principale différence est que dans CP-ABE, la politique d'accès est incluse dans le chiffré et les attributs sont inclus dans la clé de déchiffrement, alors que dans KP-ABE, c'est exactement l'inverse.



(a) CP-ABE (b) KP-ABE

FIGURE 2.2 – Schémas fonctionnels de CP-ABE et KP-ABE

2.2.3.1 ABE dans l'écosystème IoT

Les fonctionnalités d'ABE semblent être intéressantes pour une solution assurant la protection de la vie privée [140], notamment en implémentant un système de sécurité centré sur le propriétaire des données [113], [78] où le fournisseur de services ne pourra avoir accès aux données en clair. Ceci est possible grâce au chiffrement des données sous le contrôle du propriétaire des données tout en assurant, comme l'a relevé Ambrosin *et al.* [7], le partage de ces dernières avec une grande granularité.

TABLE 2.1 – Entrés/Sorties des algorithmes de CP-ABE

Algorithme	Description des entrées/sorties
$Setup$:	*Entrée :* Un paramètre de sécurité l. *Sortie :* Une clé publique de chiffrement Pk et une clé secrète principale MSk qui servira à générer les clés secrètes de déchiffrement.
Enc :	*Entrée :* Message à chiffrer m, la clé publique Pk et un ensemble d'attributs a_i dans le cas de KP-ABE ou une politique d'accès P dans le cas de CP-ABE. *Sortie :* Le chiffré c.
$KeyGen$:	*Entrée :* La clé secrète principale MSk et un ensemble d'attributs a_i dans le cas de CP-ABE ou une politique d'accès P dans le cas de KP-ABE. *Sortie :* une clé de déchiffrement secrète Sk, liée à un ensemble d'attributs a_i dans le cas de CP-ABE ou à une politique d'accès P dans le cas de KP-ABE.
Dec :	*Entrée :* Le chiffré c et une clé de déchiffrement Sk. *Sortie :* Si l'ensemble d'attributs a_i satisfont la politique P alors sortie m sinon $\perp$

Aussi, le chiffrement par attribut commence à être assez mature pour être incorporé comme module fonctionnel d'une solution de sécurité [106]. Cependant, ABE est une méthode de chiffrement coûteuse en terme de capacité de calcul. En effet, si elle ne pose pas de problèmes particuliers pour une utilisation dans le Cloud, ceci n'est pas le cas pour une implémentation sur les dispositifs ayant des contraintes de ressources [7].

Ainsi, les schémas de chiffrement basés sur les attributs sont généralement gourmands en calcul et ne sont pas adaptés aux équipements à ressources limitées, comme l'ont démontré Wang *et al.* [141], en évaluant l'implémentation d'ABE en Java sur un smartphone sous Android. Ambrosin *et al.* [7], quant à eux, ont réalisé une implémentation en C d'ABE également sous Android. Dans [6], la même évaluation est effectuée pour des dispositifs de l'IdO, pour un cas d'utilisation de la e-santé (surveillance à distance des patients). Dans leur expérimentation, les auteurs de [6] ont utilisé jusqu'à 10 attributs et un niveau de sécurité équivalent à 80 bits. La conclusion est que même si l'ABE est réalisable sur des dispositifs contraints, la faisabilité dépend fortement des exigences de l'application, principalement des niveaux de sécurité et du nombre d'attributs utilisé dans la politique d'accès.

L'adaptation du chiffrement par attribut à l'écosystème IdO nécessite de prendre en

compte plusieurs aspects liés aux limites des objets connectés (mémoire, processeur, énergie, bande passante). Par exemple, Guo *et al.* [47] ont proposé une solution adaptée à l'IdO mais qui se focalise uniquement sur la taille des clés à sauvegarder dans les dispositifs contraints. Négligeant de fait, le problème de la capacité de calcul, d'énergie et de bande passante. En ce qui concerne notre cas d'utilisation, qui est la surveillance à distance des patients, nous porterons notre attention sur l'efficacité du chiffrement de données qui est effectué par les dispositifs de collecte (capteurs connectés).

Le schéma commun proposé dans la littérature est d'avoir des périphériques limités en ressources qui implémentent un chiffrement léger et d'externaliser la lourde charge de calcul vers des périphériques non contraints. Ce type de solution est proposé par Touati *et al.* [137] où les périphériques contraints impliquent des voisins de confiance non soumis aux limitations de ressources pour prendre en charge la partie lourde du calcul ABE (exponentiation). Les auteurs proposent de scinder mathématiquement un paramètre secret en n parties et de partager des clés secrètes avec les n nœuds de confiance non contraints. Le principal inconvénient de cette solution est sa mise en œuvre dans un cas réel, en prenant en considération l'infrastructure nécessaire à la gestion des clés et de la confiance et leur impact sur la bande passante du réseau. La solution présentée dans Asim *et al.* [9] vise à externaliser les opérations de calcul lourd vers le Cloud. Cependant, cette solution nécessite l'accès à un tiers de confiance et doit établir et maintenir la connexion entre les périphériques de l'IdO et le proxy, ce qui peut générer une surcharge en matière d'échanges de messages. Quant à Nguyen *et al.* [101], ils proposent une modification de CP-ABE avec délégation d'une partie de la fonction de chiffrement. L'intérêt de ce travail est de considérer la partie délégataire comme honnête mais curieuse. L'analyse de sécurité présentéé montre que le schéma proposé est sûre sur la base de problème du logarithme discret. Enfin, les performances en termes de temps d'exécution et d'énergie consommée ont été évaluées sur une plate-forme simulant un objet ayant des contraintes.

L'utilisation du paradigme du Fog ou de Edge Computing est l'un des meilleurs moyens d'implémenter le chiffrement par attributs dans un écosystème IdO, en externalisant les opérations de calcul. Zhou et Huang [157] proposent un schéma CP-ABE préservant la confidentialité en externalisant des opérations de calcul lourd au fournisseur de services

de chiffrement. La complexité de l'algorithme de chiffrement n'est pas liée au nombre d'attributs utilisés dans la structure d'accès ; elle dépend toutefois du nombre d'attributs utilisés par le propriétaire de données pour chiffrer. A notre connaissance,au moment de la rédaction de ce manuscrit, Zhang *et al.* [153] proposent le schéma de chiffrement le plus efficace pour une utilisation dans un environnement IdO. Néanmoins, dans leur protocole, le propriétaire des données transmet d'abord au nœud de déportation la structure d'accès qui génère un cryptogramme intermédiaire qui est renvoyé au nœud de l'IdO pour calculer le cryptogramme final avant de le retransmettre vers le fournisseur de services de Cloud, ceci implique une plus grande utilisation du réseau et donc de consommation d'énergie pour le nœud de l'IdO. En outre, la longueur du texte chiffré générée par le nœud IdO dépend du nombre d'attributs utilisés dans la politique d'accès, ce qui entraîne une consommation de mémoire et d'énergie.

Enfin, il est utile de signaler qu'une autre approche est envisagée dans la littérature pour adapter le chiffrement par attribut à l'IdO. Cette approche propose de remplacer les opérations consommatrices relatives aux applications bilinéaires par des opérations sur les courbes elliptiques Yao *et al.* [149].

2.3 Systèmes de recommandation en tant que systèmes de gestion de confiance

Plusieurs définitions sont données dans la littérature pour la notion de confiance mais la majorité des auteurs semblent s'accorder sur celle de Gambetta [40] pour qui *"la confiance (ou, symétriquement, méfiance) est un niveau particulier de la probabilité subjective avec laquelle un agent estime qu'un autre agent ou groupe d'agents effectuera une action particulière, à la fois avant de pouvoir observer cette action (ou indépendamment de sa capacité à pouvoir le faire) et dans un contexte dans lequel il agit de son propre chef. "* Ainsi, le concept de confiance est subjectif et couvre les notions de prédiction [115], de croyance personnelle en quelque chose et d'attentes [58]. Et pour évaluer la confiance, nous utilisons des informations objectives, telles que la mesure de la qualité de service, et des informations subjectives, telles que les commentaires ou évaluations des utilisateurs [94].

Dans les systèmes basés sur les services, la gestion des relations de confiance est

intrinsèquement liée au système de recommandation car il s'agit d'une classe spécifique du système de relations de confiance [31]. Par ailleurs, la recommandation fait partie du processus de prise de décision [124], c'est une technique de fouille et de filtrage des informations qui vise à sélectionner un service pertinent pour l'utilisateur actif [41]. En outre, il fournit une prédiction sur le comportement futur du fournisseur de service, tels que la qualité de service délivrée.

Les systèmes de gestion de la confiance sont appliqués à différents contextes (Cloud computing, commerce électronique, P2P), où les fournisseurs sont en concurrence pour offrir des services similaires, mais moyennant des frais et un niveau de performance différents. Récemment, le même concept est devenu un sujet de recherche important pour les services d'externalisation à l'image du Cloud Computing [102]. L'intérêt des systèmes de gestion de confiance est de permettre de créer une relation de confiance entre l'utilisateur et le fournisseur de services [80]. Du point de vue du consommateur de service, l'évaluation de la confiance fournira une recommandation pour l'aide à la décision et un niveau d'assurance sur le comportement futur du fournisseur de service. Par conséquent, le consommateur peut adapter ses mesures de sécurité [152].

Dans le périmètre de ce travail (services d'externalisation pour l'IdO), nous adoptons indifféremment le terme "système de gestion de la confiance" ou "système de recommandation". En effet, le concept de système de recommandation étudié pour les différentes applications se prêtent facilement à une adaptation au paradigme d'externalisation des services pour l'IdO. Alors que plusieurs travaux, [23; 93; 119; 147; 148] , proposent des solutions de gestion de la confiance pour le paradigme de déportation des services informatiques en s'inspirant implicitement des systèmes de recommandation basés sur les retours d'expérience des utilisateurs ou sur le contenu des services.

Enfin, il est admis que la confiance est fortement liée à la notion de réputation. La réputation est définie par Mui *et al.* [95] comme *la perception qu'un agent crée à travers des actions passées sur ses intentions et ses normes.* Cette perception est basée sur les expériences passées. La réputation peut servir à évaluer la confiance [62], qui induit des suppositions et des attentes sur le comportement futur.

2.3.1 Les systèmes de recommandation

Le filtrage collaboratif (FC) a été largement adopté dans la sélection des services de confiance pour de nombreuses applications (par exemple, commerce électronique, services Web, systèmes Pair à Pair). Pour la sélection du service dans le Cloud, Longshuai *et al.* [83] utilisent le filtrage collaboratif pour fournir un service de recommandation de confiance. Pour résoudre le problème de parcimonie des données disponibles (data-sparcity), les auteurs proposent un algorithme de FC optimisé avec une décomposition en valeurs singulières (SVD) avant d'appliquer la similarité cosinus pour regrouper les utilisateurs. Balaji et Rajkumar [11] étudient un filtrage collaboratif hybride pour résoudre le problème de sélection du service dans le Cloud. Dans ce travail, la méthode de classification K-means des utilisateurs est associée au calcul de similarité avant d'évaluer la prédiction du paramètre cible et enfin, produire la recommandation.

Benouaret *et al.* [18] proposent une solution hybride basée sur l'approche collaborative et sur un score global calculé sur la base de la note moyenne des utilisateurs ainsi que d'une somme pondérée de différents indicateurs de QoS. Huang *et al.* [59] proposent une amélioration du coefficient de similarité de Pearson (PCC) et un service de recommandation combinant un modèle basé sur les utilisateur et les items. Akinola et Adigun [5] calculent la similarité dans un contexte de Cloud Computing mobile (MCC) et propose un modèle probabiliste de prédiction. Alors que Chen *et al.* [26] explorent le voisinage géographique pour calculer la similarité dans un sous-espace réduit et dense. Zheng *et al.* [155] associent l'approche basée sur les utilisateurs et celle basée sur les items dans un seul modele (WSRec), se qui améliore les performances, néanmoins sans toutefois garantir l'évolutivité de la solution.

Le principal problème du filtrage collaboratif est la rareté et le manque de données permettant de calculer une valeur de similarité précise. Pour illustrer ce problème, Hu *et al.* [57] ont mené une étude sur un ensemble de données réelles de recommandations (MovieLens) et ont conclu que le nombre de paires d'utilisateurs et de paires de films dont la similarité est supérieure à 0.1 est de 7.80 % du total des paires des utilisateurs et de 12.14 % du total des paire de films. En outre, une autre limitation de certains travaux, tels

que [27] et [94] est leur incapacité à fournir une prédiction personnalisée pertinente pour l'utilisateur actif. Un autre problème majeur, comme l'a noté Wang *et al.* [139], est que la majorité des propositions actuelles ne parviennent pas à détecter les évaluations aléatoires et malveillantes et supposent implicitement qu'un utilisateur a toujours un bon jugement sur un service, ce qui est irréaliste. Peu de travaux comme ceux de Kuang *et al.* [72] et Su *et al.* [131] prennent en compte les utilisateurs malveillants.

2.3.2 Les modèles d'évaluation de la confiance

D'autres travaux sont basés sur un modèle mathématique pour une évaluation globale de la confiance, tel que Saeed et Shaikh [119] qui calculent un score global avec une moyenne pondérée des paramètres indicateurs de confiance (e.g. temps de réponse) à partir de plusieurs sources de données (retour d'expérience des utilisateurs, commentaires de tiers-experts et de déclarations des fournisseurs). Le niveau de confiance est basé sur cinq attributs, à savoir : Disponibilité, Fiabilité, Intégrité, Confidentialité et Authentification. Challagidad *et al.* [23] propose un algorithme qui évalue la réputation du fournisseur de services en tenant compte des commentaires des clients, du taux de rejet des serveurs et de la charge de travail du serveur. Dans le travail de Mrabet *et al.* [93], la crédibilité des avis des utilisateurs est évaluée en utilisant une corrélation entre les valeurs de divers attributs relatifs à la qualité de service (QoS) rapportés par un utilisateur donné. Ces attributs (Disponibilité, fiabilité, efficacité temporelle et intégrité des données) sont utilisés pour calculer le niveau de confiance globale du service. Saied *et al.* [121], proposent un modèle de gestion de la confiance pour l'environnement coopératif de l'IdO. Dans la solution proposée, les évaluations données par les consommateurs du service sont accompagnées par un descriptif du contexte. Cette approche permet de prendre en charge l'hétérogénéité des noeuds de l'Internet des objets. La simulation de la solution montre l'efficacité de la proposition et sa résilience envers certaines attaques contre les système de gestion de confiance. Néanmoins, il aurait été intéressant de comparer ces performances avec d'autres solutions.

Ceci étant, les relations non linéaires entre les indicateurs de confiance du service et le niveau de confiance final du service ainsi que l'évaluation subjective des utilisateurs

donnent lieu à un problème complexe de sélection de services. Aussi, Wu et Zhou [147] proposent d'utiliser les réseaux de neurones et la logique floue pour traiter les informations subjectives issues de différents profils d'utilisateurs. Le réseau de neurones est utilisé pour apprendre les biais des préférences des utilisateurs sur différents aspects de la confiance. Ce biais appris est utilisé pour ajuster les fonctions d'appartenance et les règles d'inférence de la logique floue. Ainsi, le modèle de logique floue, adapté, prend en entrée les données de retour-d'expérience des utilisateurs et génère un score de fiabilité globale. Ce score de sortie est comparé à une valeur donnée par un tiers-expert. Si les deux valeurs sont proches l'une de l'autre, les règles floues et les fonctions d'appartenance seront conservées pour une utilisation ultérieure. Sinon, les règles et la fonction d'appartenance seront ajustées jusqu'à ce que les résultats soient cohérents. Pour la classification des services, Yang *et al.* [148] proposent un réseau de neurones qui extrait et organise automatiquement les informations discriminantes à partir de données, sans avoir besoin de connaissances préalables. La description du service correspond à l'entrée du réseau de neurones alors que sa classification est le résultat de ce réseau.

2.4 Contribution des blockchains à la protection des données

Motivé par l'intérêt récent de la communauté scientifique d'utiliser la blockchain pour améliorer la vie privée, nous présentons dans cette partie l'utilisation de ce paradigme, qui nous promet de se passer des tiers de confiance pour interagir de manière sécurisée dans un réseau pair à pair.

2.4.1 La blockchain

La Blockchain est un nouveau paradigme qui commence à être proposé par la communauté scientifique pour assurer le contrôle d'accès aux données. La technologie de la blockchain nous promet de se passer des tiers de confiance pour interagir de manière sécurisée dans un réseau qui ne l'est pas. La première description d'une blockchain se trouve dans un article écrit Satoshi Nakamoto [97], qui a donné naissance à la monnaie cryptographique Bitcoin. La blockchain est un registre publique infalsifiable et indestructible, disponible

dans un réseau pair à pair où chaque nœud peut s'ériger en une sorte de garant des informations enregistrées dans la blockchain, en effectuant un travail particulier pour lequel il sera rémunéré. Ces nœuds sont appelés mineurs. Dans la blockchain Bitcoin, chaque compte est caractérisé par une paire de clés publique/privée et identifié par pseudonyme dérivée de sa clé publique . Pour comprendre le fonctionnement de la blockchain, nous allons prendre l'exemple d'Alice qui veut transférer un montant de x Bitcoin à partir de l'un de ses comptes vers un compte appartenant à Bob. Alice va générer une transaction, qu'elle va signer avec la clé privée de son compte et qui contient comme information : les adresses des comptes source et destination, le montant du transfert et une référence à une ou plusieurs transactions précédentes qui ont servi à créditer le compte d'Alice (on n'a pas le droit au découvert). Cette transaction est partagée sur le réseau pair à pair, où elle est récupérée par les nœuds qui jouent le rôle de mineur. Ces derniers rassemblent plusieurs transactions pour constituer un bloc qu'ils valident mathématiquement avant de le chaîner à la blockchain à travers un processus de consensus spécifique entre les mineurs. Dans le cas de Bitcoin, ce processus est appelé preuve de travail. Une fois la transaction enregistrée dans la blockchain, elle devient publique, infalsifiable, irrévocable et indestructible. Ceci est garanti par la communauté du réseau et non par une autorité centrale de confiance. Une compréhension plus approfondie des principes de la blockchain est donnée dans Christidis et Devetsikiotis [29].

2.4.2 La blockchain et la sécurité

Comme il a été souligné par Gupta *et al.* [49], la résilience de la blockchain, son intégrité et l'absence d'autorité centrale de confiance (point critique en cas de défaillance) permettent de construire un service d'autorisation sécurisé. Ainsi, ces caractéristiques ont inspiré plusieurs travaux dans le domaine du contrôle d'accès aux données [10], [158] notamment dans l'environnement IdO et Cloud [51], [105] et [133]. Azaria *et al.* [10] proposent une solution qui gère le contrôle d'accès aux données de santé personnelles stockées dans différentes bases de données à travers des contrats intelligents embarqués dans une blockchain. Un autre apport de [10] est de proposer un modèle de rétribution des mineurs, basé sur Ether (la devise pour Ethereum), aspect très rarement évoqué dans la

littérature. Quant à Hashemi *et al.* [51], ils proposent une solution basée sur la blockchain, qui permet de gérer, par leur propriétaire, l'accès aux données produites par les objets connectés. Sukhodolskiy et Zapechnikov [133] proposent un modèle de contrôle d'accès aux données stockées sur un Cloud non fiable. Pour ce faire, ils proposent d'utiliser un protocole basé sur CP-ABE et d'enregistrer, de façon immuable, les événements de sécurité (génération de clé, révocation de droit ect.) sur une blockchain.

Dans le travail de Jeon *et al.* [60], l'infrastructure conventionnelle (e.x serveurs/MySql) utilisant une blockchain, est conçue comme une base de données. Quant à Singla et Bertino [129], ils proposent de remplacer l'autorité de certification (CA) dans l'infrastructure à clé publique par une blockchain. Ils ont démontré que l'utilisation de la blockchain est beaucoup plus efficace en termes de calcul et de stockage, en plus de fournir une infrastructure à clés publiques plus robuste et plus évolutive.

Le schéma général des solutions proposées est l'utilisation de blockchain pour la gestion du contrôle d'accès, alors que les données elles-mêmes sont manipulées hors-chaîne, en d'autres termes, sur les serveurs et Cloud dédiés à cela. L'intérêt de cette démarche est de séparer, comme l'illustre la figure 2.3, le plan de gestion du contrôle d'accès du plan de stockage, réduisant ainsi les privilèges du fournisseur de service de stockage au profit d'une autorité décentralisée, incarnée par la blockchain. Un autre intérêt de l'utilisation des blockchains est la protection de la vie privée à travers l'utilisation de pseudonymes dans les transactions, tout en assurant la sécurité de ces dernières (processus de signature/vérification) [30].

2.4.3 Mise en œuvre dans la pratique

Ceci étant, dans le monde réel, une attention particulière doit être accordée à certains aspects de déploiement des blockchains. D'abord le passage à l'échelle, par exemple dans [51] il est fait état d'une solution qui s'adapte à de grandes échelles en termes de nombre d'objet connectés. Mais comme le souligne Herrera-Joancomartí et Pérez-Solà [55], l'une des faiblesses actuelles de la blockchain est le nombre de transactions possibles par unité de temps (actuellement au maximum 7 transactions par seconde pour la blockchain Bitcoin).

Un autre aspect à prendre en compte est la préservation de la vie privée en évitant

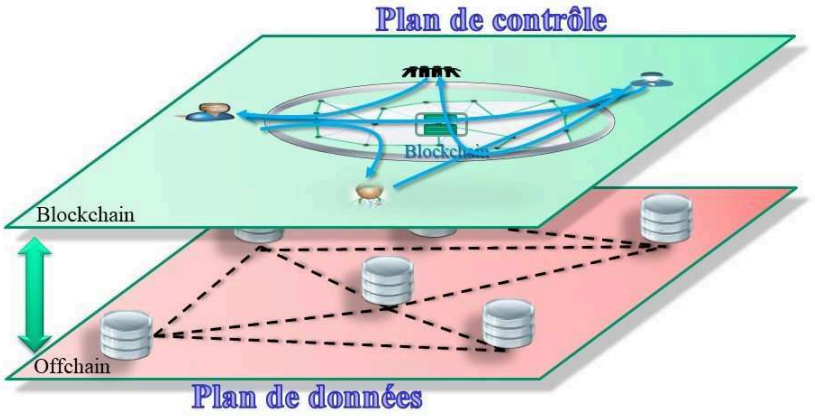


FIGURE 2.3 – Plan de gestion du contrôle d'accès et plan de stockage

les fuites d'informations à travers la blockchain. En effet, étant par essence un registre publique, les transactions doivent être construites de façon à rendre impossible la déduction d'informations sur l'individu à travers l'observation de l'utilisation qui est faite des données. Se pose également le problème de l'anonymat dans les blockchains, car même si en principe les utilisateurs ont la possibilité de se cacher derrière autant d'adresses Blockchain qu'ils peuvent générer de clés, des méthodes existent pour mettre à mal cet anonymat [70]. A ce sujet, Henry *et al.* [54] détaillent des techniques de sauvegarde de l'anonymat dans le réseau Blockchain. Enfin, la mise en œuvre des Blockchain implique plusieurs choix à faire, en fonction des fonctionnalités attendues : la nature de la Blockchain (publique ou privée), les mécanismes utilisés pour le consensus, l'incitation des mineurs, etc.

L'objet de cette thèse n'étant pas la Blockchain, nous nous contenterons de proposer un modèle de Blockchain qui répond à un besoin fonctionnel sans avoir la prétention d'améliorer le concept de la Blockchain, qui sort du champ de recherche de notre thèse.

2.5 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art de la sécurité des données personnelles dans notre environnement d'étude, à savoir, l'écosystème de l'Internet des objets et le Cloud, ou plus généralement sous les paradigmes de déportation des services informatiques. Nous avons, dans un premier temps, discuté les solutions cryptographiques,

notamment, celles utilisant le chiffrement basé sur les attributs, puis nous avons abordé l'aspect gestion de la confiance, avant de terminer par l'utilisation des Blockchains comme moyen de sécuriser des données.

Ainsi, nous terminons cette partie introductive. Dans la partie suivante nous allons présenter notre modèle de protection des données personnelles dans un environnement IdO-Externalisation des traitements centré sur le propriétaire des données.

Deuxième partie

Protection cryptographique

Chapitre 3

Pré-requis scientifiques pour le chiffrement basé sur les attributs

3.1 Introduction

La construction d'un schéma de chiffrement fait appel à des outils mathématiques spécifiques, par exemple l'arithmétique modulaire pour le RSA. Il est également nécessaire d'énoncer des hypothèses de base pour leur sécurité, à l'image de la difficulté de factoriser les grands nombres entiers. Ceci étant, la compréhension formelle du chiffrement basé sur les attributs nécessite des connaissances préalables de certaines constructions mathématiques et notions relatives à la cryptographie.

Dans ce chapitre, nous allons introduire ces notions nécessaires à la compréhension formelle du chiffrement basé sur les attributs (ex. groupe cyclique et couplage symétrique bilinéaire), mais également des notions et définitions qui seront utilisées pour caractériser la sécurité du schéma de chiffrement proposé (ex. sécurité calculatoire et problème difficile).

Notons que ce chapitre n'a pas vocation à être exhaustif ou strictement formel mais de fournir les pré-requis nécessaires pour la bonne compréhension du modèle proposé. Par conséquent, sans verser dans les généralités, nous nous efforcerons de donner les explications les plus claires et précises possibles, avec pour objectif principal d'aider le lecteur à comprendre la suite du travail. Enfin, la rédaction de ce chapitre fait référence, entre autres, au cours de Pierre-Alain Fouque [39], de Stéphane Ballet et Alexis Bonecaze [130] sur la cryptographie ainsi qu'au livre de Katz et Lindell [66].

3.2 Notions utilisées dans ABE

3.2.1 Définitions utiles

Définition 3 (Algorithme polynomial ou algorithme efficace)

Un algorithme $\mathcal{A}$ est dit polynomial (ou efficace) s'il existe un polynôme P tel que pour toute entrée x, dont la taille est définie par un paramètre de sécurité $\lambda \in \mathbb{N}$, le temps d'exécution se termine au plus après $P(\mid x \mid)$ étapes de calcul ; avec $\mid x \mid$ la taille de l'entrée x.

Par la suite, λ sera appelé paramètre de sécurité et la taille de l'entrée x sera exprimée en bits. Nous utiliserons également des algorithmes efficaces *probabilistes*.

Un algorithme est dit *probabiliste*, s'il fait appel à une certaine étape de son exécution à des données tirées aléatoirement et qui vont influencer son comportement. Cette notion est importante pour la cryptanalyse mais aussi pour la cryptographie étant donné que les parties ont besoin de générer des aléas (ex. clé de chiffrement). Souvent, on ne peut pas définir un algorithme déterministe plus efficace qu'un algorithme probabiliste pour résoudre un problème donné. Ceci permet, par la même occasion, de prendre en compte un plus large spectre d'adversaires lors de la définition du modèle de sécurité, que nous verrons par la suite.

Définition 4 (Fonction négligeable)

Une fonction $f : \mathbb{N} \to \mathbb{R}$ est dite négligeable, si pour tout polynôme positif P, on peut trouver un entier N tel que $\forall n > N, f(n) > 1/P(n)$.

Par la suite, nous parlerons de probabilité ou d'avantage négligeable plutôt que de fonction négligeable.

3.2.2 Les Groupes

Définition 5 (Groupe)

Un groupe noté $(\mathbb{G}, \cdot)$ est une structure algébrique, composé d'un ensemble $\mathbb{G}$ muni d'une loi de composition interne $\cdot : \mathbb{G} \times \mathbb{G} \longrightarrow \mathbb{G}$, associative, symétrique et admettant un élément neutre. L'élément neutre est noté 1 en notation multiplicative ($\times$) et 0 en notation additive ($+$).

Un élément $g \in \mathbb{G}$ est dit générateur du groupe $\mathbb{G}$, si $\forall x \in \mathbb{G}, \exists n \in \mathbb{N} \setminus x = g^n$ ($g \cdot g ... g$ n fois).

Un groupe $\mathbb{G}$ est dit *fini*, si sa cardinalité[1] est finie. Dans ce cas, son cardinal est appelé ordre de $\mathbb{G}$. Si p est cet ordre et qu'il est premier, alors on dira que le groupe $\mathbb{G}$ est d'ordre premier p.

Enfin, le groupe $\mathbb{G}$ est dit *cyclique*, s'il est *fini* et qu'il admet un seul élément générateur. Concrètement, les groupes cycliques sont construits à partir des corps finis comme par exemple $\mathbb{Z}/\mathbb{Z}_p$[2] avec p premier (également noté $\mathbb{Z}_p^*$), ou des courbes elliptiques.

Remarque 1

Sauf mention contraire, dans la suite de ce document, nous désignerons par groupe le groupe fini symétrique d'ordre premier muni de la loi de composition interne multiplication.

3.2.3 Couplage symétrique bilinéaire

Définition 6

Un couplage bilinéaire symétrique, noté $e : \mathbb{G}_0 \times \mathbb{G}_0 \to \mathbb{G}_T$, *est une application qui associe à deux éléments de* $\mathbb{G}_0$, *un élément de* $\mathbb{G}_T$. *Avec* $\mathbb{G}_0$ *et* $\mathbb{G}_T$ *des groupes multiplicatifs cycliques, d'ordre premier* p *et* g *un générateur de* $\mathbb{G}_0$. *L'application* e *satisfait les propriétés suivantes :*

1. *Bilinéarité :* $\forall u, v \in \mathbb{G}_0$ *et* $a, b \in \mathbb{Z}_p : e(u^a, v^b) = e(u, v)^{ab}$.
2. *Non-dégénérescence :* $e(g, g) \neq 1$
3. *Symétrie :* $\forall u, v \in \mathbb{G}_0 : e(u, v) = e(v, u)$

Un couplage bilinéaire symétrique e peut être entièrement caractérisé en donnant sa valeur pour $e(g, g)$, du fait de la propriété de non-dégénérescence. La difficulté est de trouver un tel couplage qui soit efficacement calculable.

Remarque 2

L'application de couplage ainsi que les opérations dans $\mathbb{G}_0$ *doivent être calculables par un algorithme fonctionnant en temps polynomial.*

1. Nombre d'éléments
2. Ensemble des classes d'équivalence pour la congruence modulo p.

Dans la pratique, les courbes elliptiques sont utilisées pour le couplage à des fins de chiffrement. Joux [63], dans sa variante du protocole de Diffie-Hellman, est le premier à utiliser le couplage à base de courbes elliptiques. En effet, l'ensemble des points d'une courbe elliptique forment un groupe.

3.2.4 Structure d'accès dans le contexte d'ABE

Définition 7 (Structure d'accès [153])

Soit $\{A_1, A_2, ..., A_n\}$ *un ensemble d'attributs. Une collection* $\mathbb{A} \subseteq 2^{\{A_1, A_2, ..., A_n\}}$ *est dite monotone*[3]*, si* $\forall B, C$*, si nous avons* $B \in \mathbb{A}$ *et* $B \subseteq C$ *alors,* $C \in \mathbb{A}$ *. Une structure d'accès (ou-bien, une structure d'accès monotone) contient une collection (respectivement, une collection monotone) de* $\mathbb{A}$*, sous ensemble non vide de* $\{A_1, A_2, ..., A_n\}$*.*

L'ensemble $\mathbb{A}$ *est appelé l'ensemble des éléments autorisés et les éléments qui ne sont pas dans* $\mathbb{A}$ *sont appelés les éléments non-autorisés.*

Dans la définition originale de Beimel [14], l'ensemble d'attributs est désigné par un ensemble de parties. Dans notre cas, la structure d'accès $\mathbb{A}$ est appelée structure d'accès monotone si elle ne contient pas la négation des attributs. Si c'est le cas, on l'appelle structure d'accès non monotone. Dans la suite de cette thèse, nous considérerons uniquement les structures d'accès monotones.

Dans la littérature relative à ABE, la structure d'accès prend généralement la forme d'un arbre d'accès, noté τ. La Figure 3.1 illustre un tel arbre pour un exemple de politique d'accès simple : $P = Att1$ OR (($Att2$ AND $Att3$) OR ($Att4$ AND $Att5$)). Chaque feuille représente un attribut et chaque nœud interne est une porte logique (ET, OU).

Dans le chapitre suivant, cette structure d'accès sera combinée à un schéma de partage de secret à seuil (Threshold Secret-Sharing Scheme) de Shamir [125], qui fait appel à l'interpolation de Lagrange. Cette construction assurera la possibilité de ne reconstituer un certain secret que par les entités satisfaisant une certaine politique d'accès, définie dans la structure d'accès.

3. Intuitivement : si B est un sous ensemble des éléments autorisés alors, tous les sous ensembles C contenant B sont des éléments autorisés

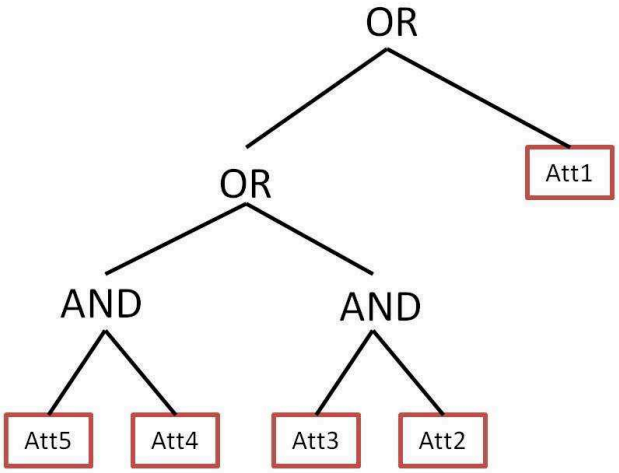


FIGURE 3.1 – Arbre d'accès pour un exemple de politique simple

3.3 Courbes elliptiques

Définition 8 (Équation de Weierstrass)

Soit $\mathbb{F}$ un corps, l'équation de Weierstrass[4] *sur $\mathbb{F}$ est une forme simplifiée de l'équation d'une courbe elliptique, elle prend la forme générale suivante :*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

Avec $a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}$

Une courbe algébrique satisfaisant l'équation de Weierstrass est dite lisse si les dérivées partielles en x et y de la fonction $f(x, y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$ ne s'annulent pas en même temps.

Définition 9

Une courbe elliptique E définie sur un corps $\mathbb{F}$ est une courbe lisse définie par une équation de Weierstrass augmentée d'un point $O \in \mathbb{F}$ (qui n'appartient pas à la courbe) appelé point à l'infini, qui sera l'élément neutre de l'addition géométrique des points de la courbe.

4. Karl Theodor Wilhelm Weierstrass, mathématicien allemand (1815-1897)

3.3.1 Avantage des courbes elliptiques en cryptographie

L'utilisation des courbes elliptiques en cryptographie a été proposée par Miller [92] et Neal [99] dans les années quatre-vingt comme alternative au chiffrement asymétrique existant, en utilisant des courbes elliptiques sur des corps finis d'ordre premier $\mathbb{F}_q$ avec $q = p^t$ et p premier.

Le principal avantage des courbes elliptiques est la difficulté accrue pour résoudre le problème mathématique sous-jacent (voir 3.4.0.2) par rapport aux autres schémas à longueur de clé égale. Ce qui revient à dire, que pour un niveau de sécurité donné, nous avons des chaînes de bits plus petites et par là, de meilleures performances en terme d'implémentation [50].

La longueur des chaînes de bits manipulées dépend de la taille du groupe utilisé, d'où l'importance de connaître la taille du groupe. Cependant, calculer le nombre de points d'une courbe elliptique sur un corps fini est en général difficile. Toutefois, le théorème de Hasse[5] sur les courbes elliptiques donne un intervalle pour le nombre de points de la courbe elliptique. A noter que le nombre de points d'une courbe elliptique est le nombre de solutions de l'équation de Weierstrass qui caractérise cette courbe.

Théorème 1 (Théorème de Hasse sur les courbes elliptiques)
Soit une courbe elliptique E sur un corps fini $\mathbb{F}$ à q éléments, alors le nombre de points de cette courbe, noté N satisfait :

$$|N - (q + 1)| \leq 2\sqrt{q}$$

En d'autres termes N est borné :

$$q + 1 - 2\sqrt{q} \leq N \leq q + 1 + 2\sqrt{q}$$

On admet généralement le fait que le nombre de points d'une courbe elliptique est proche de la taille de l'ordre premier sélectionné. Par exemple pour un ordre premier de la taille de 256 bits, le nombre de points relatifs à cette courbe elliptique est approximativement de 2^{256}.

5. Mathématicien allemand (1898-1979).

Les meilleures attaques connues contre les courbes elliptiques (pas-de-bébé pas-de-géant [127] et la méthode rho de Pollard [111]) ont une complexité moyenne de l'ordre de $\sqrt{q}$ étapes pour résoudre le problème du logarithme discret sur les courbes elliptiques (voir 3.4.0.3), soit une complexité en temps et mémoire $O(\sqrt{q})$. Ce qui assure un niveau de sécurité équivalent à $\sqrt{q}$ étapes, avec q le nombre de points de la courbe elliptique considérée. Pour revenir à l'exemple précédent, une courbe elliptique utilisant un ordre premier de 256 bits (en moyenne $q = 2^{256}$ points), cette courbe elliptique assurera une sécurité équivalente à $\sqrt{2^{256}}$ soit 2^{128} étapes.

Enfin, il existe des algorithmes (Schoof, Atkin, Elkies) qui permettent de donner le nombre précis des points d'une courbe elliptique. Cette notion permet de définir le niveau de sécurité souhaité, en définissant l'ordre de grandeur du groupe utilisé.

La sécurité offerte par les courbes elliptiques dépend du choix de la courbe, notamment, le corps $\mathbb{F}$ sur lequel elle sera définie, l'équation de Weierstrass mais également du degré de plongement.

Définition 10 (Le degré de plongement)
Le degré de plongement (embedding degree en anglais) est le plus petit entier k tel que le problème du logarithme discret (DL) sur une courbe elliptique E qui est défini sur un corps fini $\mathbb{F}_p$ peut être transformé en un problème de logarithme discret sur un corps fini $\mathbb{F}_{p^k}$.

Le degré de plongement doit être le plus grand possible afin de se prémunir des algorithmes pouvant résoudre le problème du logarithme discret sur le corps fini $\mathbb{F}_{p^k}$ (voir 3.4.0.3) et qui pourraient être plus efficaces que si on résolvait le problème du logarithme discret sur la courbe elliptique définie sur $\mathbb{F}_p$.

Enfin, à titre indicatif, la Table 3.1, fournie par Hemalatha et Manickachezian [52] sur la base des recommandations du NIST [13], donne un ordre de grandeur de la taille des clés pour différents schémas de chiffrement. Pour la colonne relative à ABE, q donne la taille en bits du corps fini sur lequel sont définis la courbe elliptique et l'ordre premier p du point de base.

TABLE 3.1 – Ordre de grandeurs des tailles et niveau de sécurité relatives aux primitives de chiffrement principales

Security Strength(bits)	Symmetric algorithms	RSA-Key length	ABE p (prime order), q (field size)
80	2TDEA	1024	p=160, q=512
112	3TDEA	2048	p=224, q=1024
128	AES-128	3072	p=256, q=1536

3.4 Sécurité calculatoire

A défaut de prouver qu'un schéma de chiffrement présente une sécurité parfaite, même avec un adversaire disposant d'une puissance de calcul infinie (hypothèse très forte inutilement), on prouve la sécurité dite calculatoire du schéma. En faisant des suppositions sur les limites en temps (ressources) de calculs de l'adversaire et en minimisant la probabilité de son succès, ceci revient à dire, dans le cas de la sécurité calculatoire, que : la sécurité n'est garantie que pour un adversaire *efficace* ayant des ressources ou s'exécutant en un temps *borné* et que cet adversaire peut réussir avec une probabilité suffisamment *petite* pour être acceptable.

3.4.0.1 L'approche asymptotique de la sécurité calculatoire

L'approche asymptotique de la sécurité calculatoire définit un entier λ, appelé *paramètre de sécurité* qui caractérise le schéma de chiffrement. La probabilité de succès de l'adversaire sera quantifiée comme une fonction de λ. Dans cette approche, la notion d'adversaire *efficace* est remplacée par la notion *d'algorithme probabiliste s'exécutant en un temps polynomial* λ, avec λ suffisamment grand (d'où la mention d'asymptotique). Quant à la notion de succès avec une probabilité suffisamment *petite*, elle est remplacée par la notion de probabilité *négligeable*. La définition suivante est généralement admise pour la sécurité calculatoire asymptotique.

Définition 11 (sécurité calculatoire asymptotique)
Un schéma est sûr si aucun adversaire s'exécutant en un temps polynomial ne réussit à casser le schéma avec une probabilité au plus négligeable.

A noter qu'il existe une deuxième approche, qui est l'approche concrète en donnant

des valeurs explicites à la probabilité maximale de succès qu'a un adversaire, s'exécutant pendant un temps limité. Par exemple, en garantissant qu'un schéma de chiffrement ne peut être brisé par un adversaire s'exécutant pondant 100 ans sur le plus rapide des ordinateurs avec une probabilité supérieure à 2^{-50}. Cette approche est utile pour certains schémas comme le DES, où la taille des clés est fixe. Par conséquent, l'approche asymptotique n'a pas de sens dans ce sas. Néanmoins, quantifier précisément ces garanties est difficile et restrictive dans le sens où elles n'ont de sens que pour les valeurs exprimées. Par exemple, en garantissant l'inviolabilité contre un adversaire s'exécutant sur 2 ans avec une certaine probabilité, quelle serait cette probabilité pour un adversaire s'exécutant sur 3 ans ?

3.4.0.2 Problème difficile

Les différents schémas de chiffrement se basent sur des problèmes algorithmiques dit difficiles. La difficulté à résoudre nous renvoie à la définition 11 de la sécurité calculatoire asymptotique, dans le contexte de la cryptographie. En effet, ce sont des problèmes algorithmiques qui supposent qu'aucun algorithme s'exécutant en un temps polynomial ne peut résoudre ce problème avec une probabilité non négligeable (à un niveau asymptotique).

La sécurité du RSA, par exemple, est basée sur la difficulté de factoriser un grand nombre entier en facteurs premiers. L'hypothèse de sécurité de RSA dit que si un algorithme permet de résoudre ce problème de factorisation facilement alors il pourra *casser* la sécurité du RSA. Un autre problème difficile est le logarithme discret qui permet de définir un ensemble de problèmes difficiles sur les groupes cycliques d'ordre premier.

3.4.0.3 Problème du logarithme discret (DL)

Définition 12 (Problème du logarithme discret)

Soient $\mathbb{G}$ et g, respectivement un groupe cyclique d'ordre p et un générateur de ce groupe. Le problème du logarithme discret revient à trouver x, connaissant g et g^x ; avec x choisi aléatoirement dans $\mathbb{Z}_p^*$.

A partir de ce problème, plusieurs autres ont été définis, dont une classe de problèmes dite décisionnelle. Dans cette classe, le problème revient à *décider* si une solution donnée pour un problème calculatoire est bien la bonne solution. Cette notion sera utilisée par

la suite pour démontrer la sécurité de notre schéma de chiffrement en prenant comme hypothèse de sécurité le problème de Diffe-Hellman bilinéaire décisionnel, que nous allons présenter ci-après.

3.4.0.4 Le problème Diffe-Hellman bilinéaire décisionnel

L'hypothèse du problème décisionnel bilinéaire de Diffie-Hellman, noté (DBDH) est une hypothèse de difficulté algorithmique basée sur la difficulté de calcul des logarithmes discrets dans des groupes cycliques.

Définition 13 (Le problème Diffe-Hellman bilinéaire décisionnel)
Étant donné e, p, g, g^a, g^b, g^c, T, avec $a, b, c \in \mathbb{Z}_p^*$, choisis aléatoirement et p premier, décider si $T = e(g,g)^{abc}$ ou généré aléatoirement.

Il est utile également de reprendre la définition donnée par Waters [144] pour le contexte d'ABE.

Définition 14
Un challenger sélectionne un groupe $\mathbb{G}_0$ d'ordre premier p, conformément au paramètre de sécurité choisi et g son générateur. Soient $a, b, s \in \mathbb{Z}_p^*$ choisis aléatoirement. Si le challenger fournit à l'adversaire (g, g^a, g^b, g^s) alors cet adversaire ne doit pas distinguer le résultat valide $e(g,g)^{abs} \in \mathbb{G}_T$ d'un élément quelconque $Z \in \mathbb{G}_T$ avec un avantage non négligeable.

Ceci nous amène à la formulation suivante de la notion d'avantage d'un algorithme pour résoudre un problème dans l'environnement de groupes finis symétriques d'ordre premier (voir section 3.5.3 pour un formalisme plus général de la notion d'avantage).

Définition 15 (Avantage d'un algorithme [144])
Soit un algorithme $\mathcal{B}$, avec en entrée (g, g^a, g^b, g^s, T) et qui produit en sortie 0 ou 1, selon qu'il décide que T est valide ou généré aléatoirement ($T = Z$ avec Z aléatoire). $\mathcal{B}$ a comme avantage ϵ pour résoudre le problème DBDH dans $\mathbb{G}_0$ si :

$$| Pr[\mathcal{B}(g, g^a, g^b, g^s, T = e(g,g)^{abs}) = 0] - Pr[\mathcal{B}(g, g^a, g^b, g^s, T = Z) = 0] | \geq \epsilon$$

Enfin, il existe une variante calculatoire du problème Diffe-Hellman bilinéaire décisionnel, dans laquelle le problème revient à calculer g^{xy} connaissant g, g^x et g^y.

3.5 Modèle de sécurité

La caractérisation de la sécurité d'un schéma de chiffrement passe par la définition stricte des capacités de l'adversaire éventuel. Un principe fondamental sur la capacité de l'adversaire a été énoncé par Auguste Kerckhoffs[6]. Ce principe suppose que l'algorithme de chiffrement et tous les paramètres publiques sont connus de l'adversaire. Autrement dit, la sécurité d'un schéma de chiffrement ne doit reposer que sur le secret de la clef.

On distingue principalement deux modèles d'attaques, l'attaque à texte clair choisi et l'attaque à texte chiffré choisi. La définition du modèle d'attaque passe également par la notion *d'indistinguabilité* (voir plus loin).

Remarque 3

Il est également utile de doter l'adversaire de la capacité d'exécuter autant de requêtes qu'il estime utile et également de pouvoir exécuter chaque requête de façon adaptative (en fonction des précédentes).

3.5.1 Attaque à texte clair choisi (en anglais : chosen-plaintext attack)

Dans ce type d'attaques, noté CPA, l'adversaire a la possibilité d'obtenir les messages chiffrés correspondant aux messages en clair de son choix. Il est clair que cette attaque est toujours possible pour les schémas de chiffrement à clé publique. L'algorithme de chiffrement et la clé publique de chiffrement sont supposés connus de tout le monde. Dans le cas de schémas à clé secrète, cette attaque suppose que l'adversaire a la possibilité de demander à des parties partageant une clé secrète de chiffrer des messages qu'il a sélectionnés et de récupérer le chiffré correspondant.

Des exemples historiques de la deuxième guerre mondiale ont été rapportés pour ce type d'attaque, à l'image des alliés qui interceptaient les messages chiffrés des allemands sur la position des mines marines (le message chiffré), sachant que leur positionnement est à l'initiative des alliées (le message clair choisi). Cette connaissance aurait aidé les cryptanalystes de Bletchley Park à casser les chiffres allemands.

Formellement, cette attaque est modélisée par un oracle de chiffrement, qui est une boite

6. Auguste Kerckhoffs von Nieuwenhoff, cryptologue néerlandais (1835-1903).

noire à laquelle l'adversaire $\mathcal{A}$ a accès. Cet oracle procède au chiffrement des messages en clair soumis par l'adversaire en choisissant à chaque fois une nouvelle clé secrète inconnue de l'adversaire.

3.5.2 Attaque à texte chiffré choisi (en anglais : chosen-ciphertext attack)

Dans ce type d'attaques, l'adversaire a accès à un oracle de déchiffrement lui permettant de déchiffrer des messages chiffrés de son choix. Cette attaque est notée CCA1. Il existe une deuxième variante, notée CCA2, qui ne diffère de la première que par le fait que l'adversaire peut adapter ses requêtes en fonction de l'évolution de son attaque.

Ce type d'attaques est moins réalisable dans le monde réel car elle suppose que l'adversaire a la possibilité d'exécuter l'algorithme de déchiffrement sur des chiffrés choisis par lui-même, sans pour autant connaître la clé secrète.

3.5.3 L'indistinguabilité

Définition 16 (Indistinguabilité)

La notion d'indistinguabilité traduit l'incapacité d'un attaquant à qui on propose deux messages en clair et un chiffré (de l'un des deux messages) de distinguer lequel des deux messages a été chiffré.

Si un adversaire arrive à distinguer le message correspondant au chiffré (en se passant d'un choix au hasard), alors cela signifie qu'il a la capacité d'obtenir au moins un bit d'information sur le texte chiffré. Plus formellement, la notion d'indistinguabilité peut être appréhendée pour le chiffrement à clé publique (mais également pour le chiffrement à clé secrète) sous l'attaque CPA, en utilisant une expérience impliquant un adversaire $\mathcal{A}$ et un challenger $\mathcal{C}$.

Étant donné un schéma de chiffrement à clé publique $\varepsilon(Enc, Dec, KeyGen)$, un paramètre de sécurité k (voir définition 2) et une sous-routine $\mathcal{U}()$ qui retourne une valeur γ aléatoirement choisie dans $\{0, 1\}$, l'expérience notée $Exp_{\varepsilon}^{Ind-CPA-\gamma}(\mathcal{A})$, se déroule comme suit [66] :

$\mathcal{C}$		$\mathcal{A}$
$(PK, SK) \leftarrow KeyGen(k)$	$\xrightarrow{PK}$	
$\gamma \leftarrow \mathcal{U}()$	$\xleftarrow{m_0, m_1}$	Choisit(m_0 et m_1)
$C \leftarrow Enc(PK, m_\gamma)$	$\xrightarrow{C}$	
		Retourne $\gamma' \in \{0, 1\}$

Le challenger $\mathcal{C}$ génère les paramètres secrets et publiques du chiffrement et envoie les paramètres publiques à l'adversaire $\mathcal{A}$. Ce dernier choisit deux messages m_1 et m_2, avec $|m_1| = |m_2|$ et les transmet au challenger $\mathcal{C}$. Le challenger tire au hasard un bit $\gamma \in \{0, 1\}$, chiffre le message m_γ et le retourne à l'adversaire. $\mathcal{A}$ retourne un bit γ'. Le résultat de l'expérience est le suivant :

$$Exp_{\varepsilon}^{Ind-CPA-\gamma}(\mathcal{A}) = 1 \qquad \textbf{si} \qquad b = b', \qquad 0 \qquad \textbf{autrement}$$

La probabilité de succès de $\mathcal{A}$ à cette expérience ($Exp_{\varepsilon}^{Ind-CPA-\gamma}(\mathcal{A}) = 1$), est notée par $Pr[\gamma = \gamma']$

Définition 17 (Avantage de l'adversaire $\mathcal{A}$)
L'avantage de l'adversaire $\mathcal{A}$ au jeu $Exp_{\varepsilon}^{Ind-CPA-\gamma}(\mathcal{A})$, exécutant une attaque à texte en clair choisi est défini par la quantité :

$$Adv_{\varepsilon}^{Ind-CPA}(\mathcal{A}) = Pr[\gamma = \gamma'] - 1/2$$

Définition 18 (Schéma de chiffrement sûr contre une attaque à texte en clair choisi)
Un schéma de chiffrement ε avec un paramètre de sécurité λ est dit sûr contre un adversaire $\mathcal{A}$ s'exécutant en un temps polynomiale s'il existe une fonction négligeable $\epsilon(\lambda)$ vérifiant :

$$Adv_{\varepsilon}^{Ind-CPA}(\mathcal{A}) \leqslant 1/2 + \epsilon(\lambda)$$

En d'autres termes, cette définition traduit le fait qu'aucun adversaire $\mathcal{A}$ s'exécutant en un temps polynomial ne pourra réussir à l'expérience $Exp_{\varepsilon}^{Ind-CPA-\gamma}(\mathcal{A})$ avec un avantage significativement plus important que s'il tirait la valeur de γ' au hasard, donc avec $Pr[\gamma = \gamma'] = 1/2$.

Cette notion est appelée indiscernabilité sous l'attaque à texte en clair choisi, notée IND-CPA. Nous avons délibérément choisi de ne pas détailler cette notion sous l'attaque à

texte chiffré choisi (IND-CCA) car nous estimons que dans la réalité ce scénario d'attaque est une hypothèse inutilement forte et peu probable (capacité de choisir et d'avoir accès à des textes chiffrés et à leurs textes en clair simultanément). De plus, la construction d'un schéma IND-CCA est particulièrement utile pour les protocoles d'échange de clés, où un adversaire a la possibilité d'envoyer des messages chiffrés comme challenge et obtenir par conséquent leurs textes déchiffrés en retour. En ce qui concerne notre travail, le protocole proposé ne permet pas un tel comportement.

Enfin, le cas échéant, Cramer et Shoup [32] décrivent une méthode pour construire un schéma de chiffrement asymétrique prouvé IND-CCA à partir d'un schéma prouvé IND-CPA, en utilisant des fonctions de hachage.

3.5.4 Preuve de sécurité par l'absurde

Une méthode utilisée pour prouver la sécurité d'un schéma de chiffrement est la preuve par l'absurde. Elle consiste à démontrer que si un adversaire peut réussir dans une expérience Exp_{ε}^{X}, relative à une attaque X sur un schéma ε avec un avantage non négligeable, alors on peut construire un adversaire capable de résoudre le problème sous-jacent. Dans cette preuve, nous allons considérer deux parties : l'adversaire $\mathcal{A}$ contre le schéma de chiffrement considéré et un challengeur $\mathcal{C}$ pour le problème difficile sous-jacent. L'objectif étant de construire un simulateur $\mathcal{B}$ qui émule le schéma de chiffrement considéré et qui interagit avec l'adversaire de telle sorte que ce dernier puisse produire une réponse permettant de réussir l'expérience Exp_{ε}^{X} avec un avantage non négligeable. A son tour, le simulateur $\mathcal{B}$ va utiliser ce résultat pour résoudre le problème difficile considéré.

Pour terminer la preuve, on évalue l'avantage de succès de l'adversaire en sachant qu'elle est bornée par l'avantage de réussir à résoudre le problème difficile, qui lui est négligeable (hypothèse de sécurité de départ du schéma considéré).

Remarque 4

Il est utile de rappeler que toutes les parties invoquées s'exécutent en un temps polynomial.

3.6 Conclusion

Dans ce chapitre, nous avons décrit les instruments qui seront nécessaires à la construction et à la validation (en terme de sûreté) de notre schéma de chiffrement basé sur les attributs, intégrables dans un environnement de déportation informatique. Dans le chapitre suivant, nous allons décrire formellement notre schéma ainsi que la preuve de sa validité cryptographique contre un adversaire réaliste.

Chapitre 4

Modèle de protection de données personnelles dans un environnement IdO-Cloud, basé sur le chiffrement par attributs

4.1 Introduction

Le Cloud Computing fournit des capacités de stockage et de traitement pour les objets connectés disposant de ressources limitées. Récemment, pour améliorer l'efficacité du Cloud Computing, les paradigmes de Edge et de Fog Computing ont été introduits pour rapprocher les services de Cloud des utilisateurs finaux [20; 150]. Ces services sont disponibles directement à la périphérie du réseau, tels que des points d'accès, des routeurs ou des cloudlets [38; 112]. L'une des principales applications de ces technologies est le domaine de la e-santé. La e-santé tire profit des fonctionnalités des paradigmes de l'Internet des objets, du Cloud et du Cloud mobile pour permettre aux patients d'être surveillés à tout moment et n'importe où par le biais d'une multitude de dispositifs connectés, tout en partageant les données stockées dans le Cloud Dinh Thai *et al.* [35]; Gope et Hwang [44]; Ould-Yahia et Paradinas [108].

Du fait de l'externalisation des données à caractère personnel, leur sécurité et leur confidentialité sont devenues l'une des préoccupations majeures du point de vue du propriétaire des données. Par conséquent, le propriétaire des données souhaite s'assurer que ses

données collectées sont traitées en toute sécurité dans les nœuds Edge/Fog et stockées de manière sécurisée sur les serveurs du Cloud et enfin, accessibles uniquement aux utilisateurs autorisés. Cependant, la sécurité des données dans un environnement Cloud n'est pas garantie [88; 113], et les menaces peuvent provenir des fournisseurs de services eux-mêmes. En effet, de nombreuses entreprises ont un intérêt économique à collecter des données privées, notamment de santé [81]. En outre, des algorithmes et des protocoles de sécurité complexes ne peuvent pas être utilisés dans les écosystèmes IdO et de mobiles, en raison des ressources physiques et énergétiques limitées [104].

4.1.0.1 Motivation

Pour ces raisons, ces dernières années, un modèle centré sur le propriétaire des données est apparu dans la littérature pour la protection des données à caractère personnel. Conformément à la littérature, le schéma de chiffrement à base d'attributs (ABE) est un moyen viable d'améliorer la confidentialité dans de nombreux domaines tels que les applications de e-santé [68; 107; 113] mais également pour les applications de maison intelligente [28] et plus globalement dans l'environnement Cloud Computing [98]. En outre, pour sécuriser la transmission et le stockage des données, ABE fournit un contrôle d'accès détaillé et un partage flexible des données [52; 79]. Néanmoins, les schémas ABE impliquent des opérations de mapping et des exponentiations gourmandes en ressources, sachant que leur complexité augmente de manière linéaire avec le nombre d'attributs. Actuellement, même si des implémentations efficaces des schémas ABE peuvent être mises en œuvre sur des équipements "classiques" (PC, serveur, etc.), ce n'est pas toujours le cas pour les périphériques ayant des limitations en ressources (capteurs, appareils mobiles et IdO, par exemple) [7; 141]. Pour apporter une solution à ce problème, [9; 47; 137; 153; 157] avaient proposé une externalisation d'ABE, prouvant de ce fait, la viabilité de ce concept.

Nous présenterons dans ce chapitre un schéma de déportation des calculs lourds d'ABE avec de meilleures performances de chiffrement et une faisabilité dans un environnement Internet des objets. Ce schéma de chiffrement sera intégré dans un protocole à sécurité décentralisée basé sur la Blockchain.

4.2 Modèle considéré

4.2.1 Architecture du système

Commençons par définir le système pour lequel notre schéma de chiffrement et notre protocole ont été développés. Nous définirons également l'environnement de menaces considéré. Typiquement, le modèle d'utilisation d'ABE, illustré par la Figure 4.1, considère un producteur ou source de données (DS), qui produit et chiffre les données, un consommateur de données (RQ) qui reçoit les données et les déchiffre, et enfin une autorité de confiance (CA) chargée de la gestion des clés et des attributs.

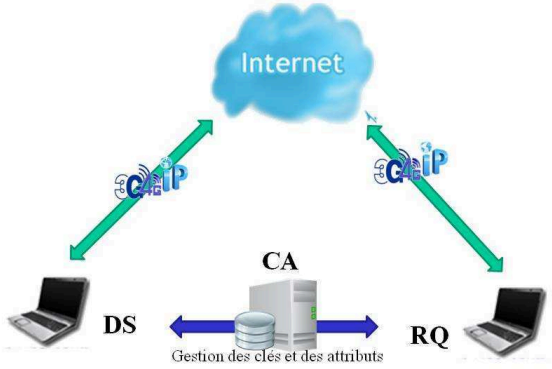


FIGURE 4.1 – Cas d'utilisation générique d'ABE

Concrètement, nous allons adopter le scénario de la surveillance médicale de l'état d'un individu à l'aide de capteurs connectés. Les données collectées sont transmises au Cloud pour leur traitement, stockage et surtout partage avec les entités autorisées. Ce scénario intègre dans son architecture le contexte de l'Internet des objets et du Cloud ainsi que de fortes contraintes de confidentialité et de préservation de la vie privée [1]. Nous considérons que les sources de données (DS) sont limitées en capacité de calcul, mémoire et énergie. D'où une architecture incluant une entité d'externalisation de certains traitements, notée (PR) pour proxy. Dans une situation réelle, ces entités seront instanciées par des technologies de Fog computing. La Figure 4.2 donne une image de ce type d'architecture.

Enfin, RQ peut être le médecin traitant doté d'une preuve cryptographique attestant

de ses attributs, par exemple une carte à puce délivrée par une autorité de confiance. Cette preuve lui permettra, entre autres, d'attester de ses attributs. Pour le partage des clés, une solution simple (mais réaliste) a été proposée par Chandu *et al.* [24], utilisant le courrier électronique et le chiffrement RSA.

Remarque 5

En France, les professionnels de la santé sont titulaires d'une carte à puce, la Carte de Professionnelle de Santé (CPS), qui permet un accès sécurisé aux données personnelles de santé. Elle est délivrée aux professionnels de santé par l'agence des systèmes d'information partagés de santé (ASIP santé).

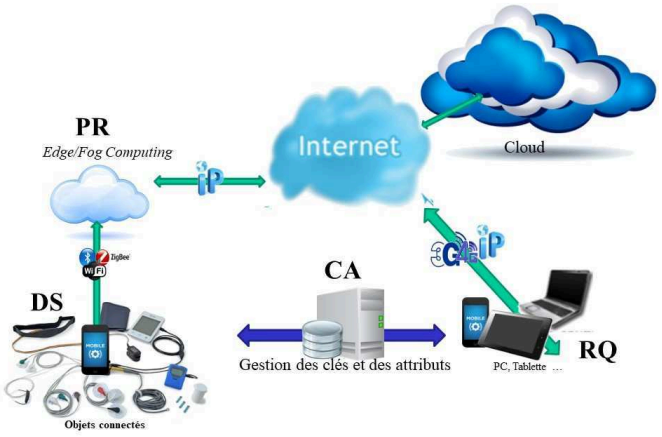


FIGURE 4.2 – Vue globale de l'architecture système considérée

4.2.2 Modèle de menace

Pour définir un modèle de menace, il convient d'envisager la réalité du monde afin, d'une part de ne pas sous-estimer la menace et d'autre part de ne pas la surestimer. En effet, toute surestimation inutile, se traduira par un surcoût de la sécurité également inutile. Le but de la solution que nous proposons est de préserver la vie privée du patient dans le scénario considéré.

L'analyse de ce scénario permet de dégager une première menace induite par le canal de communication. Souvent on utilise des communications sans fil, facilement interceptables.

Un modèle qui prend en compte cette menace est le modèle de Dolev et Yao [36], qui considère le réseau comme un intrus. Ainsi, un attaquant peut écouter, supprimer, rejouer et modifier un message. Cependant, l'attaquant ne pourra pas déchiffrer un message s'il ne possède pas la clé de déchiffrement.

Ensuite, étant donné que les infrastructures d'externalisation (Cloud, Edge, Fog, etc.) sous le contrôle d'un tiers qui pourrait ne pas être totalement sûr, nous adoptons le même modèle que Li *et al.* [79], qui considère ces infrastructures comme des acteurs honnêtes mais curieux. Cela signifie que le Cloud, par exemple, instancie correctement le protocole mais peut essayer de déduire des informations sur le patient.

En outre, ce modèle prend en charge la menace réaliste que représente le contexte du Cloud, dans lequel il peut être la cible d'entités malveillantes. Il peut également faire l'objet d'une fuite d'informations personnelles par inadvertance.

Enfin, nous considérons l'autorité de confiance (CA) comme totalement honnête.

4.3 Le Chiffrement basé sur les attributs avec l'informatique en brouillard (FCCP-ABE)

Pour réaliser un contrôle d'accès avec une fine granularité et assurant la protection des données stockées, nous proposons un nouveau schéma de chiffrement basé sur le chiffrement par attributs et dont nous avons prouvé la sûreté sous une attaque adaptative à clairs choisis. Ce schéma est un chiffrement basé sur des attributs avec une politique d'accès incluse dans le chiffré, adapté à l'environnement de l'informatique en brouillard ou Fog-Computing Cipher-Policy Attribute-Based Encryption, par abréviation FCCP-ABE[1]. Conscients que la limitation des ressources est une préoccupation majeure de l'IdO, nous proposons une conception du chiffrement avec un mécanisme de déportation des tâches gourmandes. Dans ce chapitre, nous présentons notre conception ainsi que la preuve de sécurité de notre schéma dans un contexte de sécurité réaliste.

1. Fog-Computing Cipher-Policy Attribute-Based Encryption

4.3.1 Construction proposée

Soient $\mathbb{G}_0$ un groupe multiplicatif d'ordre premier p et g un générateur de ce groupe. Soit également $e : \mathbb{G}_0 \times \mathbb{G}_0 \to \mathbb{G}_T$ un couplage bilinéaire. Nous considérons également, le coefficient de Lagrange $\Delta_{i,S}$ défini pour $i \in \mathbb{Z}_p$ et un ensemble S éléments dans $\mathbb{Z}_p$:

$$\Delta_{i,S}(x) = \prod_{j \in S, j \neq i} \frac{x-j}{i-j}$$

Le coefficient de Lagranfe permettra de donner le résultat d'un polynôme q_x au point 0, où x est un noeud de l'arbre d'accès. Ce polynôme est utilisé dans le schéma du partage de secret (voir ci-dessous).

De plus, on considère une fonction de hachage, $H : \{0,1\}^* \to \mathbb{G}_0$ et une primitive de chiffrement symétrique $Senc(M, Ks)$ qui chiffre et déchiffre le message M avec la clé secrète Ks.

Remarque 6

Pour améliorer l'efficacité de notre solution, nous chiffrons la masse de données avec un chiffrement symétrique et la clé associée avec FCCP-ABE.

Préalablement à la définition formelle de notre schéma, nous considérons les primitives suivantes, relatives à la manipulation de l'arbre d'accès τ :

— $parent(x)$ qui retourne le nœud parent du nœud x.

— $num(x)$ est le nombre d'enfants du nœud x.

— $index(x)$, donne l'ordre des enfants de chaque nœud, en numérotant chaque enfant de 1 à $num(parent(x))$, en fixant une convention de sens.

— $attrib(x)$ renvoie un identifiant numérique unique de l'attribut associé à la feuille x de l'arbre d'accès τ.

Nous définissons également une valeur de seuil $k_x = num(x)$, si x est une porte AND, et $k_x = 1$ sinon. Chaque seuil k_x sera utilisé comme degré du polynôme pour le schéma de partage de secret [17; 126].

Par la suite, nous définissons les algorithmes suivants pour notre schéma de chiffrement FCCP-ABE :

— $Setup(\lambda) \rightarrow (Pk, Msk, \alpha, \beta)$: Cet algorithme génère la clé publique Pk et la clé secrète principale Msk, correspondant à un paramètre de sécurité λ, donné en entrée. Ce paramètre détermine la taille du groupe. De plus, cet algorithme générera aléatoirement deux nombres $\alpha, \beta \in \mathbb{Z}_p$. Les deux clés Pk et Msk sont définies comme suit :

$$Pk = (\mathbb{G}_0, g, h = g^{\beta}, e(g,g)^{\alpha/\beta})$$

$$Msk = (\alpha, \beta)$$

— $EncryptCons(M, Pk, Ks) \rightarrow CT_{one}$: Cet algorithme, destiné à être exécuté sur les équipements contraints, prend en entrée un message à chiffrer M, la clé publique de chiffrement pour FCCP-ABE Pk et une clé de chiffrement symétrique Ks. L'algorithme sélectionne un élément $t \in \mathbb{Z}_p$ aléatoirement et génère un chiffré intermédiaire CT_{one}, comme suit :

$$CT_{one} = (\tau, C_1, C_2, C_3, C_4)$$

Avec : $C_1 = Ks \cdot e(g,g)^{\alpha t/\beta}$, $C_2 = h^t$, $C_3 = g^t$ et $C_4 = SEnc(M, Ks)$. Et τ une structure d'accès.

— $EncryptUncons(CT_{one}, \tau, Pk) \rightarrow CT_{two}$: Cet algorithme est exécuté par des périphériques non contraints. D'abord, nous définissons un schéma de partage de secret à seuil [125], associé à une structure d'accès (voir 3.2.4), tel que proposé dans le schéma original de CP-ABE. Nous commençons par définir un polynôme q_x avec un degré $d_x = k_x - 1$ pour chaque nœud x de l'arbre d'accès τ (k_x est une valeur de seuil définie précédemment). En partant du nœud racine R, nous sélectionnons un élément aléatoire $s \in \mathbb{Z}_p$ et on pose $q_R(0) = s$. Ensuite, afin de définir complètement le polynôme q_R, nous choisissons d_R points au hasard (nous rappelons que pour définir totalement un polynôme de degré d_R, il suffit de définir explicitement $d_R + 1$ points). Pour les autres nœuds x, nous définissons $q_x(0) = q_{parent(x)}(index(x))$ et nous choisissons d_x autres points de manière aléatoire pour la finalisation de la définition de q_x.

Soit X un ensemble de feuilles dans l'arbre τ (correspondant à un ensemble d'attributs utilisés dans la stratégie d'accès). Le texte chiffré final est construit en calculant CT_{two} :

$$C'_2 = C_2 \cdot h^s = h^{t+s}$$

$$C'_3 = C_3 \cdot g^s = g^{t+s}$$

$$\forall x \in X : C_x = g^{q_x(0)}, C'_x = H(attrib(x))^{q_x(0)}$$

$$CT_{two} = (\tau, C_1, C_3, C'_2, C'_3, C_4, \forall x \in X : C_x, C'_x)$$

— $KeyGen(Msk, C_2, S) \rightarrow Sk$: Cet algorithme prend en entrée une clé secrète principale Msk, la composante C_2 générée avec *EncryptCons* et un ensemble d'attributs S. Il associe un élément aléatoire $r_i \in \mathbb{Z}_p$ pour chaque attribut $i \in S$. Aussi, pour se prémunir des attaques par collusion, un élément aléatoire $r \in \mathbb{Z}_p$ est généré pour chaque utilisateur [120]. La clé secrète est calculée comme suit :

$$Sk = (D_1 = g^{(r+\alpha/\beta)}, D_2 = g^r \cdot C_2, \forall i \in S : D_i = g^r \cdot H(i)^{r_i}, D'_i = g^{r_i})$$

— $Decrypt(CT_{two}, Sk, S) \rightarrow M$: L'algorithme de déchiffrement fait appel à l'algorithme récursif $DecryptNode(CT_{two}, Sk, x, S)$, inspiré du travail de [19] qui prend en entrée un cryptogramme CT_{two}, une clé secrète Sk et un nœud x de la structure d'accès τ ainsi qu'un ensemble d'attributs autorisés S. La démarche est la suivante :

D'abord, il extrait τ de CT_{two}. Si le nœud x est une feuille, alors $i = attrib(x)$, ensuite si $i \notin S$, alors $\perp \leftarrow DecryptNode(CT_{two}, Sk, x)$. Autrement dit :

$$\begin{aligned} DecryptNode(CT_{two}, Sk, x) &= \frac{e(D_i, C_x)}{e(D'_i, C'_x)} \\ &= \frac{e(g^r \cdot H(i)^{r_i}, g^{q_x(0)})}{e(g^{r_i}, H(i)^{q_x(0)})} \\ &= \frac{e(g^r, g^{q_x(0)}) \cdot e(H(i)^{r_i}, g^{q_x(0)})}{e(g^{r_i}, H(i)^{q_x(0)})} \\ &= e(g, g)^{r q_x(0)} \end{aligned}$$

La récursivité est définie comme suit : Si x n'est pas un nœud feuille, alors l'algorithme $DecryptNode(CT_{two}, Sk, z)$ est appelé pour chaque nœud z fils de x. Le résultat est sauvegardé dans $F[z]$. Si au moins l'exploration de l'un des fils du nœud actif x renvoie $\perp$ (Erreur), alors, $DecryptNode(CT_{two}, Sk, x) \rightarrow \perp$. Sinon, soit S_x l'ensemble

des fils du nœud x, de taille k_x, on calcule dans ce cas $F[x]$ en utilisant l'interpolation polynomiale de Lagrange :

Soit $i = index(z)$ and $S'_x = \{index(z) : z \in S_x\}$.

$$\begin{aligned} F[x] &= \prod_{z \in S_x} F_z^{\Delta_{i,S'_x}(0)} \\ &= \prod_{z \in S_x} (e(g,g)^{r \cdot q_z(0)})^{\Delta_{i,S'_x}(0)} \\ &= \prod_{z \in S_x} (e(g,g)^{r \cdot q_{parent(z)}(index(z))})^{\Delta_{i,S'_x}(0)} \\ &= \prod_{z \in S_x} (e(g,g)^{r \cdot q_x(i)}) \cdot \Delta_{i,S'_x}(0) = e(g,g)^{r \cdot q_x(0)} \end{aligned}$$

Nous pouvons maintenant finaliser la définition de l'algorithme **Decrypt(**CT_{two}, Sk**)**. L'algorithme commence par appeler $DecryptNode(CT_{two}, Sk, R)$, nous rappelons que R est le sommet de l'arbre τ. Si S satisfait l'arbre d'accès τ, alors :

$$\begin{aligned} F &= DecryptNode(CT_{two}, Sk, R) \\ &= e(g,g)^{r \cdot q_R(0)} \\ &= e(g,g)^{r \cdot s} \end{aligned}$$

Et calcule :

$$\begin{aligned} B &= \frac{e(D_2, C'_3)}{e(C_3, C'_2)} = \frac{e(g^r h^t, g^{t+s})}{e(g^t, h^{t+s})} \\ &= e(g^r, g^{t+s}) = e(g,g)^{r(t+s)} \end{aligned}$$

$$A = \frac{B}{F} = \frac{e(g,g)^{r(t+s)}}{e(g,g)^{r \cdot s}} = e(g,g)^{rt}$$

L'algorithme retrouve alors la valeur de Ks :

$$Ks = \frac{C_1 \cdot A}{e(C_3, D_1)}$$

En effet, $\frac{Kse(g,g)^{\alpha t/\beta} \cdot e(g,g)^{rt}}{e(g^t, g^{r+\alpha/\beta})} = Ks$

Et au final, le message en clair est obtenu :

$$M = SEncr(C_4, Ks)$$

.

4.3.2 Le modèle de sécurité dans le système considéré

Nous supposons que les fournisseurs de service d'informatique en nuage ou en brouillard sont honnêtes mais curieux. Ils sont uniquement autorisés à exécuter les protocoles correctement, mais qu'ils ne sont pas autorisés à connaître les données privées. Aussi, une propriété de sécurité importante requise dans notre système est la capacité de résister aux attaques par collusion entre utilisateurs. Ces derniers pouvant tenter de combiner leurs droits afin d'augmenter leurs privilèges. Enfin, les canaux de communication sont supposés non sûrs.

Nous définissons la sécurité pour le schéma FCCP-ABE par une expérience, impliquant un challenger $\mathcal{B}$ et un attaquant $\mathcal{A}$. En ce qui concerne les objectifs de sécurité et la capacité de l'adversaire, nous adoptons le modèle de sécurité en texte clair choisi pour notre système et nous nous inspirons de l'expérience de sécurité décrite dans plusieurs travaux, entre autres Waters [144] et Goyal *et al.* [45].

— **Init :** L'adversaire (algorithme) probabiliste $\mathcal{A}$, s'exécutant en un temps polynomial, choisit un ensemble d'attributs pour générer une structure d'accès de défi P^* et l'envoie au challenger $\mathcal{B}$.

— **Setup :** Le challenger $\mathcal{B}$ exécute l'algorithme *Setup* pour générer le paramètre publique PK et le transmet à l'adversaire $\mathcal{A}$.

— **Phase 1 :** L'adversaire répète, autant qu'il estime nécessaire et de façon adaptative, des requêtes pour générer des clés secrètes, chaque fois avec un nouvel ensemble d'attributs S_i.

— **Challenge :** L'adversaire soumet deux messages de longueur égale, m_0 et m_1, et fournit une structure d'accès de défi A^*, de sorte qu'aucun des ensembles S_i de la phase 1 ne satisfasse la structure d'accès donnée. Le challenger tire au hasard une valeur $\gamma \in \{1, 0\}$ et chiffre m_γ sous A^* en exécutant les algorithmes *EncryptCons* et *EncryptUncons*. Le texte chiffré CT^* est retourné à l'adversaire.

— **Phase 2 :** La phase 1 est répétée à condition qu'aucun ensemble d'attributs sélectionné S_j ne satisfasse la structure d'accès fournie en tant que défi.

— **Guess :** L'adversaire doit décider lequel des deux messages a été chiffré, il génère une estimation γ' de γ.

Enfin, la preuve de sécurité est donnée en Annexe A.

4.4 Système de préservation de la vie privée basé sur la Blockchain

Dans notre proposition, la gestion des messages de contrôle d'accès se fait avec une Blockchain, une solution décentralisée permettant de surmonter le problème critique de l'unicité du point d'échec (du point de vue de la confiance). Le contrôle d'accès affiné et la sécurité du stockage sont obtenus grâce à un nouveau schéma de chiffrement basé sur les attributs conjugués à l'informatique en brouillard (FCCP-ABE). Dans cette section, nous allons décrire la conception basée sur le paradigme de la Blockchain.

4.4.1 Modélisation du registre des autorisations d'accès sur Blockchain (RA2-Blockchain)

La Blockchain est utilisée comme base de données distribuée, persistante et infalsifiable pour gérer les messages de contrôle d'accès. De plus, l'un des avantages de l'utilisation de la Blockchain est de fournir une solution pour la révocation des droits d'accès. Avant de décrire notre système, nous allons présenter notre enregistrement d'autorisation d'accès sur la Blockchain sous la forme d'un jeton, désigné *token*, représentant une pseudo crypto-monnaie.

token(idx,@st,@rq, @do) : est une structure de données d'identification permettant de spécifier une autorisation par le propriétaire de l'adresse Blockchain @*do* (et qui va signer

ce jeton) pour que le propriétaire de l'adresse $@rq$ puisse accéder à des données stockées dans $@st$ (adresse Blockchain du fournisseur de stockage) et identifiées par l'index idx.

Nous définissons également un actif numérique applicable, noté idx, qui représente un index permettant d'identifier un enregistrement sur le Cloud. Par ailleurs, le rattachement de cette valeur à une adresse Blockchain du propriétaire de la donnée $@do$ dans la Blockchain, permettra d'enregistrer la propriété de la donnée de façon immuable. La génération de l'idx se fait en calculant l'empreinte d'une séquence de bits. Pour notre système,

$$idx = Hash(CT_{one})$$

Où CT_{one} est le texte de chiffrement intermédiaire vu dans 4.3.1.

Nous définissons également les deux transactions suivantes pour notre Blockchain :

1. $idxGenTrans(idx, @st, @src, @do)$: est la transaction source qui génère des objets idx. Une fois que la valeur idx est calculée par un proxy (PR) ayant une adresse Blockchain $@src$, le proxy diffuse cette transaction pour transférer idx sur le compte $data-owner$ qui possède une adresse Blockchain $@dst$. Le proxy enregistre également une adresse Blockchain $@st$ correspondant au service de stockage de données dans le Cloud.
2. $grantTrans(token(idx, @st, @rq, @do), @src, @dst)$: Cette transaction est utilisée pour transférer le $token$ du compte Blockchain d'un acteur à un autre. Dans notre système, les jetons $token$ sont générés par le propriétaire des données, puis transférés sur le compte du demandeur. Le demandeur ($@rq$) l'envoie au fournisseur de stockage ($@st$) qui le renvoie au propriétaire des données (do). Ce processus, comme on peut le voir sur la Figure 4.3, garantit la traçabilité de la demande et la non duplication de la requête, tout en gardant un certain anonymat sur la Blockchain.

La figure 4.4 montre les différentes interactions au sein de la Blockchain. La génération de l'objet idx est effectuée par le proxy, qui peut être une Cloudlet, ou tout autre nœud de déportation (PR). Lors du transfert des données chiffrées vers le Cloud, le proxy (PR) calcule $idx = Hash(CT_{one})$ et diffuse la transaction $idx_{G}en_{T}rans(idx, @st, @src, @dst)$, à chaque demande d'accès approuvée.

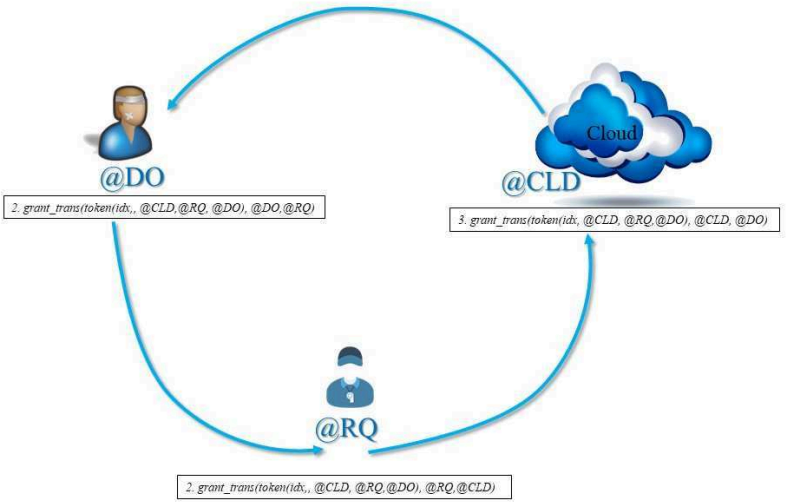


FIGURE 4.3 – Circuit du jeton d'autorisation d'accès dans la Blockchain

4.4.2 Architecture globale du système considéré

Le système que nous proposons comprend six rôles. La figure 4.5 illustre le modèle de système que nous proposons dans le contexte de la santé en ligne : Ce modèle comprend six rôles :

1. *Data-owner (DO)* : génère des données avec ses propres appareils connectés et les stocke dans le Cloud. Le propriétaire des données est le seul à avoir le droit de donner accès à ses données.
2. *Fog-proxy (PR)* : qui est la passerelle ou le proxy (ex. une Cloudlet), situé à l'extrémité du réseau et impliqué dans le processus de chiffrement. Ce proxy est approuvé uniquement pour effectuer les protocoles correctement.

   Le proxy exécute l'algorithme *EncryptUncons* (CT_{one}, τ, Pk) sans pouvoir apprendre aucune partie des données cryptées.
3. *Data-requester (RQ)* : qui est le consommateur de données qui peut être un médecin ou tout autre praticien qui demande l'accès aux données personnelles du propriétaire des données. Pour prouver l'identité du demandeur de données, nous utilisons une infrastructure PKI.

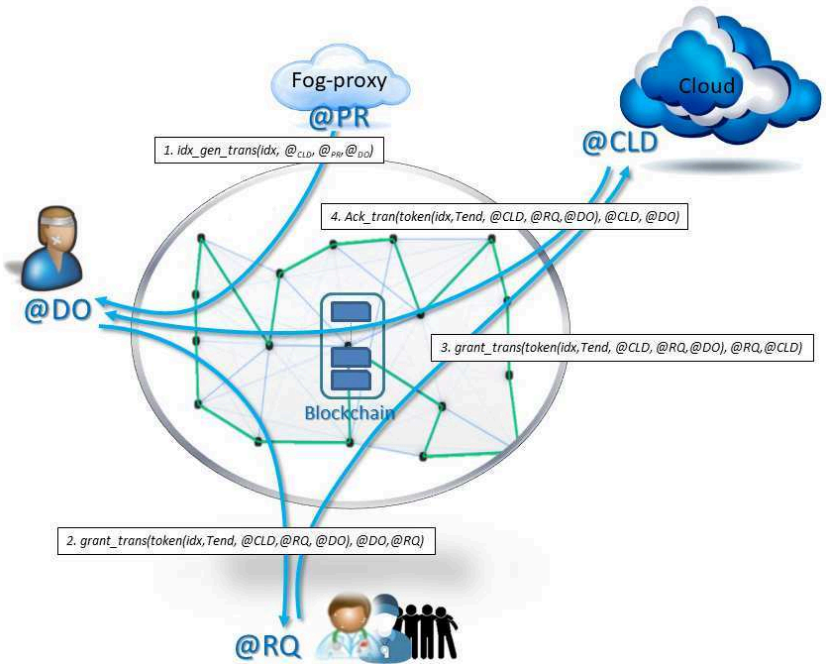


FIGURE 4.4 – Interactions dans l'enregistrement d'autorisation d'accès basée sur Blockchain

4. *Blockchain-A2R (BA2R)* : qui est l'autorité décentralisée de confiance utilisée pour assurer la vérification et la validation des messages échangés sur un réseau non sécurisé.

5. *Storage-provider (CLD)* : qui est instancié par un fournisseur de services de stockage en nuage (Cloud). Cet acteur peut uniquement vérifier si un demandeur anonyme peut fournir une preuve de l'autorisation du *DO* pour accéder à ses données.

6. *Data-sources (DS)* : qui sont les dispositifs producteurs de données (capteurs ou dispositifs de santé connectés utilisés pour collecter des mesures). Dans notre système, *DS* est un périphérique à ressources limitées.

   Notez que nous avons également besoin d'une entité pour vérifier l'identité et les attributs du demandeur. Il peut s'agir d'une infrastructure PKI avec une autorité de confiance. Par exemple, en France, cette autorité de confiance est l'agence de santé numérique française (ASIP Santé) qui tient un répertoire des professionnels de la santé.

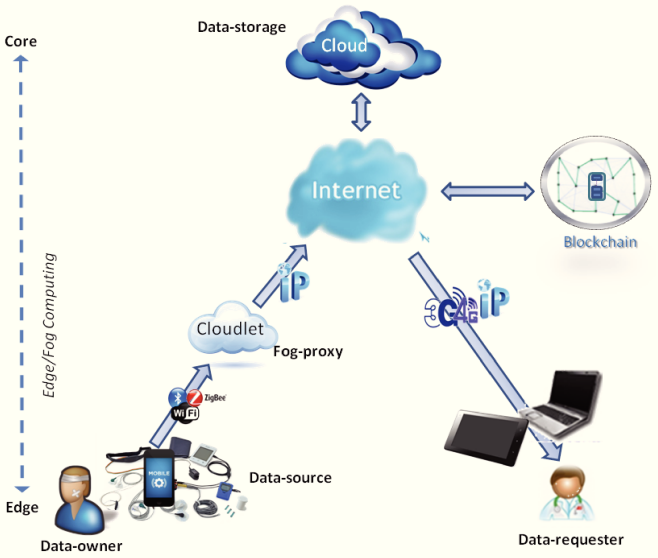


FIGURE 4.5 – Notre architecture système dans le contexte de la santé en ligne

4.4.2.1 Hypothèses sur le système

Avant de détailler notre protocole, nous émettons les hypothèses suivantes :

1. Nous utilisons des primitives de chiffrement symétriques $SEncr(M, K)$, supposées sûres. Cela signifie que pour obtenir M à partir de $SEncr(M, K)$, nous devons connaître K.

2. Les attaques physiques menées sur les équipements IdO (source de données) pour obtenir les clés secrètes stockées ne sont pas prises en compte dans le cadre de cette thèse.

3. Il existe suffisamment de nœuds honnêtes sur le réseau pour maintenir la Blockchain et la préserver de toute altération.

4. Chaque acteur X de notre système conserve, au moins, une adresse de Blockchain notée $@X$ et utilisée pour diffuser des transactions au sein de la Blockchain.

5. Les utilisateurs gèrent leurs clés Blockchain et ses clés PKI de manière sécurisée.

4.5 Protocole de protection des données personnelles, centré sur le propriétaire de données avec un contrôle décentralisé et un chiffrement adapté à l'environnement IdO/Cloud

Notre protocole, centré sur le propriétaire de données, est utilisé pour protéger ses données personnelles et renforcer son pouvoir dans la gestion de ses données. La Blockchain est utilisée comme base de données distribuée, persistante et inviolable pour le contrôle et la vérification de la validité des messages de gestion. FCCP-ABE assure un contrôle d'accès aux données à granularité fine pouvant être mis en œuvre au sein de périphériques limités en ressources. Les phases suivantes du protocole sont illustrées par les Figures 4.6 et 4.7.

— **Phase 1, *Initialisation du système*** : Au cours de cette étape, un paramètre de sécurité λ et un univers d'attributs sont choisis avant d'exécuter la procédure $Setup(\lambda)$ est exécutée pour générer les paramètres publiques et privés de FCCP-ABE(Msk, Pk). De plus, les périphériques *source de données* (DS) sont configurés avec la clé symétrique Ks et *Fog-proxy* (PR) avec Pk.

— **Phase 2, *Enregistrement des données*** (Voir Figure 4.6) : La *source de donnés* (le périphérique limité en ressources), notée DS, chiffre les données :

$$EncryptCons(data, Pk, Ks) \rightarrow CT_{one}$$

et transfère CT_{one} au *Fog-proxy (PR)* ainsi que C_2 au *Data-Owner*. Une fois reçu, PR exécute

$$EncryptUncons(CT_{one}, \tau, Pk) \rightarrow CT_{two}$$

, calcule $idx = Hash(CT_{one})$ et enfin stocke les résultats au niveau du *Fournisseur de stockage* (CLD). Au même moment, PR diffuse la transaction

$$idxGenTrans(idx, @CLD, @PR, @DO)$$

.

— **Phase 3, *Autorisation d'accès*** : Lorsqu'un utilisateur demande des données au *Data-owner* DO, il s'authentifie d'abord auprès du DO lui-même avec son ensemble

d'attributs S. Il utilise une PKI ou toute autre technique d'authentification, même face à face (à l'occasion d'une visite médicale par exemple). Si cette étape d'authentification est achevée avec succès, DO exécute l'algorithme $KeyGen$ avec les paramètres correspondants :

$$KeyGen(Msk, C_2, S) \rightarrow Sk$$

et envoie cette clé secrète à RQ via un canal sécurisée. Simultanément, DO génère le jeton $token(idx, @st, @rq, @do)$ et diffuse la transaction $grantTrans$:

$$grantTrans(token(idx, @st, @rq, @do), @do, @rq).$$

Lorsque cette transaction est approuvée par la Blockchain, cela signifie que DO autorise RQ à accéder aux données identifiées par idx et stockées dans CLD.

— **Phase 4, *Accès aux données*** : Lorsque RQ reçoit l'autorisation d'accéder aux données (*phase 3*), il diffuse une transaction $grant_Trans(token, @rq, @st)$ pour transférer le jeton $token(idx, @st, @rq, @do)$ au service de stockage (le Cloud), ayant l'adresse Blockchain $@st$. Le Cloud peut alors vérifier que le propriétaire de l'adresse $@rq$ est une entité légitimement autorisée à accéder aux données identifiées par idx. Enfin, après que RQ ait prouvé qu'il dispose de la clé secrète associée à l'adresse Blockchain $@rq$ (avec un simple protocole nonce-challenge, non détaillé ici), le Cloud envoie le texte chiffré CT_{two} à RQ et diffuse la transaction $grant_Trans$ afin de renvoyer le $token$ à $@do$ et d'informer celui-ci que ses données ont été consultées. Enfin, RQ utilise sa clé secrète Ks pour récupérer les données.

La Figure 4.7 schématise le protocole d'échange des phases 3 et 4.

4.6 Conclusion

Dans ce chapitre, nous avons d'abord présenté notre solution de chiffrement basée sur les attributs, avec possibilité de déportation des traitements lourds vers les nœuds de l'informatique en brouillard. Ensuite nous avons intégré ce schéma de chiffrement dans un protocole, dont le contrôle de la validité des échanges se fait via la Blockchain. La Blockchain permet d'assurer des échanges sécurisés de façon décentralisée et assure l'anonymat des échanges.

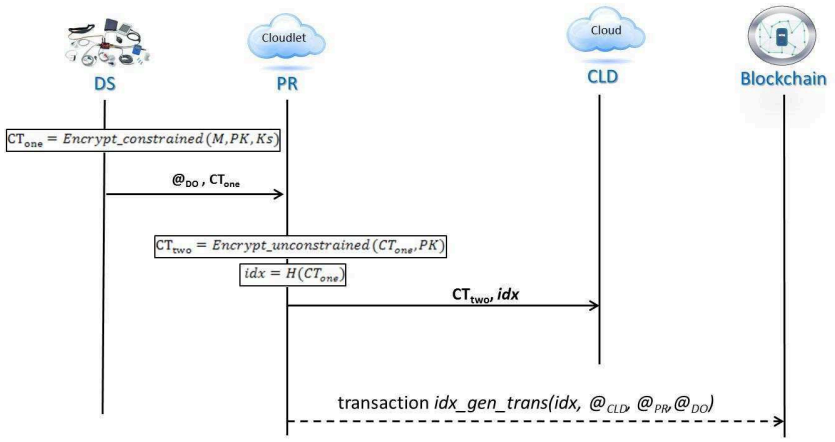


FIGURE 4.6 – Enregistrement des données sur le Cloud

Le principal obstacle pratique que nous avons identifié pour la mise en œuvre de la Blockchain est le problème de passage à l'échelle, en termes de nombre de transactions pouvant être validées par unité de temps, comme nous le verrons dans la partie évaluation des performances du chapitre 4.

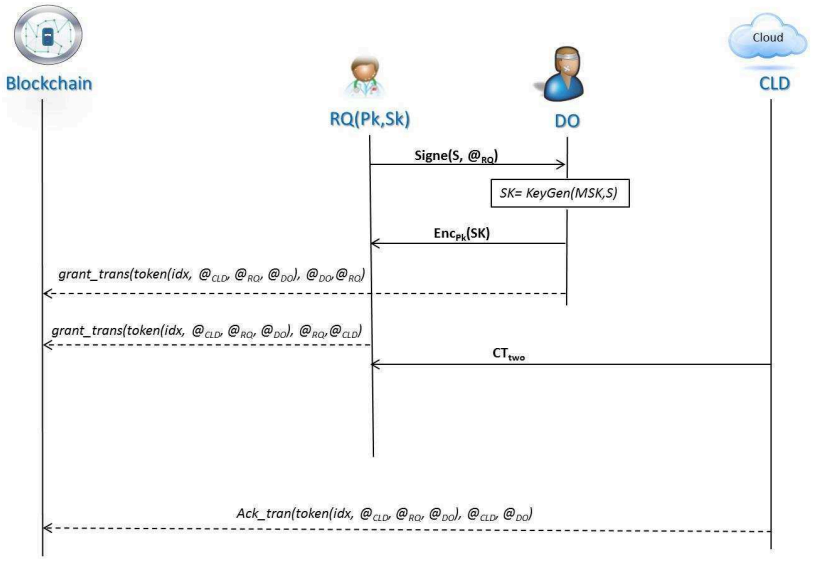


FIGURE 4.7 – Autorisation puis accès aux données

Chapitre 5

Analyse théorique et expérimentale du modèle de protection des données personnelles dans un environnement IdO/Cloud, basé sur le chiffrement par attributs

5.1 Introduction

Dans ce chapitre, nous allons effectuer une analyse théorique et une discussion de notre solution de sécurité, suivie d'une analyse expérimentale du CP-ABE sur différentes plates-formes matérielles. Nous donnerons les résultats de l'évaluation de la performance de notre schéma FCCP-ABE, comparés à d'autres propositions, en termes de performances d'opérations de chiffrement au niveau des dispositifs contraints. Il sera également question de l'illustration expérimentale des capacités de différentes plate-formes à effectuer des calculs d'exponentiation et de couplage bilinéaire. Enfin, nous analyserons à travers des modèles de cas d'utilisation, les capacités d'implémentation de notre solution avec la

Blockchain.

5.2 Analyse théorique de la sécurité et de la protection de la vie privée

Cette analyse est valable sous les hypothèses énumérées dans 4.4.2.1.

5.2.0.1 La sécurité relative à ABE

Sur la base de l'intuition de sécurité de FCCP-ABE, pour récupérer un message secret, un adversaire qui intercepte la communication, entre le périphérique source de données et le nœud proxy, doit calculer ou deviner la clé secrète symétrique Ks. Ceci n'est pas possible en raison des hypothèses 1 et 2 (chiffrement symétrique supposé sûr et absence d'attaque physique sur les noeuds). L'autre possibilité est de résoudre $e(g,g)^{\alpha t/\beta}$ et de retrouver la valeur aléatoire t. Cela n'est pas possible dans un temps polynomial grâce à l'hypothèse DBDH et aux valeurs aléatoires utilisées pour générer les clés secrètes des utilisateurs Sk. Le même raisonnement peut être appliqué à l'adversaire qui écoute toutes les communications du système.

De plus, les attaques par collusion entre utilisateurs sont inopérantes, puisque nous avons généré une valeur aléatoire pour randomiser la clé privée de chaque utilisateur. Ceci, en conformité avec la proposition de Bethencourt *et al.* [19]. Il est également important de noter que le Cloud ainsi que le nœud proxy ne peuvent pas récupérer le message secret d'origine. En effet, le Cloud n'est pas impliqué dans le processus de chiffrement/déchiffrement des données. De plus, la partie partielle du chiffré, calculée par les périphériques proxy, n'est pas suffisante pour récupérer la valeur de $e(g,g)^{\alpha t/\beta}$ ou t.

5.2.0.2 La sécurité relative à la Blockchain

Aussi, le système proposé est basé sur le paradigme de la Blockchain pour assurer une vérification et une validation autonomes des événements de contrôle d'accès. Par conséquent, il n'y a pas d'autorité de confiance centralisée, qui risque d'être un point critique pour la sécurité du système. L'utilisation des pseudonymes dans les transactions de la Blockchain

permet de préserver la confidentialité des utilisateurs. L'anonymat dans la Blockchain repose sur le fait que les utilisateurs peuvent créer un nombre illimité d'adresses anonymes [97]. Notre conception basée sur une Blockchain protège contre les adversaires qui veulent compromettre les nœuds du système. Cette exigence est garantie grâce aux transactions signées numériquement, qui garantissent qu'un adversaire ne peut pas falsifier un message de contrôle ni emprunter l'identité d'un utilisateur légitime. Par conséquent, sur la base du processus décentralisé de vérification et de validation de la Blockchain, le seul moyen de corrompre le réseau est de prendre le contrôle de la majorité des ressources du réseau.

5.3 Analyse expérimentale

Pour valider notre schéma, nous avons mené des expériences en utilisant les plates-formes Raspberry Pi en tant que périphériques contraints et une station de travail faisant office de nœud d'externalisation (nœud proxy) telle qu'une Cloudlet.

La simulation expérimentale du schéma ABE est réalisée à l'aide de la bibliothèque de cryptographie basée sur le couplage (PBC-Library) [84]. Le poste de travail fonctionne sous Ubuntu 16.04 LTS 64 bits, avec un processeur Intel(R) Core(TM) i5-4590s de 3,00 GHz et 8 Go de RAM. Le modèle B de la Raspberry Pi 3 utilise un système d'exploitation Raspbian, doté d'un processeur ARMv8 quad-core 1,2 GHz à 64 bits et d'une mémoire vive de 1 Go.

Pour atteindre un niveau de sécurité de 128 bits, nous avons modifié quelque peu les paramètres de couplage Type-A de PBC-Library d'origine afin d'utiliser des courbes elliptiques de 256 bits basées sur la courbe supersingulaire $y^2 = x^3 + x$ sur un corps fini de 1536 bits. Le nombre d'attributs est $N = \{5, 10, 20, 30, 40\}$. Nous considérons que cette gamme est représentative des applications du monde réel. Pour éviter les erreurs, les résultats expérimentaux sont la moyenne de 10 essais.

5.3.1 Évaluation des performances de CP-ABE sur différentes plates-formes

Pour montrer le défi que représente l'implémentation d'ABE dans des appareils à ressources limitées par rapport à des équipements sans contraintes, nous simulons l'algorithme

de chiffrement d'origine du schéma CP-ABE [19] avec le nombre d'attributs définis à partir de $N = \{5, 10, 20, 30, 40\}$.

En ce qui concerne le cas d'utilisation de la surveillance de l'état des individus, nous analysons les performances de l'opération de chiffrement, car il s'agit de l'opération la plus dimensionnante effectuée par les périphériques chargées de collecter les données. Les résultats sont donnés dans la figure 5.1.

La figure 5.2 montre le temps d'exécution des opérations de calcul significatives dans $\mathbb{G}_0$ et $\mathbb{G}_T$ (voir le tableau 5.1).

Toutes ces expériences sont effectuées à la fois sur un équipement contraint (Raspberry Pi) et sur un autre non contraint (PC). Comme on peut s'y attendre, les opérations cryptographiques exécutées dans un Raspberry Pi sont nettement plus lentes que leur exécution sur un poste de travail. Ces résultats expérimentaux motivent notre choix pour le modèle d'architecture avec déportation des calculs lourds vers des dispositifs sans contraintes de ressources.

TABLE 5.1 – Liste des opérations informatiques significatives dans $\mathbb{G}_0$ et $\mathbb{G}_T$

Abbreviations	**Meanings**
Expo G0	Exponentiation dans $\mathbb{G}_0$
Expo GT	Exponentiation dans $\mathbb{G}_T$
Mul G0	Multiplication dans $\mathbb{G}_0$
Mul GT	Multiplication dans $\mathbb{G}_T$
Rand G0	Random generation dans $\mathbb{G}_0$
Rand GT	Random generation dans $\mathbb{G}_T$
Pairing	Couplage bilinéaire $e(\mathbb{G}_0, \mathbb{G}_0)$

5.3.2 Comparaison des performances

Afin de démontrer la validité du FCCP-ABE proposé, nous comparons ses performances de chiffrement, sur un périphérique contraint, à une sélection de schémas issue de l'état de l'art : Asim *et al.* [9], Bethencourt *et al.* [19] , Zhou et Huang [157] et Zhang *et al.* [153]. On s'est focalisé sur les schémas répondant aux besoins de notre cas d'étude, à savoir une capacité de déportation relative à l'opération de chiffrement.

Comme le montre la Figure 5.3, le temps d'exécution de notre schéma est constant et

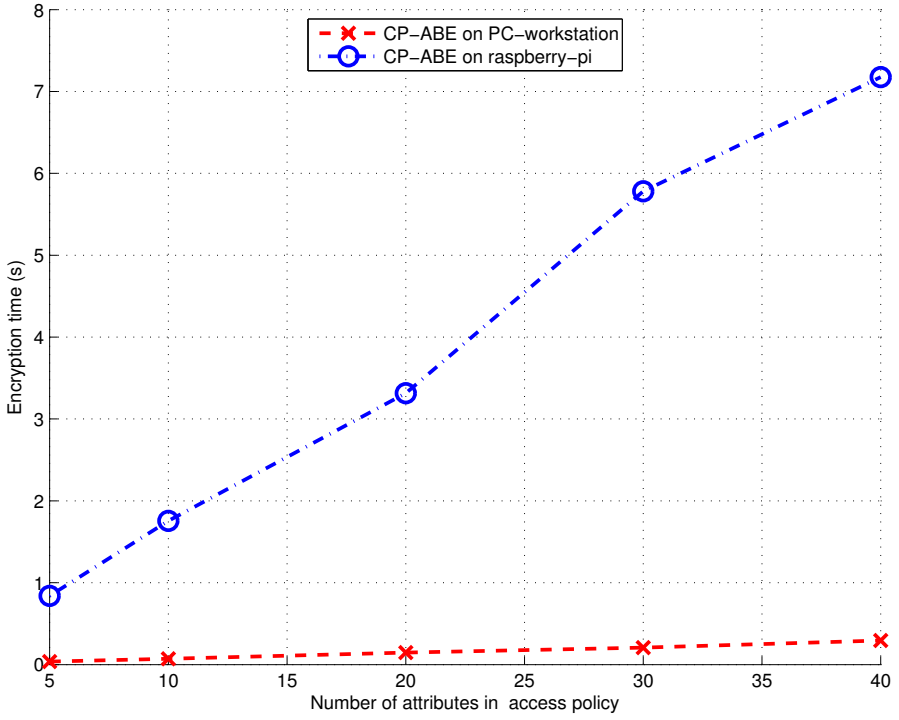


FIGURE 5.1 – Comparaison entre le temps d'exécution du chiffrement CP-ABE sur la station de travail PC et sur une carte (raspberry Pi)

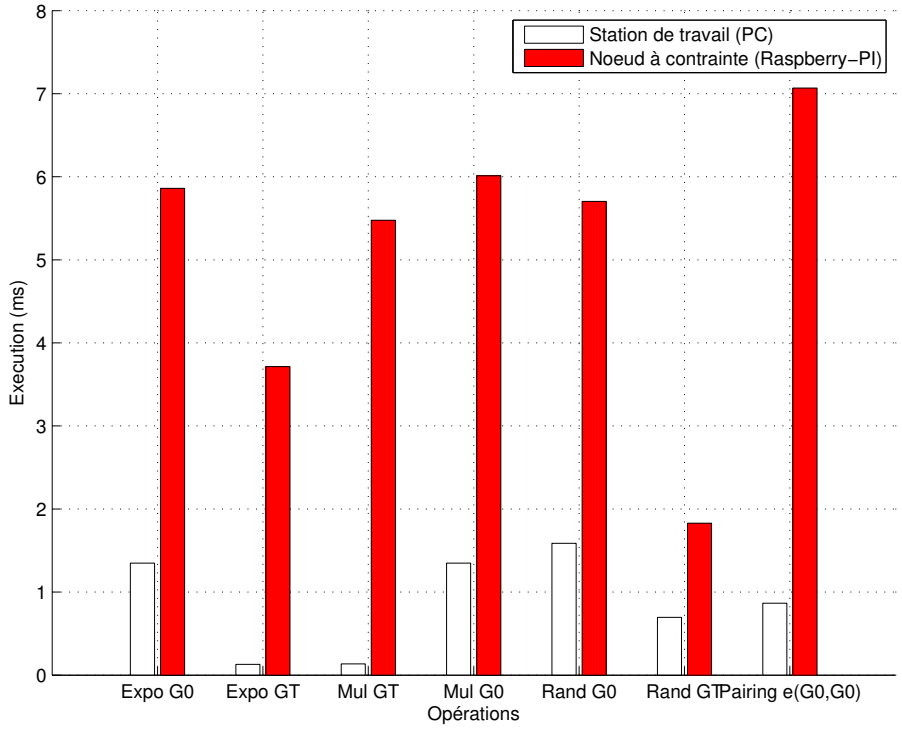


FIGURE 5.2 – Temps d'exécution des opérations du groupe cyclique sur la station de travail et sur Raspberry-Pi

indépendant du nombre d'attributs. De plus, comparé aux solutions existantes [19] et [9], FCCP-ABE est le plus efficace dans le contexte de périphériques contraints.

Notre modèle est légèrement meilleur que celui proposé par Zhang *et al.* [153] (qui se trouve être à notre connaissance le plus efficace au moment de la rédaction de cette thèse) en termes de temps d'exécution. Cependant, la conception proposée dans [153] implique deux échanges de messages entre le périphérique IdO et le nœud proxy, en plus des messages nécessaires à l'établissement et le maintien de la connexion, alors que notre système n'a besoin que de transférer CT_{one} du périphérique IdO vers le nœud proxy.

De plus, comme le montre le tableau 5.2, pendant le processus de chiffrement, la longueur du texte chiffré générée par le périphérique IdO dépend du nombre d'attributs utilisés dans la structure d'accès n, tandis que dans notre contribution, la longueur du texte chiffré généré par les dispositifs IdO est indépendante du nombre d'attributs. Ceci suggère que notre proposition utilise moins de ressources mémoire et consomme moins d'énergie pour la transmission des données.

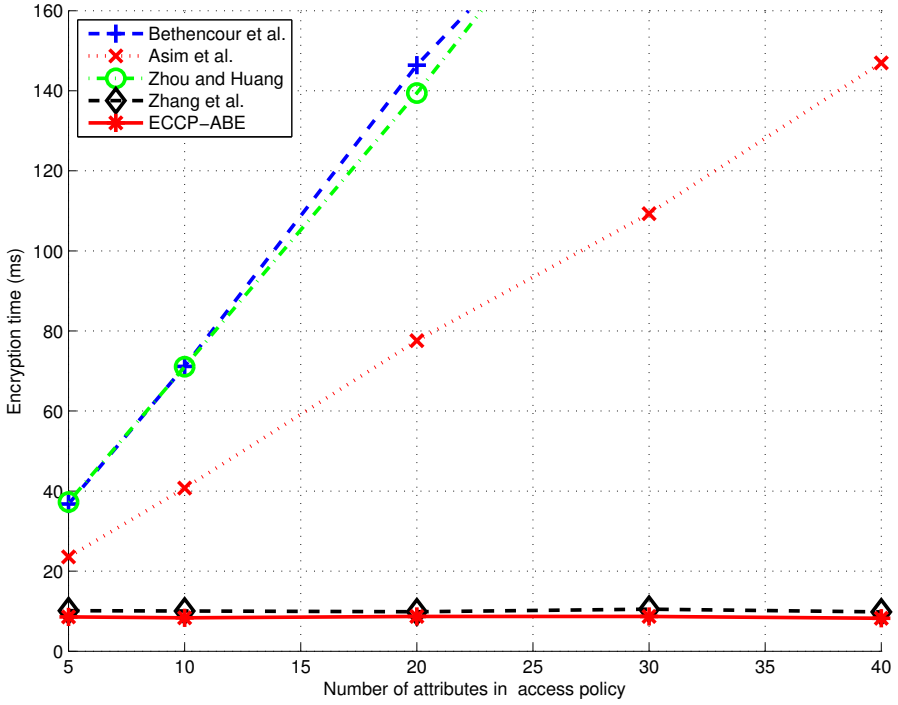


FIGURE 5.3 – Comparaison des performances de chiffrement de FCCP-ABE avec d'autres schémas populaires

TABLE 5.2 – Comparaison de la taille du chiffré de FCCP-ABE et du schéma de Zhang *et al.* [153]

Schéma	**Taille du chiffré**
FCCP-ABE	$\mid \tau \mid + \mid SEncr(M, Ks) \mid +2 \mid \mathbb{G}_0 \mid + \mid \mathbb{G}_T \mid$
Zhang et al.	$\mid \tau \mid + \mid SEncr(M, Ks) \mid +(3+n) \mid \mathbb{G}_0 \mid + \mid \mathbb{G}_T \mid$

$\mid * \mid$ *: taille en Bits dans* $*$*.* n *: le nombre d'attributs dans la structure d'accès.*

5.4 Analyse de l'adaptabilité de la Blockchain à différentes échelles

L'évaluation de la Blockchain va porter sur la capacité de cette technologie à prendre en charge le nombre de transactions nécessaires au bon fonctionnement de notre protocole. L'analyse de notre protocole fait ressortir que le nombre de transactions $idxGenTrans$ est le plus dimensionnant pour évaluer l'ordre de grandeur de ce nombre de transactions, car cette transaction est générée à chaque enregistrement effectué par le nœud proxy (PR) sur le Cloud, alors que la transaction $grantTrans$ n'est généré que pour transmettre l'autorisation d'accès, quand cela est requis.

5.4.1 Méthodologie

Nous avons confronté des scénarios traduisant différents ordres de grandeur du nombre d'enregistrements nécessaires par jour (et donc de transaction *idx_gen_trans*) aux capacités actuelles des Blockchains à valider des transactions par unité de temps.

Pour la Blockchain, nous avons utilisé un premier modèle qui est la Blockchain Bitcoins. A partir des données disponibles en ligne[1], nous avons calculé une moyenne de 24 8244.85 transactions par jour avec un écart-type de 53 859.27, soit en moyenne 0.348 seconde par transaction. La moyenne calculée est proche de celle obtenue par simulation par Memon *et al.* [91].

Nous avons, également, pris en considération le cas des Blockchains privées, basées sur Ethereum, dont Rouhani et Deters [117] ont analysé les performances de deux logiciels client

1. www.blockchain.com

Geth[2] et Parity[3]. Bien que l'utilisation d'une Blockchain privée (faible décentralisation) s'écarte de l'esprit de notre solution, à savoir, une confiance décentralisée, elle reste un cas d'implémentation possible pour notre protocole (bien que non recommandé).

5.4.2 Évaluation des capacités de la Blockchain

Les résultats expérimentaux de [117] pour le client Parity d'Ethereum sont donnés dans le Tableau 5.3.

TABLE 5.3 – Performance d'une Blockchain privée sous Ethereum avec le client Parity [117]

Nombre de transactions (t_x)	**Temps pour la validation T(minutes)**
1000	1.74
2000	3.48
3000	5.13
4000	6.91
5000	8.71
10000	18.52

Ces résultats nous permettent de déduire le modèle linéaire par régression suivant :

$$T = 0.0019t_x - 0.3905 \tag{5.1}$$

Avec T le temps en minutes nécessaire pour valider t_x transactions dans la Blockchain.

5.4.3 Évaluation du nombre de transactions

Pour le nombre de transactions à traiter, nous avons fait varier le nombre d'individus suivis de 10 à 500. En l'absence de statistiques adéquates, nous avons considéré que chaque population a des besoins d'enregistrement de ses données normalement distribuées sur différents scénarios :

— S0 : toutes les 5 secondes, soit 17 280 enregistrements par jour.

— s1 : toutes les minutes, soit 1 440 enregistrements par jour.

— s2 : toutes les 15 minutes, soit 96 enregistrements par jour.

2. www.geth.ethereum.org
3. www.parity.io

TABLE 5.4 – Estimation des performances de la Blockchain de type Bitcoin

Nb d'individus	**Nb de transactions t_x**	**T(heure)**	**T(jour)**
10	37 564	3.63	0.15
50	247 137	23.89	1.00
100	358 361	34.65	1.44
200	661 532	63.96	2.66
250	880 063	85.08	3.55
500	1 899 485	183.64	7.65

TABLE 5.5 – Estimation des performances de la Blockchain Ethereum avec le client Parity

Nb d'individus	**Nb de transactions t_x**	**T(heure)**	**T(jour)**
10	37 564	1.18	0.05
50	247 137	7.82	0.33
100	358 361	11.34	0.47
200	661 532	20.94	0.87
250	880 063	27.86	1.16
500	1 899 485	60.14	2.51

— s3 : toutes les heures, soit 24 enregistrements par jour.

— s4 : une fois par jour, soit un enregistrement par jour.

Le nombre de transactions par jour à prendre en compte en fonction du nombre d'individus à surveiller, ainsi que le temps nécessaire à leur validation sont donnés respectivement par les tableaux 5.4 pour la Blockchain de type Bitcoin et 5.5 pour la Blockchain Ethereum.

Les résultats portés sur les tableaux 5.4 et 5.5 ainsi que la Figure 5.4 confirment la difficulté du passage à l'échelle des Blockchains. Cette problématique est un important axe de réflexion pour la communauté scientifique.

Pour ce qui nous concerne, les résultats obtenus montrent déjà la supériorité de la Blockchain Ethereum sur la Bitcoin. La Figure 5.5 montre les limites de cette technologie pour rester dans un cas d'utilisation réaliste, à savoir un nombre d'individus à surveiller ne dépassant pas un ordre de grandeur de 200 individus. Ce choix correspond à un traitement par la Blockchain ne dépassant pas une journée pour valider une transaction. Il est clair que cette limite va dépendre des besoins finaux réels.

Ceci étant, la solution reste viable si le besoin d'enregistrement des données est assez raisonnable. Le tableau 5.6 indique le nombre d'individus pouvant être surveillés sous

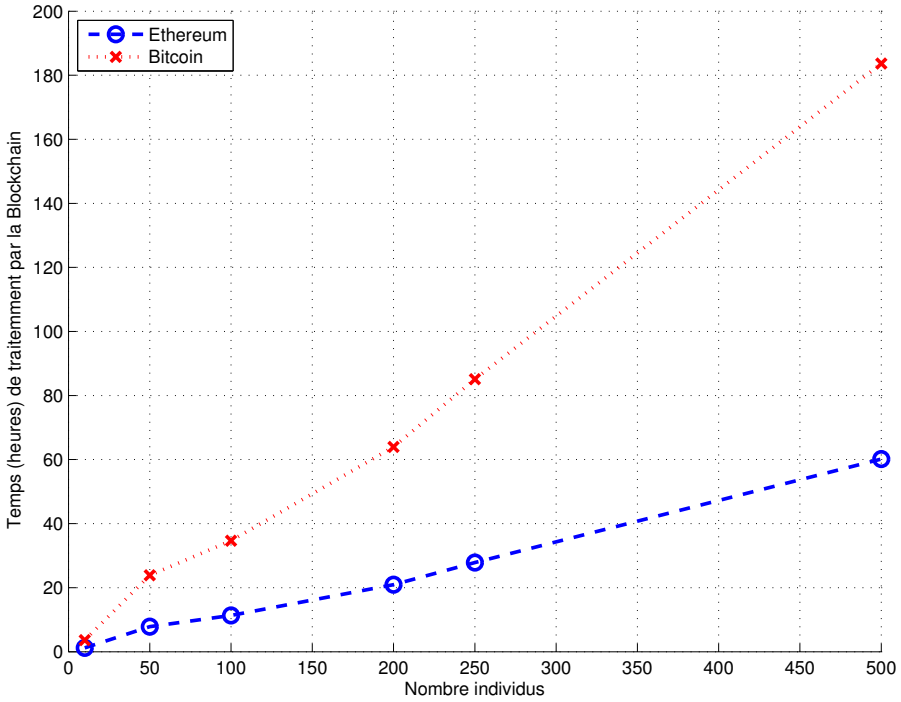


FIGURE 5.4 – Temps de traitement par la Blockchain du nombre de transactions requis

TABLE 5.6 – Nombre d'individus pris en charge sous différents scénarios d'enregistrement de données

Scénarion	Bitcoin	Ethereum
s0	14	44
s1	172	526
s2	2586	7 897
s3	10344	31 588
s4	248245	758 100

différents scénarios avec l'hypothèse d'une capacité de 248244,85 transactions par jour pour la Blockchain Bitcoin et 758100.26 transactions/jour pour Ethereum (calculé à partir de l'expression 5.1).

5.5 Impacte de l'algorithme de génération de clés de déchiffrement sur le système

L'algorithme $KeyGen(Msk, C_2, S)$ de génération de clé secrète, nécessite de disposer de la composante C_2. Ceci implique que l'entité exécutant cet algorithme (le propriétaire de

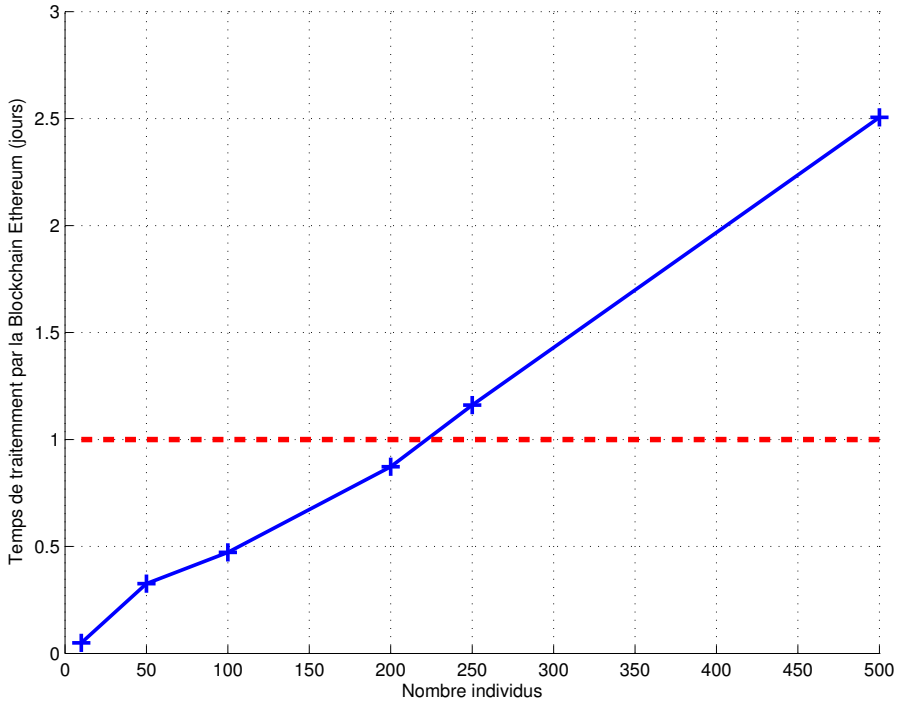


FIGURE 5.5 – Temps de traitement par la Blockchain Ethereum du nombre de transactions requis

la donnée ou une autorité de gestion des clés) doit sauvegarder pour chaque enregistrement la valeur de C_2 ou à défaut la valeur de t utilisée pour générer le CT_{one} et ceci même si on utilise la même structure d'accès. Pour rappel, l'algorithme $EncryptCons(M, Pk, Ks)$ sélectionne un élément $t \in \mathbb{Z}_p$ aléatoirement et calcule $C_2 = h^t$, qui est une des composante du chiffré intermédiaire CT_{one}.

En reprenant les scénarios précédents, en termes de nombre d'enregistrement, il est facile de calculer la surcharge de messages à transmettre que l'objet connecté devra assurer.

TABLE 5.7 – Messages à transmettre par seconde en fonction des différents scénarios

Scénarion	**Fréquence (Messages/seconde)**
s0	0.200000
s1	0.016667
s2	0.001111
s3	0.000278
s4	0.000012

Le tableau 5.7 montre que la surcharge de messages induite par notre modèle FCCP-

ABE, en fonction des différents scénarios, reste assez maîtrisable. Néanmoins, l'entité chargée de la génération des clés de déchiffrement aura à sauvegarder un indexe des valeurs de C_2, qui va dépendre de la taille en bits d'un élément de $\mathbb{G}_0$. Nous considérons deux cas de figure :

— c'est le propriétaire de la donnée qui se charge de générer les clés de déchiffrement, dans ce cas nous considérons le nombre d'enregistrement par jour.

— c'est une autorité externe qui se charge de la gestion des clés de déchiffrement, dans ce cas nous considérons le nombre d'individus à prendre en charge (voir le nombre de transactions par jour dans les tableaux 5.4 et 5.5).

Les tableaux 5.8 et 5.9 donnent respectivement la taille en mémoire à prévoir pour chacun des deux cas de figure. On suppose que la taille en bits d'un élément de C_2 est de 160 bits, correspondant à un niveau de sécurité équivalent à 80 bits.

TABLE 5.8 – Espace mémoire pour sauvegarder les valeurs de C_2 selon les différents scénarios

Scénarion	**Nombre d'enregistrements / jours**	**Mémoire (Mbyte/jour)**
s0	17280	0.345600
s1	1440	0.028800
s2	96	0.001920
s3	24	0.000480
s4	1	0.000020

TABLE 5.9 – Espace mémoire pour sauvegarder les valeurs de C_2 suivant le nombre d'utilisateur (cas d'une autorité de gestion des clés)

Nbr d'individus	**Nombre d'enregistrements / jours**	**Mémoire (Mbyte/jour)**
10	37 564	1
50	247 137	5
100	358 361	7
200	661 532	13
250	880 063	18
500	1 899 485	38
50 000	187 820 413	3 756

On peut considérer que la nécessité de transmettre la composant C_2, à chaque enregistrement, est une insuffisance de notre schéma de chiffrement. En effet, bien que notre évaluation montre que l'impacte en termes de messages supplémentaires à généré par l'objet

connecté peut être maîtrisé, il reste une charge supplémentaire pour la gestion des clés de déchiffrement pour l'entité en charge de cette opération. Ceci en plus du coût en mémoire.

Nous proposons, pour réduire ces coûts liés à la nécessité de disposer de la valeur de C_2 pour générer une clé de déchiffrement, d'initialiser les objets à contraintes (chargé de calculer CT_{one}) par une valeur pré-calculé en fonction de t et de la clé secrète principal. Ces éléments seront fixe pour une session défini.

5.6 Conclusion

Dans ce chapitre, nous avons effectué une analyse expérimentale et théorique pour valider notre modèle de protection des données personnelles. D'abord, nous avons effectué une analyse de la sécurité de notre schéma à l'ombre des hypothèses de travail établies. Ensuite, nous avons mis en évidence l'opportunité de notre choix d'externaliser les traitements ABE à travers la comparaison des performances du schéma CP-ABE et des principales opérations mathématiques sur une station de travail et une carte Raspberry-Pi. Il s'est ensuivi une comparaison expérimentale des performances de chiffrement de notre schéma FCCP-ABE avec d'autres propositions données dans la littérature. Enfin, nous avons terminé note chapitre par une analyse et une simulation des performances de la Blockchain qui nous a permis de constater les limites de cette technologie, bien que prometteuse, en termes de passage à l'échelle.

Dans la partie suivante, nous allons aborder la deuxième problématique de recherche de notre thèse, à savoir, l'évaluation de la confiance.

Troisième partie

Évaluation de la confiance

Chapitre 6

Pré-requis pour l'évaluation de la confiance à base des cartes de Kohonen

6.1 Introduction

Dans ce chapitre, nous allons présenter les pré-requis nécessaires d'une part, pour appréhender la problématique de l'évaluation de la confiance dans un système basé sur l'externalisation des services, et d'autre part, pour maîtriser suffisamment les concepts nécessaires à la construction de notre solution. Nous commencerons par présenter les systèmes de recommandation et leur utilisation comme système d'évaluation de la confiance. Par la suite, nous présenterons brièvement les paradigmes des cartes auto-organisatrices et des K-moyennes.

Enfin, nous tenons à signaler qu'en plus des références issues de la littérature, la rédaction du présent chapitre s'inspire également des thèses de Rousset [118] et Charif [25]

ainsi que du cours Cnam de Crucianu et Yacoub [33].

6.2 Système de recommandation, outil d'évaluation de la confiance

6.2.1 Terminologie relative aux systèmes de confiance

Évaluation : Une évaluation est une valeur numérique ou sémantique représentant l'appréciation d'un utilisateur pour un service. Il peut s'agir d'une mesure ou d'une estimation donnée à un paramètre indicateur de la qualité de service (e.x. bande passante, latence, temps de réponse) ou bien d'une évaluation qui peut être globale ou qui concerne un ou plusieurs attributs du service concerné.

Prédiction d'une évaluation : la prédiction d'une évaluation consiste à calculer l'évaluation la plus probable qu'un utilisateur sélectionné aurait attribué à un service donné.

Recommandation : La recommandation consiste à évaluer une liste de services et de proposer, par la suite, à l'utilisateur sélectionné celui qui lui conviendrait le plus. La recommandation exploite la prédiction de l'évaluation de chaque service d'une certaine sélection de services, en vue de lui recommander celui qui a la meilleure évaluation.

6.2.2 Système de recommandation

Dans la littérature, les systèmes de recommandation suivent généralement trois approches populaires : le filtrage fondé sur le contenu, le filtrage collaboratif et le filtrage hybride [74; 124].

S'agissant des systèmes de recommandation basés sur le contenu, l'idée de base est de trouver les services dont le contenu optimise la correspondance avec le profil de l'utilisateur [110]. Les profils des utilisateurs sont construits explicitement en interrogeant les utilisateurs sur leurs préférences ou implicitement en créant un modèle basé sur l'historique des services qu'ils ont appréciés.

Cette approche présente l'avantage de l'indépendance des utilisateurs entre eux et de la capacité de recommander des items qui n'ont pas encore été évalués par aucun utilisateur

[124]. Cependant, les approches de recommandation basées sur le contenu présentent également certaines limitations telles que le démarrage à froid de l'utilisateur (nouvellement introduit), l'exigence d'un nombre suffisant d'attributs descriptifs et le problème de sur-spécialisation. Ce dernier problème résulte du fait que le système ne peut recommander que des éléments similaires au profil de l'utilisateur ou au comportement antérieur de l'utilisateur [124].

Le filtrage collaboratif (FC) est une approche basée sur le partage d'opinions entre les utilisateurs pour prédire une opinion sur un élément donné qui n'a pas été apprécié par un utilisateur donné. Ce modèle repose sur l'hypothèse que si les utilisateurs ont les mêmes préférences pour un ensemble d'éléments, ils auront probablement les mêmes préférences pour un autre ensemble d'éléments [132]. Enfin, l'approche hybride combine les deux techniques dans un seul modèle.

Pour les besoins de notre travail, qui se base sur l'approche collaborative pour évaluer la confiance d'un service, nous détaillons ci-après le filtrage collaboratif.

6.2.3 Le filtrage collaboratif (FC)

Le filtrage collaboratif est un sous-domaine de l'apprentissage automatique [135], qui a pour objectif de prédire les paramètres sélectionnés en fonction des données historiques connues sur ce paramètre et s'appuie sur la notion de similarité entre utilisateurs ou entre services. L'entrée du FC est une matrice R d'évaluation ou de mesures, de taille $n \times m$ (n utilisateurs, m éléments/items/services). La sortie est la prédiction de l'évaluation que l'utilisateur actif u aurait probablement donné au service s inconnu de u. Deux approches sont utilisées pour le filtrage collaboratif : l'approche mémoire et l'approche modèle [34; 124].

L'approche mémoire ou voisin utilise les retours d'expérience des utilisateurs (évaluations) stockés en mémoire pour permettre la prédiction. Deux méthodes sont utilisées pour cette approche : la méthode basée sur l'utilisateur, ex. Resnick *et al.* [114] et la méthode basée sur les items/services ex. Sarwar *et al.* [122]. Dans le premier cas, la prédiction d'une note qui sera donnée par un utilisateur sélectionné u pour un service donné s est calculée à l'aide des valeurs données par les voisins (en termes de similarité d'utilisateurs) de l'utilisateur sélectionné u. Pour le deuxième cas, on utilise les notes données par l'utili-

sateur u aux voisins (en termes de similarité de services) de l'élément/service s sélectionné. L'approche mémoire est simple à mettre en œuvre mais devient coûteuse en calcul dans les cas d'utilisation avec un nombre élevé d'utilisateurs et/ou de services (problème de passage à l'échelle), car nécessitant un grand nombre de calculs pour établir les similarités.

L'approche modèle, a été développée pour surmonter les limitations de l'approche mémoire. Dans cette approche, un modèle hors ligne est construit (appris) à partir des données d'entrée initiales, permettant de réduire la dimension de l'espace. Le calcul de la similarité et de la prédiction se fera dans ce nouvel espace, qui est plus dense. De nombreuses méthodes sont utilisées pour construire le modèle telles que la décomposition en valeurs singulières (SVD) Bell *et al.* [15], l'analyse en composantes principales (PCA) Goldberg *et al.* [42], la classification bayésienne Breese *et al.* [21] et les k-means Wu *et al.* [146]. La principale limite de cette approche est la sensibilité aux données manquantes ou clairsemées.

L'une des limites du filtrage collaboratif est la difficulté à prendre en charge les nouveaux services ou utilisateurs. Pour remédier à ce problème, des solutions sont proposées dans la littérature, comme s'appuyer sur un ensemble représentatif des utilisateurs ou des services ou l'utilisation de l'approche basée sur le contenu pour démarrer.

6.2.4 Prédiction de l'évaluation

Pour le filtrage collaboratif, la prédiction d'une évaluation qu'un utilisateur u aurait donné à un service s inconnu de u est obtenue en calculant une moyenne des évaluations, pondérée par un coefficient qui mesure la similarité. Pour le filtrage basé sur les utilisateurs, nous retrouvons dans la littérature [59; 114] les deux formules 6.1 et 6.2 suivantes, qui sont généralement admises. Le tableau 7.1 contient les abréviations utilisées dans ce chapitre et également utilisées dans le chapitre suivant.

$$r'_{us} = \frac{\sum_{v \in U_s} Sim(u,v) \times r_{vs}}{\sum_{v \in U_s} Sim(u,v)} \tag{6.1}$$

La formule 6.2 permet d'ajuster les évaluations des utilisateurs, qui faut-t-il le rappeler peuvent être subjectives.

TABLE 6.1 – Liste des abréviations

Abréviations	**Définitions**
u	Identifiant de l'utilisateur
v	Identifiant du service ou de l'item
r_{uv}	Évaluation donnée par l'utilisateur u au service/item v
$\overline{r_u}$	Moyenne des évaluations données par l'utilisateur u
r'_{uv}	Valeur prédite pour l'évaluation de l'utilisateur u pour le service/item v
U_v	Ensemble des utilisateurs ayant évalué le service v.
$U_{i,j}$	Ensemble des utilisateurs ayant évalué à la fois le service/item i le service/item j
$I_{x,y}$	Ensemble des services/items co-évalués par l'utilisateur x l'utilisateur y
$Sim(i,j)$	Mesure de similitude entre les entités (services/items ou utilisateurs) i et j

$$r'_{us} = \overline{r}_u + \frac{\sum_{v \in U_s} Sim(u,v) \times (r_{vs} - \overline{r_v})}{\sum_{v \in U_s} Sim(u,v)} \tag{6.2}$$

Où r'_{us} est l'évaluation à prédire de l'utilisateur u sur le service s. U_s est l'ensemble des utilisateurs qui ont évalué le service s. r_{xs} est l'évaluation de l'utilisateur x pour le service s. $\overline{r}_x$ est la moyenne des évaluations données par l'utilisateur x.

S'agissant de l'approche mémoire basée sur les items, la prédiction de la note de l'utilisateur u pour le service s est calculée de manière similaire à précédemment mais en prenant en considération les évaluations données par u, pondérées par la similarité des services évaluée avec le service cible, noté s [122].

6.2.5 Mesures de similarité

L'une des mesures de similarité les plus populaires est le coefficient de corrélation de Pearson (PCC) défini par :

$$Sim(u,v) = \frac{\sum\limits_{i \in I_{u,v}} (r_{ui} - \overline{r}_u)(r_{vi} - \overline{r}_v)}{\sqrt{\sum\limits_{i \in I_{u,v}} (r_{ui} - \overline{r}_u)^2} \sqrt{\sum\limits_{i \in I_{u,v}} (r_{v,i} - \overline{r}_v)^2}} \tag{6.3}$$

Ce coefficient de similarité est populaire car il permet d'obtenir une précision élevée dans des conditions favorables et peut être facilement mis en œuvre. Une autre mesure

de similarité est la similarité de cosinus (COS) entre les vecteurs $\vec{A}$ et $\vec{B}$, composés des évaluations données respectivement pour les services/items i et j :

$$Sim(i,j) = \frac{\vec{A} \cdot \vec{B}}{\| \vec{A} \|^2 \cdot \| \vec{B} \|^2} \tag{6.4}$$

Et sa variante, le cosinus ajusté, qui est recommandée dans le filtrage collaboratif basé sur les items [122; 123]. Le cosinus ajusté est légèrement différent de PCC et est donné par :

$$Sim(i,j) = \frac{\sum\limits_{u \in U_{i,j}} (r_{ui} - \overline{r}_u)(r_{uj} - \overline{r}_u)}{\sqrt{\sum\limits_{u \in U_{i,j}} (r_{ui} - \overline{r}_u)^2} \sqrt{\sum\limits_{u \in U_{i,j}} (r_{uj} - \overline{r}_u)^2}} \tag{6.5}$$

Certains problèmes rencontrés dans les algorithmes de filtrage collaboratif sont liés à l'espace de données [48]. Comme le calcul de similarité utilise la mesure statistique, la précision de la mesure de similarité est sensible à la qualité et la quantité des données disponibles et peut conduire à la similarité surestimée [59]. Le deuxième problème principal est lié aux considérations de passage à l'échelle et de calcul, dans le cas d'un système comptant un nombre important d'utilisateurs ou d'éléments [124].

6.3 Les cartes auto-organisatrices

Les cartes auto-organisatrices, ou par abréviation SOM pour *Self-organizing map* en anglais, également appelées cartes de Kohonen, ont été introduites par Teuvo Kohonen[1] en 1980, comme un type de réseau de neurones artificiels. L'algorithme de Kohonen utilise un apprentissage non supervisé et compétitif pour produire une structure facile à visualiser (généralement deux dimensions) appelée carte. Cette carte est un ensemble de M nœuds ou neurones, identifiés par un ID d'index compris entre 1 et M et chaque nœud est la représentation d'un ensemble de données d'entraînement, produites en entrée.

1. Chercheur finlandais, né le 11 juillet 1934, connu pour ses travaux sur les réseaux neuronaux artificiels et sur la quantification vectorielle.

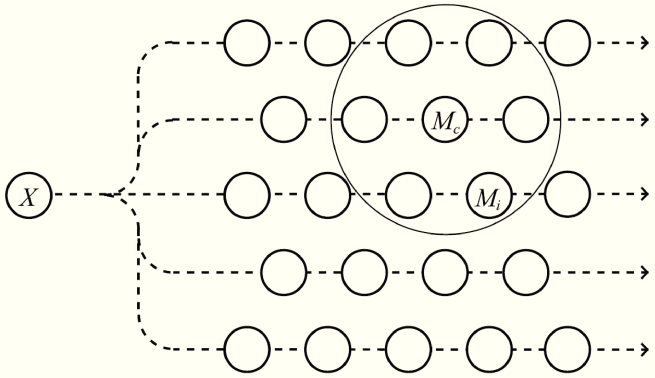


FIGURE 6.1 – Illustration d'une carte auto-organisatrice : Les données d'entrée X sont diffusées vers un ensemble de représentations M_i. La représentation M_c correspond le mieux à la donnée d'entrée X [69]

6.3.1 Algorithme des cartes auto-organisatrices

Soit X le nombre d'observations dans un espace à n dimensions $\mathbb{R}^n$, l'algorithme à Nt itérations de Kohonen se déroule selon les étapes suivantes :

- Initialiser aléatoirement (ou suivant une autre méthode plus efficace) les $m_i \in M$ représentants ou neurones dans l'espace $\mathbb{R}^n$.
- à chaque itération t, pour $t = 1$ à Nt :
  - Choisir au hasard une observation $x(t+1)$ (ou selon les variantes, suivant une loi probabiliste).
  - Mettre en compétition les neurones m_i et on détermine le gagnant m_c de façon à minimiser la distance :

$$|x(t+1) - m_c| = \underset{1 \leq i \leq M}{\operatorname{argmin}} |x(t+1) - m_i|$$

  - Sélectionner l'ensemble des voisins m_j de m_c (suivant la définition du voisinage à l'itération t)
  - Mettre à jour le code du neurone gagnant et de celui de ses voisins :

$$m_j(t+1) = m_j(t) + \epsilon(t)(x(t+1) - m_j)$$

    Où $\epsilon(t)$ est un paramètre constant ou décroissant vers 0

Intuitivement, comme le montre la figure illustratrice 6.2, cet algorithme permet d'avoir une représentation généralement bidimensionnelle discrète des données d'apprentissage, avec la particularité de sauvegarder la proche similarité entre neurones voisins.

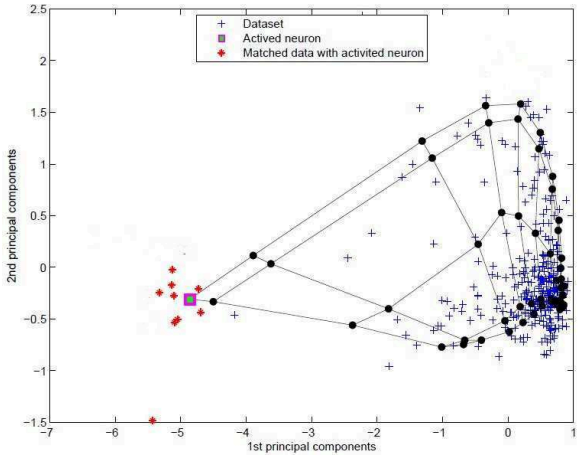


FIGURE 6.2 – Exemple pratique d'une carte auto-organisatrice

Enfin, il est à signaler que les performances de cet algorithme dépendent de l'initialisation des paramètres initiaux Mao *et al.* [89], et est sujet au problème du minimum local.

6.3.2 Avantage de la classification à base de cartes auto-organisatrices

Les cartes auto-organisatrice (SOM) présentent certains avantages que nous verrons dans le chapitre suivant et qui serviront à justifier notre choix pour cette technique :

— C'est un modèle à apprentissage non supervisé qui permet de classer l'ensemble des vecteurs dans différents clusters en fonction de leur similarité [87] et ne nécessite pas de données étiquetées.

— SOM est également une méthode qui permet de projeter un espace de données de grande dimension sur un espace plus petit et plus dense. Ce nouvel espace permettra de mettre en évidence de nouvelles relations entre individus, qui n'étaient pas évidentes dans l'espace initial.

— SOM exploite la notion de voisinage entre classes dans son apprentissage. Cette caractéristique préserve la structure topologique de l'espace d'entrée, de sorte que des représentations similaires seront associées à des nœuds plus proches dans la grille, comme l'illustre la figure 6.1 de [69].

— SOM permet de faire une analyse sur des données comprenant des relations non linéaires [90].

— Enfin, SOM s'accommode assez bien des valeurs manquantes [75].

6.4 K-moyennes

L'algorithme des K-moyennes ou K-means proposé par Macqueen [85] est un algorithme de regroupement largement utilisé dans la recherche scientifique (e.x. analyse de données, le traitement du signal et divers domaines de la production industrielle). Contrairement au clustering par carte de Kohonen, K-means nécessite d'initialiser le nombre de clusters. L'algorithme K-means vise à partitionner les vecteurs d'observations donnés en k groupes, de manière à minimiser la distance euclidienne au sein du même groupe. Dans la figure 6.3, nous pouvons voir un exemple de K-means sur user-profile (détaillé sur 7.2.2.5) appliqué au jeu de données et qui est utilisé dans nos expériences. Dans cet exemple, le paramètre $k = 4$.

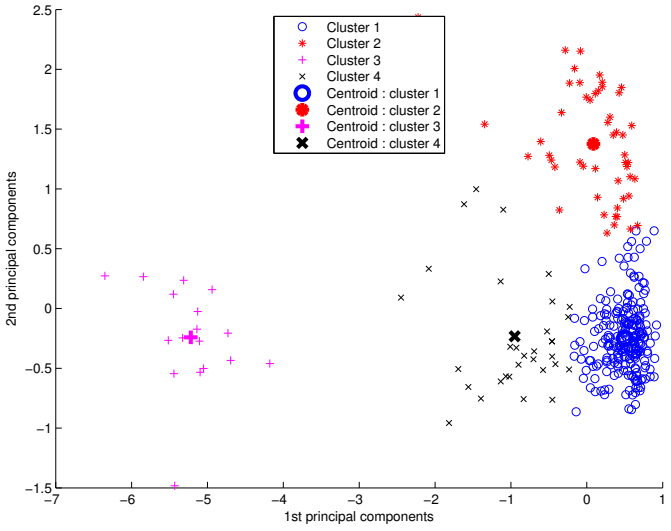


FIGURE 6.3 – Exemple K-means appliqué à un jeu de données

6.4.1 Algorithme des K-moyenne

Soit X le nombre d'observations dans un espace à n dimensions $\mathbb{R}^n$, l'algorithme des K-moyenne ou K-means se déroule suivant les étapes suivantes :

— Initialiser aléatoirement (ou suivant une autre méthode) K centroïdes dans l'espace $\mathbb{R}^n$.

— Tant que la condition d'arrêt n'est pas atteinte faire :

  — Assigner chaque vecteur d'entrée x à une centroïde k_i, de façon à minimiser la distance :

$$|x - k_i| = \underset{1 \leq i \leq K}{\operatorname{argmin}} |x - k_i|$$

  — Mettre à jour chaque centroïde en calculant la moyenne des vecteurs qui lui sont assignés :

$$k_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j$$

  Où S_i l'ensemble des vecteurs assignés à la centroïde k_i.

  La condition d'arrêt peut être un nombre d'itérations ou la stabilité du mouvement des centroïdes.

Enfin, la variante de regroupement de K-means proposée par Lloyd [82] est très populaire pour sa vitesse observée et sa simplicité de mise en œuvre [8]. Sa complexité est linéaire dans le temps et dans l'espace [73].

6.5 Conclusion

Dans ce chapitre, nous avons décrit la technique du filtrage collaboratif dans un contexte d'évaluation de la confiance d'un fournisseur de service. Puis nous avons abordé sommairement les deux algorithmes des cartes de Kohonen et des K-moyennes qui serons utilisés pour construire notre modèle de recommandation de confiance.

Chapitre 7

Modèle SOM-BTR d'évaluation de la confiance à l'aide de cartes auto-organisatrices

7.1 Introduction

Dans ce chapitre, nous allons décrire notre modèle de recommandation de confiance basé sur les cartes auto-organisatrices, noté SOM-BTR, pour *SOM Based Trust Recommendation model*. Notre approche est axée sur des utilisateurs similaires évaluant des services similaires, plutôt que de cibler un utilisateur sélectionné pour un service sélectionné. L'intuition derrière cette idée est que les utilisateurs similaires devraient se comporter de la même manière avec des services similaires. Nous traiterons également à travers ce modèle le problème du passage à l'échelle. En effet, comme relevé dans le chapitre précédent, les modèles du filtrage collaboratif traditionnels nécessitent de calculer la similarité entre chaque paire d'entités. Nous réduirons les coûts de calcul en définissant des modèles hors ligne pour les utilisateurs et les services à l'aide de l'algorithme des cartes auto-organisées et une description de la tendance des données basée sur les K-moyennes. Pour plus de clarté, nous allons utiliser dans ce chapitre les même abréviations que le chapitre précédent, données

dans le Tableau 7.1, que nous complétons par le tableau 7.1.

TABLE 7.1 – Liste des abréviations

Abbreviations	**Definitions**
N	Ensemble de données
C^i	$i^{ème}$ classe de l'ensemble N
$\|X\|$	Cardinalité de l'ensemble X (nombre d'éléments de X)
$\text{argmax}_{1\leq x\leq k}\, f(x)$	Retourne le point x du domaine de définition sur lequel la fonction f est maximale
$\text{argmin}_{1\leq x\leq k}\, f(x)$	Retourne le point x du domaine de définition sur lequel la fonction f est minimal
R	Matrice des évaluation utilisateurs - services
R_u^x	Matrice des évaluations de l'utilisateur u données à un ensemble x de services similaires

7.2 Modèle de recommandation de confiance basé sur les cartes auto-organisatrices (SOM-BTR)

Dans un scénario typique, comme illustré dans la figure 7.1, un utilisateur exprimant un besoin pour un service d'externalisation informatique, va émettre une requête pour le moteur de recommandation. Ce dernier, sur la base du profil du demandeur et des profils des services disponibles, va recommander à l'utilisateur le service le plus approprié. Après consommation du service, l'utilisateur va fournir une évaluation, qui sera stockée dans la base des retours d'expériences. Cette évaluation servira pour la mise à jour future des connaissances du module de recommandation.

Le schéma proposé comporte trois phases principales. La première phase est une opération hors ligne. Dans cette phase, les modules SOM construisent deux modèles à base de profils et de cartes de Kohonen, l'un pour les utilisateurs (SOM_U) et l'autre pour les services (SOM_S). La deuxième phase est une phase en ligne, elle commence quand un utilisateur émet une demande de recommandation pour un service disponible et de confiance. La dernière phase est également une phase hors ligne, elle est dédiée à la mise à jour des modèles construits lors de la première phase. Elle prend en compte les données récentes recueillies par le système sur les utilisateurs et les services.

Comme nous pouvons le voir sur la figure 7.1, notre modèle de confiance comprend

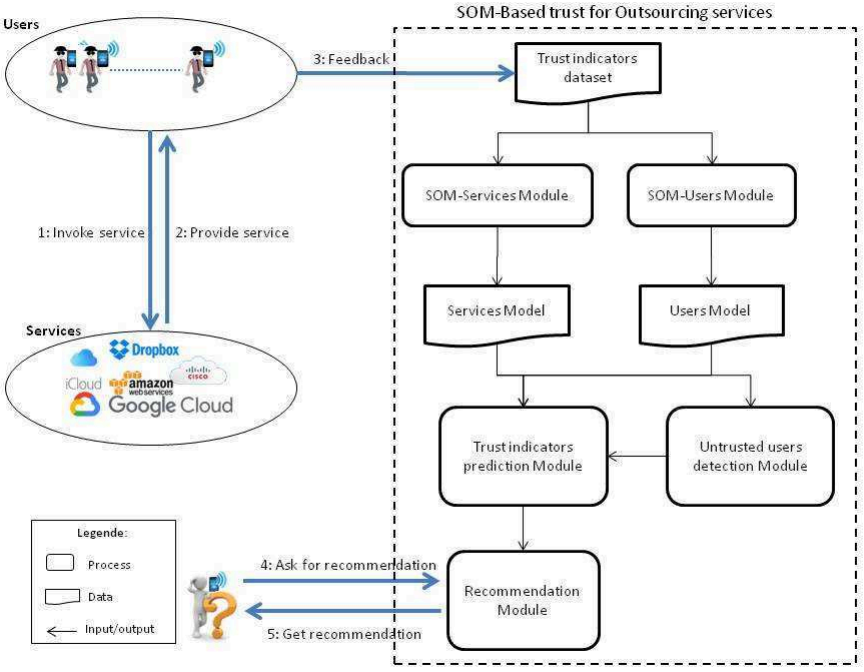


FIGURE 7.1 – Schéma fonctionnel du modèle SOM-BTR

cinq modules :

— **Module SOM-Users**, ce module génère le modèle de cartes auto-organisées qui réduit la masse des données des utilisateurs, en les modélisant à l'aide d'une carte de profils, avec un nombre limité de neurones (nœuds).

— **Module SOM-Services**, comparable au module précédent, il se charge de générer un modèle pour les services à base de profils représenté par une carte de Kohonen.

— **Module de détection des utilisateurs non fiables ("Untrusted users detection module" en anglais)**, utilise les deux modèles générés, précédemment, pour identifier les utilisateurs dont les évaluations sont non fiables.

— **Module de prédiction des indicateurs de confiance (Trust-indicator prediction module)** , qui se charge de calculer les valeurs prédites des indicateurs de confiance comme le temps de réponse ou la bande passante.

— **Module de recommandation**, qui est l'interface avec les utilisateurs, il produit des recommandations en fonction de la demande des utilisateurs et des résultats du

processus de prédiction.

7.2.1 Moyenne Descriptive

Dans une distribution de données avec un écart-type élevé, le calcul de la moyenne ne donne pas une image juste de la réalité des données. Par exemple, la séquence suivante $N = \{1.8, 1.2, 1.7, 2, 1, 7, 47, 75, 7, 8, 7, 8\}$ donnera une moyenne égale à 14.72 alors qu'il est évident que la valeur la plus courante est de 1.70.

Pour cette raison, nous introduisons la moyenne descriptive basée sur la classification des K-moyennes. Par exemple, pour une collection donnée de q valeurs $N = \{r_1, n_2, ..., n_q\}$, nous utilisons la méthode des K-moyennes pour regrouper ces notes en classes. Les résultats sont un ensemble de classes C_N^i avec $0 \leq i \leq k$, classés du plus grand au plus petit, $|C_N^1| \geqslant |C_N^2| \geqslant ... \geqslant |C_N^k|$.

Définition 19 (Moyenne Descriptive)
La Moyenne-Descriptive est définie comme suit : Soit l'ensemble des sous-ensembles ξ de N tel que

$$\xi = \bigcup_{i=1}^{j} C_N^i \quad \text{avec} \quad j = \underset{1 \leq m \leq k}{\operatorname{argmin}} \sum_{m=1}^{j} \frac{|C_N^m|}{|N|} \geqslant \alpha, \quad 0 < \alpha \leqslant 1$$

Alors, la Moyenne-Descriptive est donnée par :

$$DAv_k^\alpha(N) = \frac{\sum_{n_i \in \xi} n_i}{|\xi|} \tag{7.1}$$

En appliquant la formule 7.1 à l'exemple précédent, avec $k = 3$, nous trouvons $DAv_k^\alpha(N) = 1.70$ pour $\alpha = 0.6$ et $DAv_k^\alpha(N) = 3.15$ pour $\alpha = 0.8$. Dans ce résultat, nous pouvons voir que la valeur la plus appropriée pour décrire le jeu de données est 1.70, ce qui correspond à $\alpha = 0.6$ et $k = 3$.

Dans la suite de ce travail, nous utilisons notre calcul DAv_k^α au lieu du calcul de la moyenne traditionnelle. Les paramètres k et α seront initialisés de façon empirique en fonction de la taille et de la distribution de l'ensemble N.

7.2.2 Recommandation basée sur les cartes de Kohonen

Les trois phases du modèle de recommandation proposé sont détaillées ci-dessous :

7.2.2.1 Phase 1 : Construction du modèle

Cette phase se déroule hors-ligne. Elle vise à construire deux modèles décrivant les utilisateurs et les services et permettant de déduire les similarités sans effectuer de calculs intensifs en ligne. Supposons que nous avons m utilisateurs et n services. Idéalement, nous avons une matrice notée R de dimension $n \times m$, où chaque entrée r_{uv} représente la valeur de l'indicateur enregistrée par l'utilisateur u pour le service v. Les modules SOM pour les utilisateurs et pour les services sont utilisés pour construire les modèles d'utilisateurs et de services SOM à partir de la base des profils. Enfin, les deux modèles SOM sont sauvegardés en mémoire persistante.

Les profils de services et d'utilisateurs utilisés pour créer les modèles SOM sont détaillés dans 7.2.2.4 et 7.2.2.5.

7.2.2.2 Phase 2 : Recommandation de confiance

Cette phase commence lorsqu'un utilisateur donné est confronté à un ensemble de services candidats disponibles S et souhaite sélectionner celui qui lui conviendrait le mieux.

L'utilisateur envoie une demande de recommandation au module de recommandation. Ce dernier calcule le score ou l'évaluation de chaque service candidat en fonction des modèles SOM préalablement calculés et recommande le service ayant la meilleure appréciation estimée.

Après consommation du service, l'utilisateur enregistre un retour sous forme d'évaluation. Ce retour sera collecté par le générateur de modèle pour être utilisé lors de la prochaine phase de mise à jour des modèles SOM.

Concrètement, la recommandation de confiance est effectuée en prédisant les mesures de l'indicateur de confiance r_{uv} (telles que le temps de réponse) que l'utilisateur sélectionné u aurait données pour le service candidat v. Dans notre approche, pour prédire cet indicateur de confiance, nous sélectionnons l'ensemble des utilisateurs $\mathcal{U}_u$ similaires à l'utilisateur

sélectionné, après filtrage des utilisateurs malveillants, et nous nous concentrons sur les valeurs que ces utilisateurs ont attribuées à l'ensemble des services $\mathcal{S}_v$ semblables au service candidat.

Soit R la matrice des évaluations des services par les utilisateurs, U l'ensemble des utilisateurs et S l'ensemble des services. Pour un utilisateur donné u et un service donné v, les ensembles des utilisateurs et des services similaires respectivement à u et v sont notés $\mathcal{U}_u$ et $\mathcal{S}_v$.

Définition 20 (Meilleur-neurone-correspondant BMN)
Nous définissons la fonction meilleur-neurone-correspondant, notée $BMN(i) \xrightarrow{n}$ *pour Best-Matching-Neurone, qui renvoie l'index* n *de nœud de la carte (à* M *neurones) qui correspond le mieux à un vecteur donné* i*. C'est-à-dire le nœud qui représente le mieux le vecteur d'entrée* i*. Formellement,* n *doit minimiser la distance avec* i *:* $|i-n| = \text{argmin}_{1\leq n\leq M} |i-n|$ *(voir également l'algorithme 6.3.1)*

L'ensemble des services similaires est donné par :

$$\mathcal{S}_m = \{s \in \mathcal{S} \quad | \quad BMN(s) = m\} \tag{7.2}$$

Aussi, $\mathcal{U}_u$ et $\mathcal{S}_v$ sont définis comme suit : $\mathcal{U}_u = \{i \in U \mid BMN(i) = BMN(u)\}$ et $\mathcal{S}_v = \{j \in S \mid BMN(j) = BMN(v)\}$.

Soit R_{uv} l'ensemble des évaluations données par les utilisateurs $\mathcal{U}_u$ aux services dans $\mathcal{S}_v$, comme illustré dans la Figure 7.2.

$$R_{uv} = \{r_{xy} \in R \mid x \in \mathcal{U}_u \wedge y \in \mathcal{S}_v\}$$

La valeur prédite r'_{ij} est calculée à l'aide de la *Moyenne-Descriptive* (voir 7.2.1), en initialisant les paramètres α et K, puis nous appliquons l'algorithme des K-moyennes sur R_{ij}.

Nous obtenons ainsi K classes, notées $C^k_{R_{ij}}$ avec $0 \leq k \leq K$, que nous classons du plus grand au plus petit, $|C^1_{R_i}| \geqslant |C^2_{R_i}| \geqslant ... \geqslant |C^K_{R_{ij}}|$. Enfin, la valeur prédite est calculée comme suit :

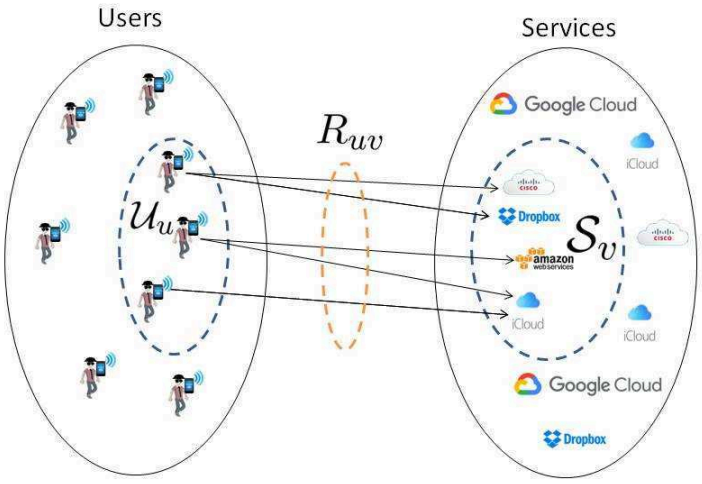


FIGURE 7.2 – Ensemble des évaluations R_{uv} d'un groupe d'utilisateurs similaires pour un groupe de services similaires

$$r'_{ij} = DAv_k^{\alpha}(R_{ij}) = \frac{\sum_{r_{xy} \in \xi} r_{xy}}{|\xi|} \tag{7.3}$$

$$\text{Avec} \quad \xi = \bigcup_{k=1}^{j} C_{R_{ij}}^{k} \quad \text{et} \quad j = \underset{1 \leq m \leq K}{\operatorname{argmin}} \sum_{m=1}^{j} \frac{|C_{R_{ij}}^{m}|}{|R_{ij}|} \geqslant \alpha \quad , \quad 0 < \alpha \leqslant 1.$$

K et α sont initialisés selon la distribution de R_{uv}. Dans la partie expérimentation (Chapitre 8), nous utilisons $\alpha = 0.7$ et K est sélectionné suivant la taille de l'ensemble R_{uv}.

7.2.2.3 Phase finale : Mise à jour des modèles SOM

La confiance est une valeur dynamique car le comportement des fournisseurs de services et les commentaires des utilisateurs peuvent changer au fil du temps. Nous définissons deux fenêtres, l'une pour les retours des utilisateurs ΔR et l'autre temporelle ΔT, pour mettre à jour les modèles SOM des utilisateurs et des services. Ces fenêtres ΔR et ΔT sont affectés par l'expert en fonction de sa perception du changement. La phase de mise à jour est détaillée dans l'algorithme 1.

Algorithm 1 Description du haut niveau de l'algorithme de mise à jour

Input : ΔT and ΔR
Output : Updated model
Initialization : $t = 0$ time initialization, $f = 0$ number of feedback, M numbers of neurons for services SOM-model
while $t < \Delta T$ and $f < \Delta R$ **do**
  if n new feedbacks are reported **then**
    $f = f + n$
  end if
  increment(t)
end while
// execute phase one :
for all $s \in S$ **do**
  perform expressions (7.4) to (7.7)
end for
SOM update for services
for $1 < i < M$ **do**
  perform expressions (7.9) and (7.10)
end for
SOM update for users
return SOM_U and SOM_S

7.2.2.4 Profil de Service

Nous utilisons l'ensemble des évaluations des utilisateurs pour le service s, noté $R_s = \{r_{is} \in R, 1 \leq i \leq n\}$ pour obtenir le profil du service s. Dans ce travail, nous adoptons un profil de service basé sur la moyenne descriptive définie précédemment. Toutes les évaluations relatives à chaque service sont classées avec l'algorithme des K-moyennes avec $k = 3$. Le modèle de profil de service comprend un terme quadruple $P_s = \langle Mc_s^1, Mc_s^2, Ec_s^1, Ec_s^2 \rangle$, avec Mc_s^1, Mc_s^2 respectivement la moyenne descriptive des première et deuxième classes, ayant le plus grand nombre d'effectifs, définis comme suit :

$C_s^1 = \{r_i \in R_s \mid r_i \in C_s^j, \quad j = \text{argmax}_{1 \leq i \leq 3} |C_s^i|\}$ et
$C_s^2 = \{r_i \in R_s \mid r_i \notin C_s^1, \, r_i \notin C_s^j, \, j = \text{argmin}_{1 \leq i \leq 3} |C_s^i|\}$.

Mc_s^1 et Mc_s^2 sont données par :

$$Mc_s^1 = DAv_k^{\alpha}(C_s^1) \tag{7.4}$$

et

$$Mc_s^2 = DAv_k^{\alpha}(C_s^2) \tag{7.5}$$

Et Ec_s^1, Ec_s^2 sont respectivement, les tailles relatives des classes C_s^1 et C_s^2 par rapport à la taille de l'ensemble des données R_s :

$$Ec_s^1 = \frac{|C_s^1|}{|R_s|} \tag{7.6}$$

$$Ec_s^2 = \frac{|C_s^2|}{|R_s|} \tag{7.7}$$

L'entrée du module SOM-Service est l'ensemble des profils précédemment définis. La sortie est le modèle SOM_S. Ce modèle est une carte avec M nœuds (voir section 6.3). Chaque nœud représente un groupe de services similaires $\mathcal{S}_m, 1 \leq m \leq M$. Nous utilisons la fonction best-matching-node ($BMN(i)$), voir définition 20 pour caractériser l'ensemble des services similaires :

$$\mathcal{S}_m = \{s \in \mathcal{S} \quad | \quad BMN(s) = m\} \tag{7.8}$$

7.2.2.5 Profil des utilisateurs

Dans le modèle proposé, nous captons le comportement des utilisateurs vis-à-vis de chaque groupe de services similaires via deux paramètres, la moyenne descriptive et la fréquence descriptive.

Supposons que nous ayons M groupes de services, chaque groupe est noté S_s avec $1 \leq s \leq M$, tel que défini précédemment. Le profil d'un utilisateur sera défini comme une suite de couples composés chacun d'une moyenne descriptive et d'une fréquence $P_u = \langle M_u^1, F_u^1, M_u^2, F_u^2, ..., M_u^M, F_u^M \rangle$

La moyenne descriptive est calculée pour chaque groupe s de services comme suit :

$$M_u^s = DAv_k^{\alpha}(R_u^s), \qquad 1 \leq s \leq M \tag{7.9}$$

Où $R_u^s = \{r_{uj} \in R \mid j \in \mathcal{S}_s\}, \qquad 1 \leq s \leq M$

Et la fréquence relative est définie comme suit :

$$F_u^s = \begin{cases} \frac{|R_{uj'}|}{|R_{uj}|} & \text{Si} \quad \exists j \in S, r_{uj} \neq NULL \\ \\ 0 & \text{Sinon} \end{cases} \tag{7.10}$$

Où $R_{uj'}$ et R_{uj} deux ensembles de valeurs extraites de la matrice de notations R comme suit : $R_{uj'} = \{r_{uj'} \in R | j' \in S_s\}$ et $R_{uj} = \{r_{uj} \in R | 1 \leq j \leq m\}$.

Rappelons que R est une matrice d'ordre $n \times m$, où m est le nombre de services et r_{ij} est l'évaluation ou la mesure de l'utilisateur u donnée pour le service v. Le générateur SOM pour les utilisateurs prend en entrée l'ensemble des profils des utilisateurs et crée une carte SOM_U avec N neurones. Chaque neurone représente un groupe d'utilisateurs similaires $\mathcal{U}_n$, $1 \leq n \leq N$.

7.3 Identification des utilisateurs non fiables

Certains services peuvent recevoir des notations incorrectes ou injustes pour diverses raisons, généralement malveillantes. Dans ce travail, nous supposons que les utilisateurs honnêtes constituent toujours la majeure partie d'un groupe d'utilisateurs. Cette hypothèse est également énoncée par Li *et al.* [77]. Contrairement aux solutions proposées dans la littérature, telles que Whitby *et al.* [145], qui font appel à des calculs probabilistes pour supprimer les notations injustes ou incohérentes, nous fournissons un moyen de détecter les utilisateurs malveillants à travers le calcul simple d'un coefficient de crédibilité. Nous définissons trois paramètres, la densité d'invocation de l'utilisateur D, le coefficient d'aberration de l'utilisateur Abr et le coefficient de crédibilité de l'utilisateur CrC.

7.3.1 Densité d'appels de l'utilisateur $D_{u,s}$

La densité d'appels de l'utilisateur, notée $D_{u,s}$ donne une indication sur l'intensité avec laquelle un utilisateur donné u invoque un groupe de services similaires S_s (noté par commodité s). L'utilisation moyenne du groupe de services ciblés par l'utilisateur u est comparée à l'utilisation moyenne de la majorité des utilisateurs pour ce groupe de services S_s. Plus la valeur est proche de un, plus l'utilisateur invoque normalement ce groupe de services.

$$D_{u,s} = \frac{n_{us} + 1}{DAv_k^{\alpha}(N_s) + 1} \tag{7.11}$$

Où n_{us} est le nombre de fois que l'utilisateur u invoque des services du groupe S_s. $DAv_k^{\alpha}(N_s)$ est calculé à partir de l'équation 7.1. N_s est la distribution d'utilisation du groupe de services S_s.

Pour éclairer le lecteur, nous allons illustrer ce concept par un exemple. Supposons qu'un groupe donné de services similaires S_s a été utilisé suivant la distribution suivante : $N_s = \{5, 12, 2, 4, 3, 4\}$ par respectivement les utilisateurs u_1, u_2, u_3, u_4, u_5 et u_6. En d'autres termes, l'utilisateur u_1 à fait appel à des services du groupe S_s 5 fois et u_2 12 fois, et ainsi de suite.

La densité d'appels (en appliquant la formule 7.11) donne : $D_{2,s} = \frac{13}{4} = 3.25$, tandis que $D_{1,s} = \frac{6}{4} = 1.5$. Cela signifie que l'utilisateur u_2 appelle ce groupe de services de manière anormalement élevée par rapport à la majorité des utilisateurs. Ce qui peut s'expliquer par une préférence personnelle ou par une tentative d'attaque. Quant à l'utilisateur u_1, sa densité d'utilisation est proche de 1. Ce qui signifie que cet utilisateur se comporte normalement avec ce groupe de services (en termes d'utilisation).

7.3.2 Le coefficient d'aberration de l'utilisateur $Abr_{u,s}$

Le coefficient d'aberration de l'utilisateur, noté $Abr_{u,v}$, quantifie la déviation des feedbacks d'un utilisateur donné u par rapport à la majorité du groupe de services similaires S_s.

Premièrement, nous utilisons les K-moyennes pour classer l'ensemble des évaluations R_{us} données par cet utilisateur aux services d'un groupe sélectionné S_s. k est fixé empiriquement en fonction de la taille de R_{us}. Ensuite, nous calculons les distances euclidiennes (nous pouvons utiliser d'autres mesures) $Dc_{i,s}$ entre chaque centre de gravité de la classe i et la valeur de Mc_s^1 vue dans 7.2.2.4. Enfin, nous calculons le coefficient d'aberration $Abr_{u,s}$ sous la forme d'une somme pondérée de ces distances. La pondération est donnée pour chaque distance euclidienne par l'effectif de la classe correspondante. Formellement, Abr est défini comme suit :

$$Abr_{u,s} = \sum_{i=1}^{k} |C_i| \times Dc_{i,s} \tag{7.12}$$

C_i est la $i^{ème}$ classe de l'ensemble des évaluations R_{us}.

7.3.3 Coefficient de crédibilité de l'utilisateur $CrC_{u,s}$

Le coefficient de crédibilité de l'utilisateur u pour un groupe de service similaire S_s, noté $CrC_{u,s}$, est donné par :

$$CrC_{u,s} = Abr_{u,s} \times e^{D_{u,s}} \tag{7.13}$$

La formule 7.13 est motivée par le fait qu'un utilisateur ne peut être considéré comme malveillant, si ses évaluations sont occasionnellement injustes. Dans ce cas, le $CrC_{u,s}$ faible, même si $Abr_{u,s}$ est élevé car $D_{u,s}$ est petit.

Cependant, si un utilisateur donne un nombre important d'évaluations injustes à un groupe de services et qu'il s'acharne anormalement à invoquer ce groupe de services (valeur $D_{u,s}$ élevée), alors $CrC_{u,s}$ sera très élevé. La valeur de $CrC_{u,s}$ va permettre de détecter une attaque localisée sur le groupe de services similaires S_s.

Un CrC_u global pour l'utilisateur u est calculé avec la *Moyenne-Descriptive* (formule 7.1), appliquée à tous les $CrC_{u,s}$ de l'utilisateur u. Les utilisateurs non fiables qui ont un comportement malveillant sont détectés en utilisant un seuil noté γ. Si le coefficient de crédibilité CrC_u calculé pour l'utilisateur y dépasse le seuil γ, cet utilisateur est identifié comme non fiable. Ainsi, l'ensemble des utilisateurs non fiables UN est défini avec l'expression suivante :

$$UN = \{u \in U \quad | \quad CrC_u > \gamma\} \tag{7.14}$$

7.4 Conclusion

Dans ce chapitre, nous avons présenté notre modèle d'évaluation de la confiance basé sur les cartes de Kohonen. Ce modèle est basé sur le filtrage collaboratif avec les cartes de Kohonen, qui, à notre connaissance n'a jamais été utilisé dans ce domaine. Notre solution

permet également de filtrer les utilisateurs ayant un comportement anormal, en termes d'évaluation des services.

Dans le chapitre suivant, nous allons montrer l'efficacité de notre modèle à travers une analyse expérimentale conduite sur des données issues du monde réel.

Chapitre 8

Évaluation expérimentale du modèle SOM-BTR

8.1 Introduction

Dans ce chapitre, nous avons réalisé des évaluations expérimentales de l'approche de gestion de la confiance proposée (SOM-BTR). Les expériences consistent principalement à mesurer la précision et l'efficacité des prévisions par rapport aux méthodes de recommandation traditionnelles et celles présentées récemment dans la littérature. Nous avons également mesuré l'efficacité de la détection des utilisateurs non fiables et l'impact qu'ils ont sur le système de recommandation en termes d'exactitude de la prédiction. Toutes les expériences sont réalisées sur un PC doté d'un système d'exploitation Windows 8.1 64 bits avec un processeur Intel Core i5 à 3,0 GHz et 8 Go de RAM. Nous avons utilisé l'environnement MATLAB avec la version R2007b. Les expériences sont réalisées avec le paramètre $\alpha = 0.7$ (pour prendre en compte une majorité de plus de 2/3 de la population), alors que le paramètre k est défini en fonction de la taille des données à traiter par la fonction DAv_k^{α} (plus la population est importante, plus k est grand).

8.2 Données utilisées

Nous avons réalisé les expériences sur un jeu de données du monde réel en libre accès (WS-DREAM) Zheng et Lyu [156]. Cette base de données contient les temps de réponse des services Web recueillis auprès de 339 utilisateurs (situés dans 30 pays) ayant invoqué

5825 services Web (situés dans 73 pays), soit une matrice d'évaluation, notée R, de taille 339×5825 (1 974 675 évaluations). Les valeurs de RT sont comprises entre 0 et 20 secondes.

Pour être plus réaliste, étant donné que la matrice des évaluations est très clairsemée dans le monde réel, nous retirons au hasard un nombre donné de la matrice R initiale, afin d'obtenir différentes densités de matrice (MD). Concrètement, par exemple pour une densité de matrice MD=30%, cela indique que seulement 30% des valeurs de la matrice R sont renseignées, soit 592 402 évaluations sur un total de 1 974 675 cases. Dans un deuxième temps, pour simuler les utilisateurs non fiables, nous sélectionnons aléatoirement un nombre d'utilisateurs et nous remplaçons leurs évaluations par des valeurs générées aléatoirement dans la même plage de données initiales (0 à 20).

8.3 Métrique d'évaluation

Nous adoptons l'erreur quadratique moyenne (RMSE) pour évaluer l'exactitude de la prévision, ainsi que le taux de faux positifs (FPR) et le taux de faux négatifs (FNR), afin d'évaluer l'efficacité de la détection des utilisateurs non fiables dans le système. Plus les valeurs de RMSE, FPR et FNR sont basses, plus la précision de la prédiction est élevée.

Le RMSE mesure la distance entre la valeur prédite et celle observée. C'est la métrique utilisée dans le très populaire prix Netflix [132]. RMSE est défini comme suit :

$$RMSE = \sqrt{\frac{\sum_{us} |r_{us} - r'_{us}|}{N}}$$

Où r_{us} est la valeur observée, r'_{us} est la valeur prédite et N est le nombre de valeurs prédites.

Le FPR est le pourcentage d'utilisateurs de confiance classés par le modèle en tant qu'utilisateurs non fiables. Le FNR est le pourcentage d'utilisateurs non fiables classés par le système en tant qu'utilisateurs fiables. Les deux métriques sont définies comme suit :

$$FPR = \frac{|DUu \cap Tu|}{|Tu|}$$

$$FNR = \frac{|DTu \cap Uu|}{|Uu|}$$

Où, DUu et DTu sont respectivement les ensembles d'utilisateurs non fiables et des utilisateurs fiables selon le modèle de détection. Tu et Uu sont les ensembles des utilisateurs respectivement fiables et non fiables, réellement présents dans le système.

8.4 Robustesse du modèle face aux utilisateurs non fiables

Dans cette expérience, nous montrons que le modèle de prédiction proposé est stable en présence d'un certain nombre d'utilisateurs non fiables dans le système. Nous fixons le nombre d'utilisateurs non fiables à 10, 20, 30, 40, 50 et 60 (respectivement 2,9 %, 5,9 %, 8,8 %, 11,8 %, 14,7 % et 17,7 % du total des utilisateurs dans la base de données), choisis à chaque fois de façon aléatoire. Les expériences sont réalisées avec différentes densités de matrice, 5%, 10%, 20% et 30% (initialisées à chaque fois de façon aléatoire). Les résultats sont représentés sur la figure 8.1.

On remarque que les valeurs de RMSE sont linéaires et stables. Cela prouve que l'approche proposée est efficace même en présence d'utilisateurs non fiables. Nous rappelons, à ce niveau, l'hypothèse selon laquelle le nombre d'utilisateurs non fiables ne doit pas gagner la majorité.

8.5 Impact du seuil γ sur la détection d'utilisateurs non fiables

Comme défini dans le chapitre précédent (section 7.3), γ est une valeur de seuil pouvant être initialisée pour identifier les utilisateurs non fiables, dont l'ensemble est noté UN. Dans cette expérience, nous étudions l'impact de la valeur de γ sur l'efficacité de l'identification des utilisateurs non fiables. Pour ce faire, nous initialisons la densité de la matrice (MD) avec différentes valeurs (5 %, 10 %, 20 % et 30 %), le nombre d'utilisateurs non fiables du système étant de 10, 20, 30, 40 utilisateurs, coisis aléatoirement, soit respectivement, 2.9%, 5.9%, 8.7% et 11.6% du total des utilisateurs, à chaque fois choisis de façon aléatoire. En faisant varier la valeur du seuil γ, nous calculons FPR et FNR. La figure 8.2 montre les

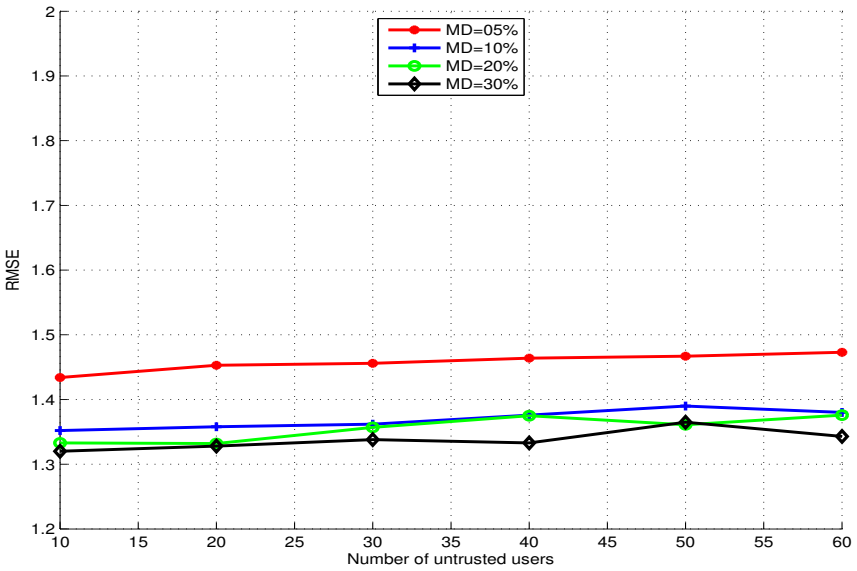


FIGURE 8.1 – Robustesse de modèle SOM-BTR en présence d'utilisateurs non fiables

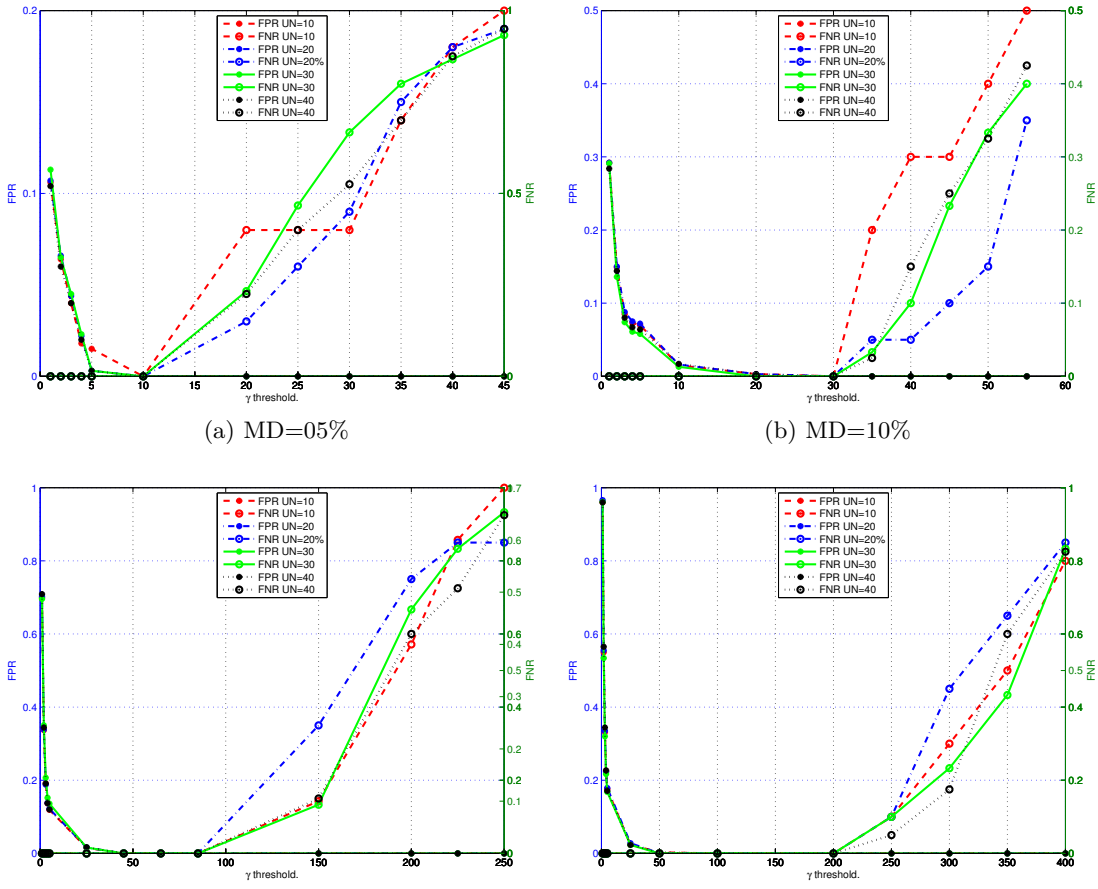
résultats de ces expériences.

Pour MD = 5%, la valeur optimale de γ est égale à 10 (figure 8.2a) alors que pour MD = 10%, la valeur optimale $\gamma \in [20, 30]$ (figure 8.2b). Avec MD = 20%, optimum $\gamma \in [45, 85]$ (figure 8.2c) et avec MD = 30%, la valeur optimale $\gamma \in [100, 200]$ (figure 8.2d).

On remarque que le spectre de l'optimum du seuil γ augmente lorsque la densité de la matrice est élevée (exactement 10 pour MD=5% à $\gamma \in [100, 200]$ pour MD = 30%). Ceci peut être expliqué par le fait que le modèle dispose de plus d'informations pour détecter les utilisateurs malhonnêtes et par conséquent, une plus grande latitude pour choisir un seuil optimal.

8.6 Détection des utilisateurs non fiables

Dans cette expérience, nous appliquons notre approche à la matrice R avec MD prenant les valeurs de 5%, 10%, 20% puis 30% et à chaque fois, nous introduisons 20 utilisateurs non fiables, choisis aléatoirement pour chaque expérience. Les valeurs de γ sont les seuils optimaux identifiés dans les expériences précédentes pour chaque valeur de MD (voir section 8.5).



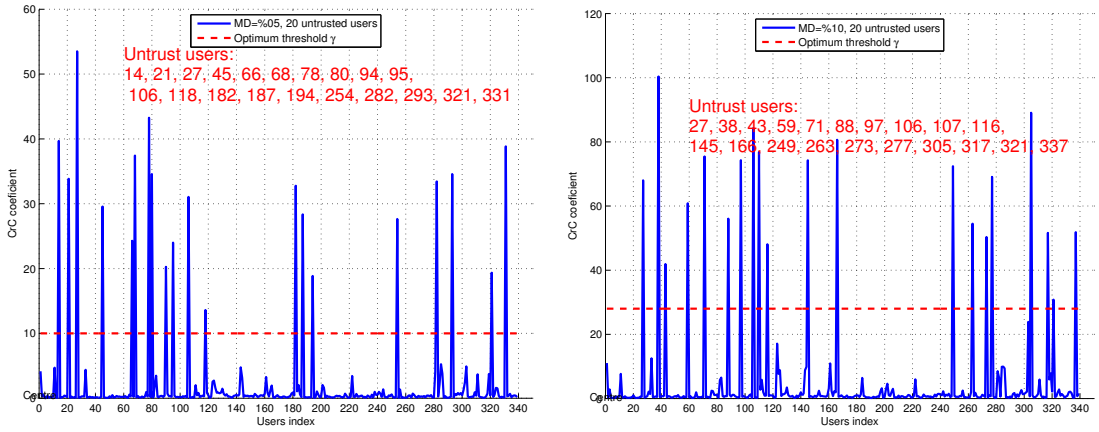
(a) MD=05%

(b) MD=10%

(c) MD=20%

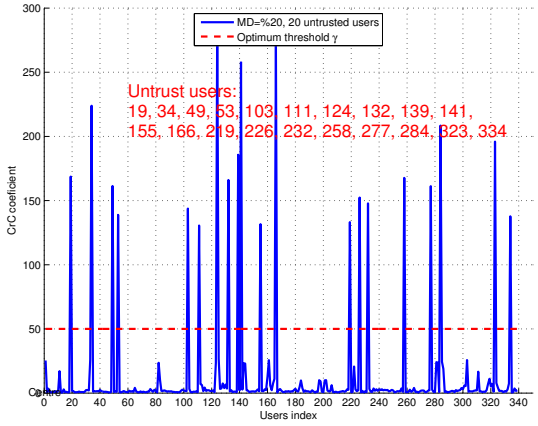
(d) MD=30%

FIGURE 8.2 – Impact de γ sur la détection d'utilisateurs non fiables

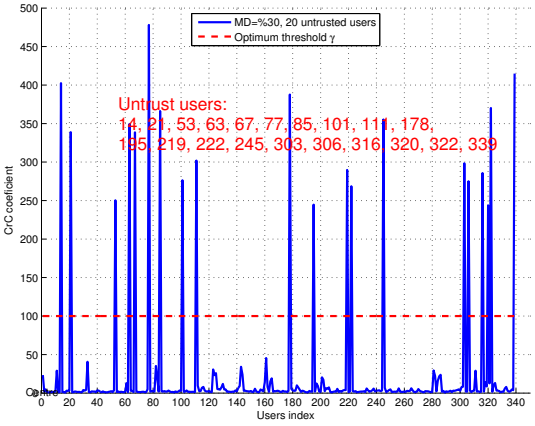


(a) MD=05%

(b) MD=10%



(c) MD=20%



(d) MD=30%

FIGURE 8.3 – Spectre de CrC pour l'ensemble des utilisateurs en présence de 20 utilisateurs non fiables, pour différentes valeurs de densité de matrice MD

La figure 8.3 montre que les utilisateurs non fiables injectés dans la matrice RT ont tous été bien détectés et qu'aucun faux positif n'est enregistré lorsque la valeur de seuil γ est bien choisie.

Par exemple, pour MD = 20%, nous initialisons $\gamma = 70$ (de la figure 8.2c on peut voir que l'optimum γ est compris entre $[50, 80]$ pour MD=20%). Dans ce cas, comme on peut le voir dans la figure 8.3c, tous les utilisateurs non fiables injectés sont détectés sans faux positifs.

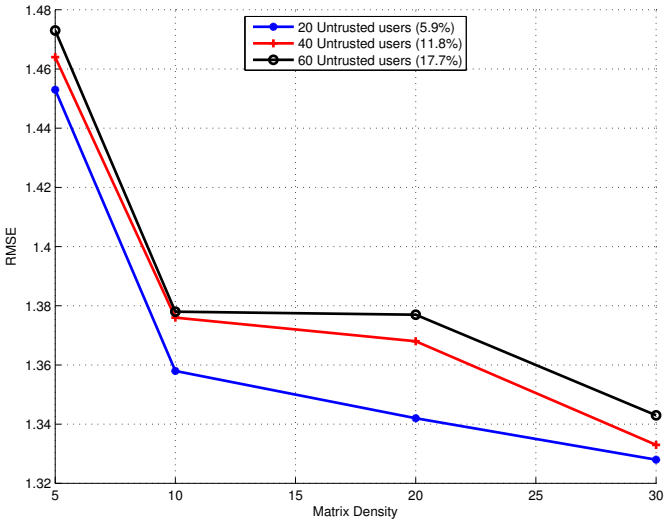


FIGURE 8.4 – Impact de la densité de matrice MD sur l'efficacité de la prédiction

8.7 Impact de la densité de données sur l'efficacité de la prédiction

Dans cette expérience, nous étudions l'effet de la densité de matricielle MD sur la prédiction des évaluations. Comme dans l'expérience précédente, la densité de la matrice est initialisée à 5 %, 10 %, 20 % et 30 %. Pour être réaliste, les utilisateurs non fiables sont initialisés à 5.9 %, 11.8 % et 17.7 %. Comme on peut le voir dans la figure 8.4, la valeur RMSE diminue lorsque la densité de la matrice augmente. Cependant, initialement, le RMSE diminue rapidement, puis lentement à mesure que le modèle converge vers la valeur de RMSE minimale. Plus le modèle dispose d'informations (valeur de MD grande) plus RMSE est faible, donc plus la prédiction est précise.

8.8 Analyse comparative de SOM-BTR avec les modèles de la littérature

Afin de démontrer la validité de notre proposition qui concerne la gestion de la confiance à base des cartes de Kohonen (SOM-BTR), nous comparons notre solution à une sélection des méthodes proposées dans la littérature (voir notre état de l'art en section 2.3.1). Nous avons sélectionné des méthodes traditionnelles du filtrage collaboratif (UPCC, IPCC et

WSRec) et des méthodes issues de travaux récents (GNMF, TAP et GURAP) :

1) *UPCC* [114] : Cette approche de filtrage collaboratif utilise la similarité entre deux utilisateurs calculée avec le coefficient de similarité de Pearson PCC.

2) *IPCC* [122] : Contrairement à la méthode précédente, cette approche du filtrage collaboratif utilise la similarité entre items en la calculant avec la technique du cosinus ajusté.

3) *WSRec* [155] : Cette technique combine les filtres de collaboration UPCC et IPCC pour calculer la prédiction.

4) *GNMF* [26] : Dans cette méthode une classification sur la base de voisinage géographique est combinée à la similarité PCC, puis intégrée dans un modèle de factorisation matricielle pour calculer la prédiction.

5) *TAP* [131] : Dans ce travail, la QoS est prédite en utilisant la similarité à base de service, après avoir identifié les utilisateurs fiables à l'aide de la classification avec les K-moyennes.

6) *GURAP* [72] : Dans cette approche, les auteurs utilisent l'emplacement géographique des utilisateurs et des services pour prévoir la qualité de service. Cette solution comprend un calcul de la réputation des utilisateurs.

Comme indiqué précédemment, pour être plus réaliste, nous effectuons une comparaison avec différents taux de densité de matrice MD (5%, 10%, 20% et 30%) et le nombre d'utilisateurs non fiables est initialisé à 10 utilisateurs sur un total de 339, choisis aléatoirement.

Pour une comparaison pertinente, les mêmes conditions initiales (matrice d'évaluation et utilisateurs nos fiables) sont utilisées pour chacun des modèles.

Le tableau 8.1 montre le RMSE, calculé pour l'approche proposée (SOM-BTR) et les méthodes citées plus haut. Nous pouvons remarquer que notre modèle présente des valeurs RMSE inférieures à celles des autres méthodes.

Cela démontre la précision de la méthode proposée basée sur la similarité utilisant les cartes de Kohonen couplée à un filtrage efficace des utilisateurs non fiables. Comparée au modèle de Kuang *et al.* [72], la méthode que nous proposons permet de meilleures

TABLE 8.1 – Résultats de l'étude comparative de la précision de la prédiction RMSE (le nombre des utilisateurs non fiables est initialisé à 10)

Methods	**MD=5%**	**MD=10%**	**MD=20%**	**MD=30%**
IPCC	2.305	2.183	2.149	2.079
UPCC	2.200	2.106	2.149	2.096
GNMF	2.275	2.078	1.703	1.655
WSRec	2.052	2.036	2.029	1.967
TAP	1.664	1.631	1.652	1.685
GURAP	1.494	1.433	1.332	1.284
SOM-BTR	**1.396**	**1.352**	**1.333**	**1.320**

performances avec une densité de matrice faible, MD = 5% et MD = 10% (qui constituent le cas d'utilisation extrême), tout en conservant les mêmes performances pour MD = 20%. Néanmoins, notre approche donne des résultats légèrement inférieurs pour une densité supérieure (MD = 30%). Cela est, probablement, dû au fait que les paramètres du modèle sont définis manuellement. Une amélioration du modèle proposé, liée à la sélection des paramètres de l'algorithme, est prévue dans les travaux futurs.

8.9 Cas d'utilisation illustratif

Afin d'illustrer la validité pratique de notre modèle de confiance, nous présentons le cas d'utilisation de surveillance de l'état de santé d'un individu, dans un scénario utilisant les objets connectés. Considérons un utilisateur avec des capteurs connectés (mouvements, pression artérielle...). Ces capteurs doivent transmettre les données collectées à un nœud intermédiaire (nœud de déportation de l'informatique en brouillard ou Fog-Computing), comme décrit dans la première partie de cette thèse, pour un pré-traitement sur le Cloud. Préalablement, l'utilisateur doit sélectionner un service de déportation. Dans cette optique, il demande au gestionnaire de confiance ou une autorité de confiance de sélectionner le fournisseur disponible le plus approprié. Supposons que les modèles SOM_S / SOM_U ont déjà été générés (phase une du modèle SOM-BTR). Supposons dans cet exemple qu'il existe trois fournisseurs disponibles. Pour chacun d'eux, le gestionnaire de confiance effectue une projection sur le modèle SOM_S, afin d'identifier les services similaires. La même tâche est

effectuée pour l'utilisateur avec le modèle SOM_U. Ensuite, le gestionnaire de confiance filtre les utilisateurs non fiables dans le même groupe d'utilisateurs (formules 7.11 à 7.13). Ensuite, le gestionnaire de confiance calcule de façon prédictive l'évaluation probable que l'utilisateur aurait donnée à chaque fournisseur disponible (formule 7.3) et recommande le meilleur fournisseur à l'utilisateur (deuxième phase du modèle SOM-BTR). Après avoir consommé le service sélectionné, l'utilisateur transmet son retour au gestionnaire de confiance. Ce retour, ainsi que d'autres seront utilisés dans la prochaine mise à jour hors ligne de SOM_S et SOM_U (phase trois de SOM-BTR).

Validation

Nous validons le cas d'utilisation précédent avec un exemple tiré des données utilisées dans ce chapitre (voir 8.2). Pour être plus réaliste, nous supprimons au hasard 80 % de la matrice d'origine et introduisons au hasard 20 utilisateurs malveillants. Supposons que l'utilisateur sélectionné est l'utilisateur avec l'ID=12. Cet utilisateur demande au gestionnaire de confiance de lui recommander un service parmi un ensemble de trois services disponibles dans son rayon de connexion radio (par exemple, les services avec les ID : 3918, 5454 et 1456), répertoriés dans la table 8.2 avec leurs profils calculés.

Le profil utilisateur sélectionné est calculé par le système (voir 7.2.2.5) et le résultat est un vecteur de dimension 770 : $P_{12} = \langle M_{12}^1 = 0.135, F_{12}^1 = 0.0068, ..., M_{12}^{770} = 0.192, F_{12}^{770} = 0.0017 \rangle$.

TABLE 8.2 – Exemple de services de la base de données WS-DERAM [156] et leurs profils

Service ID	**Fournisseur du service**	**Pays**	Mc_s^1	Mc_s^2	Ec_s^1	Ec_s^2
3918	FNIS	État-Unis	0.0523	0.1637	0.5224	0.4478
5454	moss2007.be	Belgique	0.101	0.3021	0.8507	0.1194
1456	centraline.com	Allemagne	1.8272	6.7392	0.8824	0.0882

Après la projection de ces profils sur les modèles SOM et le filtrage des évaluations non fiables (Formules 7.11 à 7.13 et expression 7.14), les résultats du calcul de la valeur de l'évaluation probable pour chaque service selon le modèle SOM-BTR sont listés dans le tableau 8.3.

La première observation est que les valeurs prédites sont proches des valeurs réelles

observées. Nous pouvons affirmer que le modèle proposé recommandera, pour l'ID utilisateur = 12, le service ID = 3918, qui a la valeur la plus basse (meilleur temps de réponse évalué).

TABLE 8.3 – La prédiction pour l'utilisateur ID=12

Service ID	**Valeur prédite**	**Valeur observée dans la base de données originale**
3918	**0.061**	**0.078**
5454	0.191	0.172
1456	1.736	1.712

D'après les expérimentations et l'étude de cas décrites ci-dessus, nous pouvons affirmer que le modèle de confiance proposé :

1. fournit une recommandation personnalisée pour chaque utilisateur, 2. implique un modèle précalculé pour réduire les besoins calculatoires, 3. inclut les évaluations antérieures avec une mise à jour périodique hors ligne des modèles SOM et 4. détecte les utilisateurs non fiables ou malveillants avant de calculer la recommandation.

8.10 Conclusion

Les expérimentations ont été effectuées sur un ensemble de données réelles et ont montré que la proposition est efficace et robuste face aux utilisateurs malveillants et au manque de données. Une étude de cas a été décrite avec ce modèle de confiance et a montré que la solution présente un certain nombre d'avantages dans une situation de cas d'utilisation réelle. Ce travail est une première approche utilisant les cartes de Kohonen pour la recommandation de services de confiance. Néanmoins, nous prévoyons d'améliorer les performances du modèle notamment en permettant la sélection dynamique des paramètres k et α.

Quatrième partie

Conclusion et perspectives

Conclusion

Dans cette thèse nous nous sommes attelés à proposer un modèle intégré pour assurer la protection des données personnelles centrée sur le propriétaire de données et ce, dans un environnement d'Internet des objets. Ce modèle a été construit suivant deux axes : la sécurité cryptographique et la gestion de la confiance. Deux axes complémentaires, à travers lesquels nous avons employé des techniques prometteuses pour la protection de données personnelles, qui sont le chiffrements basé sur les attributs et la Blockchain mais aussi les cartes auto-organisatrices, qui sont un outil issu de l'intelligence artificielle, pour la gestion de la confiance. Aussi, nous rappelons, ci-après, les contributions de notre thèse.

Contributions

La première contribution consiste à adapter un schéma de chiffrement basé sur les attributs pour une utilisation dans le contexte de l'Internet des objets. Cette adaptation permet l'implémentation de la primitive de chiffrement sur des objets ayant des contraintes d'énergie, de mémoire et de capacité de calcul. Ceci est réalisé à travers notre schéma de chiffrement FCCP-ABE, où les tâches de calcul lourdes sont déportées vers des nœuds non soumis aux contraintes de capacité.

Dans un second temps, et c'est notre deuxième contribution, nous nous sommes penchés sur l'intégration du schéma FCCP-ABE dans un protocole où les échanges sont contrôlés et validés via la Blockchain. L'utilisation de la Blockchain permet d'assurer la fonction d'autorité de confiance de façon décentralisée, évitant ainsi la vulnérabilité d'une autorité de confiance centralisée. Par ailleurs, la solution proposée permet de réduire les privilèges du fournisseur de services de déportation des fonctionnalités informatiques, qui ont aucun

accès à l'information en clair durant tout le processus de collecte, chiffrement, transfert, stockage et puis partage des données.

La troisième contribution de notre thèse résulte de l'interaction des utilisateurs avec l'environnement de l'informatique en brouillard (Fog-Computing) et plus généralement avec les services d'externalisation des fonctionnalités informatiques (ex. Cloud computing). Cette contribution consiste en la gestion de la confiance envers ces services. Cette gestion de la confiance, nous l'avons réalisée à travers un modèle de filtrage collaboratif avec détection des utilisateurs non fiables SOM-BTR. Pour se faire, nous avons fait appel aux outils de l'intelligence artificielle qui sont des cartes auto-organisatrices pour évaluer la similarité et les K-moyennes pour améliorer l'efficacité du filtrage collaboratif.

Perspectives et travaux futurs

Un travail préliminaire, a été initialisé durant cette thèse de doctorat. Ce travail consistait à explorer les possibilités offertes par l'algorithme d'optimisation par colonies de fourmis (ou ACO, pour Ant-colony Optimisation) pour détecter les nœuds vulnérables dans un réseau d'objets connectés. Ce travail que nous avons choisi de ne pas décrire dans cette thèse, a fait l'objet de la publication d'un chapitre d'ouvrage [109]. Il pourrait être davantage développé et approfondi en vue de proposer un module d'alerte dédié à l'environnement de l'Internet des objets, ce qui pourrait être une première perspective à notre travail. En effet, l'algorithme ACO est un algorithme inspiré du comportement social des fourmis pour la recherche du plus court chemin menant à la source de nourriture. L'idée que nous pourrions développer serait l'utilisation d'un graphe aléatoire, représentant les interconnexions des objets connectés, sur lequel les fourmis se déplaceraient. Le mouvement des fourmis serait influencé par certains paramètres caractérisant le nœud (temps d'accès, bande passante, etc.) et en observant leurs mouvements on pourrait détecter les nœuds sûrs de ceux présentant un risque pour la sécurité.

Concernant notre modèle SOM-BTR, en vue d'améliorer l'évaluation de la similarité dans notre modèle SOM-BTR, un travail a été engagé en vue de permettre à l'algorithme des cartes auto-organisatrices de converger vers un minimum global au lieu d'un minimum local.

Pour se faire, une réflexion est en cours pour utiliser une stratégie hybride qui s'inspirerait du comportement des lucioles (firefly algorithm en anglais) pour initialiser de façon la plus optimale possible les paramètres initiaux de l'algorithme des cartes auto-organisatrices.

Enfin, pour ouvrir d'autres possibilités d'application pour notre schéma de chiffrement FCCP-ABE, nous pourrions le faire évoluer pour qu'il puisse également prendre en charge la fonction de déchiffrement sur des objets contraints. Ceci permettra, par exemple, d'exploiter les données sur des dispositifs mobiles (type tablette ou smartphone). Nous visons particulièrement à déporter l'exécution de l'algorithme récursif $DecryptNode$, qui prend comme argument la clé secrète Sk. Or, pour rester dans l'esprit de notre modèle, le nœud de déportation ne devrait pas avoir accès à cette clé pour pouvoir accomplir ses calculs.

Publications

Revues

— Youcef Ould-Yahia, Meziane Yacoub, Samia Bouzefrane et Hanifa Boucheneb : *Self Organized Map and trust-aware-based quality of service prediction for reliable services selection in distributed computing environment*. International Journal of Advanced Intelligence Paradigms, (Article accepté en 2019, à paraître).

— Youcef Ould-Yahia, Samia Bouzefrane, Hanifa Boucheneb et Soumya Banerjee : *A data-owner centric privacy model with blockchain and adapted attribute-based encryption for Internet-of-Things and Cloud environment*. International Journal of Information and Computer Security, (Article accepté en 2019, à paraître).

Conférences

— Youcef Ould-Yahia, Samia Bouzefrane et Hanifa Boucheneb : *Towards privacy and ownership preserving of outsourced health data in iot-cloud context*. In 2018 International Symposium on Programming and Systems (ISPS), pages 1–6, April 2018.

— Youcef Ould-Yahia et Pierre Paradinas : *Applications e-santé, le contrôle des données personnelles un enjeu majeur pour la protection de la vie privée*. In Computer and Electronics Security Applications Rendez-vous (C&ESAR2017), November 2017.

Chapitre d'ouvrage

— Youcef Ould-Yahia, Soumya Banerjee, Samia Bouzefrane et Hanifa Boucheneb : *Exploring formal strategy framework for the security in iot towards e-healthcontext using computational intelligence*. Internet of Things and Big Data Technologies for Next Generation Healthcare, Springer International Publishing, pages 63–90, 2017.

Bibliographie

[1] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), avril 2016. 79

[2] J. H. Abawajy et M. M. Hassan : Federated Internet of Things and Cloud Computing Pervasive Patient Health Monitoring System. *In IEEE Communications Magazine*, volume 55, pages 48–53, janvier 2017. 38

[3] Mohammed Riyadh Abdmeziem et Djamel Tandjaoui : An end-to-end secure key management protocol for e-health applications. *In Comput. Electr. Eng.*, volume 44, pages 184–197, 2015. 38

[4] Arif Ahmed et Ejaz Ahmed : A survey on mobile edge computing. *In 10th IEEE International Conference on Intelligent Systems and Control, (ISCO 2016)*, 01 2016. 28, 31

[5] A. T. Akinola et M. O. Adigun : Feedback-based service selection in ad-hoc mobile cloud computing. *In 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, pages 172–177, 2016. 52

[6] M. Ambrosin, A. Anzanpour, M. Conti, T. Dargahi, S. R. Moosavi, A. M. Rahmani et P. Liljeberg : On the feasibility of attribute-based encryption on internet of things devices. *In IEEE Micro*, volume 36, pages 25–35, Nov 2016. 48

[7] Moreno Ambrosin, Mauro Conti et Tooska Dargahi : On the Feasibility of Attribute-Based Encryption on Smartphone Devices. *In Proceedings of the 2015*

Workshop on IoT Challenges in Mobile and Industrial Systems, pages 49–54. ACM, 2015. 47, 48, 78

[8] David Arthur et S Vassilvitskii : How slow is the k-means method ? *Proceedings of the Annual Symposium on Computational Geometry*, pages 144–153, 01 2006. 122

[9] M. Asim, M Petkovic et T Ignatenko : Attribute-based encryption with encryption and decryption outsourcing. *In Proceedings of the 19th Conference on Innovations in Clouds, Internet and Networks*, pages 21–28, 2014. 39, 49, 78, 100, 103

[10] A. Azaria, A. Ekblaw, T. Vieira et A. Lippman : Medrec : Using blockchain for medical data access and permission management. *In 2016 2nd International Conference on Open and Big Data (OBD)*, pages 25–30, 2016. 55

[11] S Balaji et K Rajkumar : A personalized cloud service recommendation system using collaborative filtering. *International Journal of Pure and Applied Mathematics*, 119:14173–14180, 01 2018. 52

[12] Balamurugan Balamurugan et P Krishna : Extensive survey on usage of attribute based encryption in cloud. *Journal of Emerging Technologies in Web Intelligence*, 6:263–272, 01 2014. 46

[13] Elaine B. Barker, William C. Barker, William E. Burr, W. Timothy Polk et Miles E. Smid : Sp 800-57. recommendation for key management, part 1 : General (revised). 2007. 67

[14] Amos Beimel : *Secure Schemes for Secret Sharing and Key Distribution*. Thèse de doctorat, Israel Institute of Technology, Technion, Haifa, Israel, 1996. 64

[15] Robert Bell, Yehuda Koren et Chris Volinsky : Modeling relationships at multiple scales to improve accuracy of large recommender systems. *In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, pages 95–104, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-609-7. 116

[16] Mihir Bellare : Practice-oriented provable-security. *In Lectures on Data Security Lecture Notes in Computer Science*, volume 1396, page 1â15. Springer Verlag, Berlin, 1999. 179

[17] J. Benaloh, J. et Leichter : Generalized secret sharing and monotone functions. *In Proceedings on Advances in Cryptology*, CRYPTO '88, pages 27–35, New York, NY, USA, 1990. Springer-Verlag New York, Inc. ISBN 0-387-97196-3. 82

[18] K. Benouaret, I. Benouaret, M. Barhamgi et D. Benslimane : Top-k cloud service plans using trust and qos. *In 2017 IEEE International Conference on Services Computing (SCC)*, pages 507–510, 2017. 39, 52

[19] John Bethencourt, Amit Sahai et Brent Waters : Ciphertext-Policy Attribute-Based Encryption. *In Proceedings of the 2007 IEEE Symposium on Security and Privacy*, SP '07, pages 321–334, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 978-0-7695-2848-9. 46, 84, 98, 100, 103, 179

[20] Flavio Bonomi, Rodolfo Milito, Jiang Zhu et Sateesh Addepalli : Fog computing and its role in the internet of things. *In Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, MCC '12, pages 13–16, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1519-7. 31, 77

[21] John S. Breese, David Heckerman et Carl Kadie : Empirical analysis of predictive algorithms for collaborative filtering. *In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 43–52, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc. ISBN 1-55860-555-X. 116

[22] CGTN : Jorf no 0129 du 6 juin 2010 p. texte no 42. vocabulaire de l'informatique et de l'internet. URL `https://www.legifrance.gouv.fr`. 29

[23] Praveen S. Challagidad, Vani S. Reshmi et Mahantesh N. Birje : Reputation based trust model in cloud computing. *In Internet of Things and Cloud Computing*, volume 5, pages 5–12, 2017. 51, 53

[24] Y. Chandu, K. S. R. Kumar, N. V. Prabhukhanolkar, A. N. Anish et S. Rawal : Design and implementation of hybrid encryption for security of iot data. *In 2017*

International Conference On Smart Technologies For Smart Nation (SmartTechCon), pages 1228–1231, 2017. 42, 80

[25] ALCHIEKH-HAYDAR Charif : *Les systèmes de recommandation à base de confiance.* Thèse de doctorat, Université de Lorraine, France, 2014. 113

[26] Zhen Chen, Limin Shen et Feng Li : Exploiting web service geographical neighborhood for collaborative qos prediction. *Future Generation Computer Systems*, 68:248 – 259, 2017. ISSN 0167-739X. 52, 144

[27] Matin Chiregi et Nima Jafari Navimipour : A new method for trust and reputation evaluation in the cloud environments using the recommendations of opinion leaders' entities and removing the effect of troll entities. *Computers in Human Behavior*, 60:280 – 292, 2016. ISSN 0747-5632. 53

[28] Rasel Chowdhury, Hakima Ould-Slimane, Chamseddine Talhi et Mohamed Cheriet : Attribute-based encryption for preserving smart home data privacy. *In* Mounir Mokhtari, Bessam Abdulrazak et Hamdi Aloulou, éditeurs : *Enhanced Quality of Life and Smart Living*, pages 185–197, Cham, 2017. Springer International Publishing. ISBN 978-3-319-66188-9. 78

[29] K. Christidis et M. Devetsikiotis : Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4:2292–2303, 2016. ISSN 2169-3536. 55

[30] M. Conoscenti, A. VetrÃ et J. C. De Martin : Blockchain for the internet of things : A systematic literature review. *In 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pages 1–6, 2016. 56

[31] Flavio Corradini, Francesco De Angelis, Fabrizio Ippoliti et Fausto Marcantoni : A survey of trust management models for cloud computing. *In CLOSER 2015 - 5th International Conference on Cloud Computing and Services*, 05 2015. 51

[32] R. Cramer et V. Shoup : Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal on Computing*, 33(1):167–226, 2003. 74

[33] Michel Crucianu et Meziane Yacoub : Cours sur la reconnaissance des formes et méthodes neuronales (rcp208) du conservatoire nationale des arts et métiers, paris. URL http://cedric.cnam.fr/vertigo/Cours/ml/index.html. 114

[34] Christian Desrosiers et George Karypis : A comprehensive survey of neighborhood-based recommendation methods. *In* Francesco Ricci, Lior Rokach, Bracha Shapira et Paul B. Kantor, éditeurs : *Recommender Systems Handbook*, pages 107–144. Springer US, 2011. ISBN 978-0-387-85820-3. 115

[35] Hoang Dinh Thai, Chonho Lee, Dusit Niyato et Ping Wang : A survey of mobile cloud computing : Architecture, applications, and approaches. *In Wireless Communications and Mobile Computing*, volume 13, 12 2013. 77

[36] D. Dolev et A. Yao : On the security of public key protocols. *IEEE Transactions on Information Theory*, 29(2):198–208, Mar 1983. ISSN 0018-9448. 81

[37] Charalampos Doukas, Thomas Pliakas et Ilias Maglogiannis : Mobile healthcare information management utilizing Cloud Computing and Android OS. *In Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, volume 2010, pages 1037–1040, 2010. 27

[38] A. E. H. G. El-Barbary, L. A. A. El-Sayed, H. H. Aly et M. N. El-Derini : A cloudlet architecture using mobile devices. *In 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pages 1–8, 2015. 77

[39] Pierre-Alain Fouque : Cours de cryptographie. URL www.di.ens.fr. 61

[40] D. Gambetta : Trust. making and breaking cooperative relations. *Tijdschrift Voor Filosofie*, 52(4):740–740, 1990. 50

[41] A. Gatzioura et M. SÃ nchez-MarrÃ : A case-based recommendation approach for market basket data. *IEEE Intelligent Systems*, 30(1):20–27, 2015. ISSN 1541-1672. 51

[42] Ken Goldberg, Theresa Roeder, Dhruv Gupta et Chris Perkins : Eigentaste : A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001. 116

[43] T. Gong, H. Huang, P. Li, K. Zhang et H. Jiang : A Medical Healthcare System for Privacy Protection Based on IoT. *In 2015 Seventh International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, pages 217–222, 2015. 38

[44] P. Gope et T. Hwang : BSN-Care : A Secure IoT-Based Modern Healthcare System Using Body Sensor Network. *In IEEE Sensors Journal*, volume 16, mars 2016. 26, 77

[45] Vipul Goyal, Omkant Pandey, Amit Sahai et Brent Waters : Attribute-based Encryption for Fine-grained Access Control of Encrypted Data. *In Proceedings of the 13th ACM Conference on Computer and Communications Security*, CCS '06, New York, NY, USA, 2006. ACM. ISBN 978-1-59593-518-2. 45, 86

[46] O. M. Guillen, T. Appelmann, J. M. Bermudo Mera, E. F. Bongenaar, G. Sigl et J. Sepulveda : Towards post-quantum security for iot endpoints with ntru. *In Design, Automation Test in Europe Conference Exhibition (DATE), 2017*, pages 698–703, 2017. 44

[47] F. Guo, Y. Mu, W. Susilo, D. S. Wong et V. Varadharajan : CP-ABE With Constant-Size Keys for Lightweight Devices. *In IEEE Transactions on Information Forensics and Security*, volume 9, pages 763–771, mai 2014. 49, 78

[48] L. Guo, X. Zheng, C. Ding, D. Mu et Z. Li : Cloud service recommendation : State of the art and research challenges. *In 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 761–764, 2015. 118

[49] Y. Gupta, R. Shorey, D. Kulkarni et J. Tew : The applicability of blockchain in the internet of things. *In 2018 10th International Conference on Communication Systems Networks (COMSNETS)*, pages 561–564, 2018. 55

[50] Darrel Hankerson, Alfred Menezes et Scott Vanstone Springer : *Guide to Elliptic Curve Cryptography*, volume 332. Springer-Verlag New York, 01 2004. 66

[51] S. H. Hashemi, F. Faghri, P. Rausch et R. H. Campbell : World of empowered iot users. *In 2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 13–24, 2016. 55, 56

[52] S. R. Hemalatha et Manickachezian : Security Strength of RSA and Attribute Based Encryption for Data Security in Cloud Computing. *In International Journal of Innovative Research in Computer and Communication Engineering*, 2014. 45, 67, 78

[53] M. S. Henriques et N. K. Vernekar : Using symmetric and asymmetric cryptography to secure communication between devices in iot. *In 2017 International Conference on IoT and Application (ICIOT)*, pages 1–4, 2017. 42

[54] R. Henry, A. Herzberg et A. Kate : Blockchain access privacy : Challenges and directions. *IEEE Security Privacy*, 16(4):38–45, 2018. ISSN 1540-7993. 57

[55] Jordi Herrera-Joancomartí et Cristina Pérez-Solà : Privacy in bitcoin transactions : New challenges from blockchain scalability solutions. *In Modeling Decisions for Artificial Intelligence*, pages 26–44. Springer International Publishing, 2016. ISBN 978-3-319-45656-0. 56

[56] Jeffrey Hoffstein, Jill Pipher et Joseph H. Silverman : Ntru : A ring-based public key cryptosystem. *In Lecture Notes in Computer Science*, pages 267–288. Springer-Verlag, 1998. 44

[57] R. Hu, W. Dou et J. Liu : A hybrid collaborative filtering approach for multi-functional service recommendation. *In 2013 International Conference on Cloud and Green Computing*, pages 54–59, 2013. 52

[58] Jingwei Huang et David M. Nicol : Trust mechanisms for cloud computing. *Journal of Cloud Computing : Advances, Systems and Applications*, 2(1):9, 2013. ISSN 2192-113X. 50

[59] S. Huang, X. Jiang, N. Zhang, C. Zhang et D. Dang : Collaborative filtering of web service based on mapreduce. *In 2014 International Conference on Service Sciences*, pages 91–95, 2014. 52, 116, 118

[60] J. H. Jeon, K. Kim et J. Kim : Block chain based data security enhanced iot server platform. *In 2018 International Conference on Information Networking (ICOIN)*, pages 941–944, 2018. 56

[61] S. J. Johnston, M. Scott et S. J. Cox : Recommendations for securing internet of things devices using commodity hardware. *In 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 307–310, Dec 2016. 44

[62] Audun Josang, Roslan Ismail et Colin Boyd : A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618 – 644, 2007. ISSN 0167-9236. 51

[63] Antoine Joux : A one round protocol for tripartite diffie-hellman. *In Algorithmic Number Theory Symposium*, pages 385–394, 01 2000. 64

[64] S. R. U. Kakakhel, L. Mukkala, T. Westerlund et J. Plosila : Virtualization at the network edge : A technology perspective. *In 2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 87–92, 2018. 39

[65] M. S. Karunarathne, S. A. Jones, S. W. Ekanayake et P. N. Pathirana : Remote Monitoring System Enabling Cloud Technology upon Smart Phones and Inertial Sensors for Human Kinematics. *In 2014 IEEE Fourth International Conference on Big Data and Cloud Computing*, pages 137–142, décembre 2014. 27

[66] Jonathan Katz et Yehuda Lindell : *Introduction to Modern Cryptography, Second Edition*. CRC Press, 2014. ISBN 9781466570269. 61, 72

[67] H. Khemissa et D. Tandjaoui : A Lightweight Authentication Scheme for E-Health Applications in the Context of Internet of Things. *In 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies*, pages 90–95, 2015. 38

[68] Ovunc Kocabas, Tolga Soyata et Mehmet K. Aktas : Emerging Security Mechanisms for Medical Cyber Physical Systems. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 13(3):401–416, juin 2016. ISSN 1557-9964. 45, 78

[69] Teuvo KOHONEN : Essentials of the self-organizing map. *Neural Networks*, 37:52 – 65, 2013. ISSN 0893-6080. 22, 119, 120

[70] Philip KOSHY, Diana KOSHY et Patrick D. MCDANIEL : An analysis of anonymity in bitcoin using p2p network traffic. *In Financial Cryptography*, 2014. 57

[71] Djamel Eddine KOUICEM, Abdelmadjid BOUABDALLAH et Hicham LAKHLEF : Internet of things security : A top-down survey. *Computer Networks*, 141:199 – 221, 2018. ISSN 1389-1286. 40

[72] Li KUANG, Long YU, Lan HUANG, Youxiang WANG, Pengju MA, Chuanbin LI et Yujia ZHU : A personalized qos prediction approach for cps service recommendation based on reputation and location-aware collaborative filtering. *In Sensors*, 2018. 53, 144

[73] K. M. KUMAR et A. R. M. REDDY : A fast k-means clustering using prototypes for initial cluster center selection. *In 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)*, pages 1–4, 2015. 122

[74] R. LAMDANI, N. BENNACER, G. POLAILLON et Y. BOURDA : A collaborative and semantic-based approach for recommender systems. *In 2010 10th International Conference on Intelligent Systems Design and Applications*, pages 469–476, 2010. 114

[75] Bassam Abdel LATIF et Gregoire MERCIER : Self-organizing maps for processing of data with missing values and outliers : Application to remote sensing images. *In Self-Organizing Maps*, chapitre 12. IntechOpen, Rijeka, 2010. 121

[76] Cheng-Chi LEE, Pei-Shan CHUNG et Min-Shiang HWANG : A survey on attribute-based encryption schemes of access control in cloud environments. *I. J. Network Security*, 15:231–240, 2013. 45, 46

[77] B. LI, R. SONG, L. LIAO et C. LIU : A user-oriented trust model for web services. *In 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering*, pages 224–232, 2013. 132

[78] Ming Li, Shucheng Yu, Kui Ren et Wenjing Lou : Securing personal health records in cloud computing : Patient-centric and fine-grained data access control in multi-owner settings. *In Security and Privacy in Communication Networks*, pages 89–106. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-16161-2. 37, 38, 47

[79] Ming Li, Shucheng Yu, Kui Ren et Wenjing Lou : Securing Personal Health Records in Cloud Computing : Patient-Centric and Fine-Grained Data Access Control in Multi-owner Settings. *In Security and Privacy in Communication Networks*, pages 89–106. Springer Berlin Heidelberg, 2010. 45, 78, 81

[80] X. Li, H. Ma, F. Zhou et X. Gui : Service operator-aware trust scheme for resource matchmaking across multiple clouds. *IEEE Transactions on Parallel and Distributed Systems*, 26(5):1419–1429, 2015. ISSN 1045-9219. 51

[81] H. Lin, J. Shao, C. Zhang et Y. Fang : CAM : Cloud-Assisted Privacy Preserving Mobile Health Monitoring. *In IEEE Transactions on Information Forensics and Security*, volume 8, pages 985–997, 2013. 78

[82] S. Lloyd : Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2), 1982. 122

[83] Zheng Longshuai, Yang Shengqi, He Jian et Huang Zhangqin : An optimized collaborative filtering recommendation algorithm. *In 2016 2nd International Conference on Cloud Computing and Internet of Things (CCIOT)*, pages 89–92, 2016. 52

[84] Ben Lynn : *On the implementation of pairing-based cryptosystems*. Thèse de doctorat, Stanford University, 2007. 99

[85] J. Macqueen : Some methods for classification and analysis of multivariate observations. *In In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967. 121

[86] Redowan Mahmud, Ramamohanarao Kotagiri et Rajkumar Buyya : *Fog Computing : A Taxonomy, Survey and Future Directions*, pages 103–130. Springer Singapore, Singapore, 2018. ISBN 978-981-10-5861-5. URL https://doi.org/10.1007/978-981-10-5861-5_5. 32

[87] J. Mai, Y. Fan et Y. Shen : A neural networks-based clustering collaborative filtering algorithm in e-commerce recommendation system. *In 2009 International Conference on Web Information Systems and Mining*, pages 616–619, 2009. 120

[88] P.d. Sheba Kezia Malarchelvi, S.s. Manikandasaran et L. Arockiam : Moncrypt : A technique to ensure the confidentiality of outsourced data in cloud storage. *International Journal of Information and Computer Security*, 11(1):1, 2019. 78

[89] C. Mao, R. Lin, C. Xu et Q. He : Towards a trust prediction framework for cloud services based on pso-driven neural network. *IEEE Access*, 5:2187–2199, 2017. ISSN 2169-3536. 120

[90] Josua Melka et Jean-Jacques Mariage : Efficient implementation of self-organizing map for sparse input data. *In Proceedings of the 9th International Joint Conference on Computational Intelligence*, volume 1, Madeira, Portugal, 2017. 121

[91] Raheel Ahmed Memon, Jian Ping Li et Junaid Ahmed : Simulation model for blockchain systems using queuing theory. *Electronics*, 8(2), 2019. ISSN 2079-9292. 104

[92] Victor S. Miller : Use of elliptic curves in cryptography. *In Advances in Cryptology -CRYPTO'85 Proceedings*, pages 417–426. Springer Berlin Heidelberg, 1985. 66

[93] M. Mrabet, Y. b. Saied et L. A. Saidane : A new trust evaluation approach for cloud computing environments. *In 2016 International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*, pages 1–6, 2016. 51, 53

[94] M. Mrabet, Y. b Saied et L. A. Saidane : A new trust evaluation approach for cloud computing environments. *In 2016 International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*, pages 1–6, 2016. 50, 53

[95] L. Mui, M. Mohtashemi et A. Halberstadt : A computational model of trust and reputation. *In Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pages 2431–2439, 2002. 51

[96] R. Nagarajan, S. Selvamuthukumaran et R. Thirunavukarasu : A fuzzy logic based trust evaluation model for the selection of cloud services. *In 2017 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5, 2017. 39

[97] Satoshi Nakamoto : Bitcoin : A peer-to-peer electronic cash system, 2008. 54, 99

[98] Suyel Namasudra : An improved attribute-based encryption technique towards the data security in cloud computing. *Concurrency and Computation : Practice and Experience*, 12 2017. 78

[99] Koblitz Neal : Elliptic curve cryptosystems. *In Mathematics of computation*, volume 48, pages 203–209, 1987. 66

[100] Kim Thuat Nguyen, Maryline Laurent et Nouha Oualha : Survey on secure communication protocols for the internet of things. *Ad Hoc Networks*, 32:17 – 31, 2015. ISSN 1570-8705. Internet of Things security and privacy : design methods and optimization. 44

[101] Kim Thuat Nguyen, Nouha Oualha et Maryline Laurent : Securely outsourcing the ciphertext-policy attribute-based encryption. *World Wide Web*, 21(1):169–183, 2018. ISSN 1573-1413. 49, 180

[102] Talal H. Noor, Quan Z. Sheng, Sherali Zeadally et Jian Yu : Trust management of services in cloud environments : Obstacles and solutions. *ACM Comput. Surv.*, 46 (1):12 :1–12 :30, juillet 2013. ISSN 0360-0300. 51

[103] B. Oniga, S. H. Farr, A. Munteanu et V. Dadarlat : Iot infrastructure secured by tls level authentication and pki identity system. *In 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pages 78–83, 2018. 44

[104] Farah Sarah Ouada, Mawloud Omar, Abdelmadjid Bouabdallah et Abdelkamel Tari : Lightweight identity-based authentication protocol for wireless sensor networks. *International Journal of Information and Computer Security*, 8(2):121, 2016. 78

[105] Aafaf Ouaddah, Anas Abou Elkalam et Abdellah Ait Ouahman : *Towards a Novel Privacy-Preserving Access Control Model Based on Blockchain Technology in IoT*, pages 523–533. 09 2017. 55

[106] N. Oualha et K.T. Nguyen : Lightweight attribute-based encryption for the internet of things. *In 2016 25th International Conference on Computer Communication and Networks (ICCCN)*. Institute of Electrical and Electronics Engineers Inc., 2016. 45, 48

[107] Y. Ould-Yahia, S. Bouzefrane et H. Boucheneb : Towards privacy and ownership preserving of outsourced health data in iot-cloud context. *In 2018 International Symposium on Programming and Systems (ISPS)*, pages 1–6, April 2018. 44, 78

[108] Y. Ould-Yahia et P. Paradinas : Applications e-santé, le contrôle des données personnelles un enjeu majeur pour la protection de la vie privée. *In Computer and Electronics Security Applications Rendez-vous (C&ESAR2017)*, Rennes, France, November 2017. 27, 40, 77

[109] Youcef Ould-Yahia, Soumya Banerjee, Samia Bouzefrane et Hanifa Boucheneb : Exploring formal strategy framework for the security in iot towards e-health context using computational intelligence. *Internet of Things and Big Data Technologies for Next Generation Healthcare*, pages 63–90, 2017. 152

[110] Michael J. Pazzani et Daniel Billsus : *Content-Based Recommendation Systems*, pages 325–341. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 114

[111] J.M. Pollard : Monte carlo methods for index computation mod p. *Mathematics of Computation*, 32, 07 1978. 67

[112] G. Premsankar, M. Di Francesco et T. Taleb : Edge computing for the internet of things : A case study. *In IEEE Internet of Things Journal*, volume 5, pages 1275–1284, 2018. 77

[113] H. S. G. Pussewalage et V. Oleshchuk : A Patient-Centric Attribute Based Access Control Scheme for Secure Sharing of Personal Health Records Using Cloud

Computing. *In 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*, pages 46–53, novembre 2016. 37, 47, 78

[114] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom et John Riedl : Grouplens : An open architecture for collaborative filtering of netnews. *In Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, pages 175–186, New York, NY, USA, 1994. ACM. ISBN 0-89791-689-1. 115, 116, 144

[115] Sebastian Ries : *Trust in Ubiquitous Computing*. Thèse de doctorat, Technische Universitat, Darmstadt, 2009. 50

[116] Rodrigo Roman, Javier Lopez et Masahiro Mambo : Mobile edge computing, fog et al. : A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78:680–698, 2018. ISSN 0167-739X. 39

[117] S. Rouhani et R. Deters : Performance analysis of ethereum transactions in private blockchain. *In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 70–74, 2017. 19, 104, 105

[118] Patrick Rousset : *Applications des algorithmes d'auto-organisation à la classification et à la prévision*. Thèse de doctorat, Paris 1, 1999. 113

[119] Othman Saeed et Riaz Ahmed Shaikh : A user-based trust model for cloud computing environment. *International Journal of Advanced Computer Science and Applications*, 9(3), 2018. 51, 53

[120] Amit Sahai et Brent Waters : Fuzzy identity-based encryption. *In SpringerLink*, pages 457–473. Springer, Berlin, Heidelberg, 2005. 45, 84

[121] Yosra Ben Saied, Alexis Olivereau, Djamal Zeghlache et Maryline Laurent : Trust management system design for the internet of things : A context-aware and multi-service approach. *Computers Security*, 39:351 – 365, 2013. ISSN 0167-4048. 53

[122] Badrul Sarwar, George Karypis, Joseph Konstan et John Riedl : Item-based collaborative filtering recommendation algorithms. *In Proceedings of the 10th Inter-*

national Conference on World Wide Web, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM. ISBN 1-58113-348-0. 115, 117, 118, 144

[123] J. Ben Schafer, Dan Frankowski, Jon Herlocker et Shilad Sen : *Collaborative Filtering Recommender Systems*, pages 291–324. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 118

[124] K. Shah, A. Salunke, S. Dongare et K. Antala : Recommender systems : An overview of different approaches to recommendations. *In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pages 1–4, 2017. 51, 114, 115, 118

[125] Adi Shamir : How to share a secret. *Commun. ACM*, 22(11):612–613, novembre 1979. ISSN 0001-0782. 64, 83

[126] Adi Shamir : How to share a secret. *Commun. ACM*, 22(11):612–613, novembre 1979. ISSN 0001-0782. 82

[127] Daniel Shanks : Class number, a theory of factorization, and genera. *In 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969)*, pages 415–440. 1971. 67

[128] A. Sharma, T. Goyal, E. S. Pilli, A. P. Mazumdar, M. C. Govil et R. C. Joshi : A Secure Hybrid Cloud Enabled architecture for Internet of Things. *In 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, pages 274–279, 2015. 38

[129] A. Singla et E. Bertino : Blockchain-based pki solutions for iot. *In 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 9–15, 2018. 56

[130] Bonecaze Stéphane, Ballet et Alexis : cours sur les courbes elliptiques application à la cryptographie. 5e année de systèmes informatiques critiques et applications. URL `www.polytech.univ-amu.fr`. 61

[131] Kai Su, Bin Xiao, Baoping Liu, Huaiqiang Zhang et Zongsheng Zhang : Tap : A

personalized trust-aware qos prediction approach for web service recommendation. *Knowledge-Based Systems*, 115:55 – 65, 2017. ISSN 0950-7051. 53, 144

[132] Xiaoyuan Su et Taghi M. Khoshgoftaar : A survey of collaborative filtering techniques. *Adv. in Artif. Intell.*, pages 4 :2–4 :2, 2009. ISSN 1687-7470. 115, 138

[133] I. Sukhodolskiy et S. Zapechnikov : A blockchain-based access control system for cloud storage. *In 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pages 1575–1578, 2018. 55, 56

[134] B. Sung, K. Kim et K. Shin : An aes-gcm authenticated encryption crypto-core for iot security. *In 2018 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–3, 2018. 42, 44

[135] Gábor Takács, István Pilászy, Bottyán Németh et Domonkos Tikk : Major components of the gravity recommendation system. *SIGKDD Explorations*, 9:80–83, 12 2007. 115

[136] Michael Till et Marco Maier : Mobile edge computing : Challenges for future virtual network embedding algorithms. *In The Eighth International Conference on Advanced Engineering Computing and Applications in Sciences*, 2014. 29

[137] L. Touati, Y. Challal et A. Bouabdallah : C-cp-abe : Cooperative ciphertext policy attribute-based encryption for the internet of things. *In 2014 International Conference on Advanced Networking Distributed Systems and Applications*, pages 64–69, June 2014. 49, 78

[138] L.M. Vaquero et Luis Rodero-Merino : Finding your way in the fog : Towards a comprehensive definition of fog computing. *HP Laboratories Technical Report*, 44, 01 2014. 29

[139] M. Wang, G. Wang, Y. Zhang et Z. Li : A high-reliability multi-faceted reputation evaluation mechanism for online services. *IEEE Transactions on Services Computing*, pages 1–1, 2018. ISSN 1939-1374. 53

[140] X. Wang, J. Zhang, E. M. Schooler et M. Ion : Performance evaluation of attribute-based encryption : Toward data privacy in the iot. *In 2014 IEEE International Conference on Communications (ICC)*, pages 725–730, 2014. 38, 47

[141] X. Wang, J. Zhang, E. M. Schooler et M. Ion : Performance evaluation of Attribute-Based Encryption : Toward data privacy in the IoT. *In 2014 IEEE International Conference on Communications (ICC)*, pages 725–730, 2014. 48, 78

[142] X. A. Wang, J. Ma et F. Xhafa : Outsourcing Decryption of Attribute Based Encryption with Energy Efficiency. *In 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, pages 444–448, novembre 2015. 39

[143] Yating Wang, Ing-Ray Chen et Ding-Chau Wang : A survey of mobile cloud computing applications : Perspectives and challenges. *Wireless Personal Communications*, 80(4):1607–1623, Feb 2015. 29

[144] Brent Waters : Ciphertext-policy attribute-based encryption : An expressive, efficient, and provably secure realization. *In* Dario Catalano, Nelly Fazio, Rosario Gennaro et Antonio Nicolosi, éditeurs : *Public Key Cryptography – PKC 2011*, pages 53–70, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-19379-8. 70, 86

[145] Andrew Whitby, Audun JÃ¸sang et Jadwiga Indulska : Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research*, 4(2):48–64, 2005. 132

[146] Chen Wu, Weiwei Qiu, Zibin Zheng, Xinyu Wang et Xiaohu Yang : Qos prediction of web services based on two-phase k-means clustering. *In Proceedings of the 2015 IEEE International Conference on Web Services*, ICWS '15, pages 161–168, Washington, DC, USA, 2015. IEEE Computer Society. ISBN 978-1-4673-7272-5. 116

[147] Z. Wu et Y. Zhou : Service trustworthiness evaluation using neural network and fuzzy logic. *In 2016 IEEE International Conference on Services Computing (SCC)*, pages 563–570, June 2016. 51, 54

[148] Yilong YANG, Peng LIU, Lianchao DING, Bingqing SHEN et Weiru WANG : Servenet : A deep neural network for web service classification. *CoRR*, abs/1806.05437, 2018. 39, 51, 54

[149] Xuanxia YAO, Zhi CHEN et Ye TIAN : A lightweight attribute-based encryption scheme for the internet of things. *Future Generation Computer Systems*, 49, 10 2014. 50

[150] Shanhe YI, Cheng LI et Qun LI : A survey of fog computing : Concepts, applications and issues. *In Proceedings of the 2015 Workshop on Mobile Big Data*, Mobidata '15, pages 37–42, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3524-9. 77

[151] W. YU, F. LIANG, X. HE, W. G. HATCHER, C. LU, J. LIN et X. YANG : A survey on the edge computing for the internet of things. *IEEE Access*, 6:6900–6919, 2018. ISSN 2169-3536. 31

[152] J. YUAN et X. LI : A reliable and lightweight trust computing mechanism for IoT edge devices based on multi-source feedback information fusion. *IEEE Access*, 6:23626–23638, 2018. 51

[153] Peng ZHANG, Zehong CHEN, Joseph K. LIU, Kaitai LIANG et Hongwei LIU : An efficient access control scheme with outsourcing capability and attribute update for fog computing. *In Future Generation Computer Systems*, volume 78, pages 753 – 762, 2018. 19, 50, 64, 78, 100, 103, 104, 179

[154] X. ZHENG, L. D. XU et S. CHAI : Qos recommendation in cloud services. *IEEE Access*, 5:5171–5177, 2017. ISSN 2169-3536. 39

[155] Z. ZHENG, H. MA, M. R. LYU et I. KING : Wsrec : A collaborative filtering based web service recommender system. *In 2009 IEEE International Conference on Web Services*, pages 437–444, 2009. 52, 144

[156] Zibin ZHENG et M. R. LYU : Ws-dream : A distributed reliability assessment mechanism for web services. *In 2008 IEEE International Conference on Dependable Systems and Networks With FTCS and DCC (DSN)*, pages 392–397, 2008. 20, 137, 146

[157] Z. Zhou et D. Huang : Efficient and secure data storage operations for mobile cloud computing. *In 2012 8th international conference on network and service management (cnsm) and 2012 workshop on systems virtualiztion management (svm)*, pages 37–45, 2012. 49, 78, 100

[158] G. Zyskind, O. Nathan et A. '. Pentland : Decentralizing privacy : Using blockchain to protect personal data. *In 2015 IEEE Security and Privacy Workshops*, pages 180–184, 2015. 55

Cinquième partie

Annexes

Annexe A

Preuve de la sécurité du schéma ECCP-ABE

Pour prouver la sécurité du schéma proposé, la méthode couramment utilisée dans la littérature consiste à réduire le problème à un problème difficile connu, tel que DBDH [16].

Cette démonstration se base sur la démarche de la preuve de sécurité de Bethencourt *et al.* [19]. Cette démarche revient à dire que s'il existe une vulnérabilité dans notre schéma de chiffrement, alors, il y a une vulnérabilité dans les propriétés mathématique du groupe symétrique et dans la fonction de hachage utilisés. Il est utile de renvoyer le lecteur à l'analyse de sécurité au niveau intuitive de Bethencourt *et al.* [19] que nous avons également développée dans la sectionne 5.2.0.1.

Par ailleurs, nous avons choisi de ne pas aborder dans cette démonstration la sécurité du chiffrement symétrique utilisé que nous supposons suffisamment sûre pour ne pas compromettre notre schéma. D'où l'absence du terme C_4 de CT_{two} dans notre preuve.

Concernant la fonction de hachage, elle est modélisée par un oracle aléatoire [19].

Concrètement, la sécurité de notre schéma revient à prouver le théorème suivant, et pour se faire nous allons procéder par un raisonnement par l'absurde. A ce niveau, il est aussi utile de renvoyer le lecteur aux définitions 17 et 18, relatives respectivement à l'avantage d'un algorithme et au schème de chiffrement sûre.

Le formalisme utilisé dans cette preuve est le même que celui développé par Zhang *et al.* [153].

Théorème 2

Si un adversaire probabiliste polynomial peut casser notre schéma avec un avantage non négligeable, alors un simulateur s'exécutant également en temps polynomial peut être construit pour distinguer le tuple DBDH d'un tuple aléatoire avec un avantage non négligeable.

Enfin, pour plus de rigueur, nous rappelons le *Lemme* suivant qui sera utilisé dans la démonstration Nguyen *et al.* [101].

Lemme : *Soit p un nombre premier, étant donné $k \in \mathbb{Z}_p$ et $r \in \mathbb{Z}_p$ uniformément distribuée, alors $a = k+r \ mod(p)$ et $b = kr \ mod(p)$ sont également uniformément distribués.*

Preuve. Supposons que nous ayons un adversaire probabiliste polynomial $\mathcal{A}$ avec un avantage non négligeable $Adv_A = \varepsilon$ susceptible de casser notre schéma. Nous allons démontrer que nous pouvons, alors, construire un simulateur $\mathcal{B}$ pouvant résoudre le problème de DBDH avec un avantage non négligeable.

Soit le challenger $\mathcal{C}$ qui va définir les groupes $\mathbb{G}_0$ et $\mathbb{G}_T$ et un couplage bilinéaire calculable e ainsi qu'un générateur g du groupe $\mathbb{G}_0$. Le challenger $\mathcal{C}$ choisit de manière aléatoire : $a, b, c \in \mathbb{Z}_p$ et sélectionne de manière aléatoire $\mu \in \{0, 1\}$. Si $\mu = 0$ alors le défi est $(A, B, C, Z) = (g^a, g^b, g^c, e(g, g)^{abc})$, sinon le défi est (A, B, C, Z) avec Z un élément aléatoire de $\mathbb{G}_T$. $\mathcal{C}$ envoie le défi à $\mathcal{B}$.

Par la suite, le simulateur $\mathcal{B}$ va utiliser $\mathcal{A}$ pour trouver la solution du problème DBDH, comme suit :

— **Init :** $\mathcal{A}$ choisit un ensemble d'attributs pour générer une structure d'accès P^* et l'envoie à $\mathcal{B}$.

— **Setup** L'algorithme $\mathcal{B}$ simule les paramètres publiques du schéma en les générant avec une distribution identique à celle du schéma d'origine. $\mathcal{B}$ choisi aléatoirement $d \in \mathbb{Z}_p$ et définit implicitement $\alpha/\beta = d + ab$ alors $e(g, g)^{\alpha/\beta} = e(g, g)^{d+ab} = e(g, g)^d e(g, g)^{ab} = e(g, g)^d e(A, B)$ et il pose également $\beta = b$, pour avoir $h = g^\beta = g^b = B$ et fourni la clé publique PK^*, ainsi simulé à $\mathcal{A}$.

— **Phase 1 :** $\mathcal{A}$ peut soumettre de manière adaptative tout ensemble d'attributs S à $\mathcal{B}$, de sorte que S ne satisfait pas la structure d'accès P^* et $\mathcal{B}$ répond avec la clé secrète

SK^* correspondant à l'ensemble soumis S.

Pour simuler des clés secrètes, $\mathcal{B}$ sélectionne aléatoirement un $r' \in \mathbb{Z}_p$ et pose implicitement $r = br' - ab$, puis définit $D_1 = g^{(\alpha/\beta)+r} = g^{d+ab+br'-ab} = g^{d+br'} = g^d B^{r'}$ et sélectionne aléatoirement un $t' \in \mathbb{Z}_p$, afin de définir implicitement $t = a + \frac{t'}{b}$. Ensuite, il calcule $D_2 = g^r h^t = g^{br'-ab} g^{b(a+\frac{t'}{b})} = g^{br'+t'} = B^{r'} g^{t'}$. Pour chaque $j \in S$, $\mathcal{B}$ sélectionne aléatoirement $r'_j \in \mathbb{Z}_p$ et implicitement calcul $H(j)^{r_j} = g^{ab+r'_j}$ et calcul $D_j = g^{br'-ab} g^{ab+r'_j} = B^{r'} g^{r'_j}$ et pose $D'_j = g^{r'_j}$. Enfin, $\mathcal{B}$ transmet la clé secrète simulée $SK' = \{D_1, D_2, \forall j \in S : D_j, D'_j\}$ à $\mathcal{A}$.

— **Challenge :** $\mathcal{A}$ soumet deux messages m_0 et m_1 de même longueur à $\mathcal{B}$. $\mathcal{B}$ sélectionne de façon aléatoire $\gamma \in \{0, 1\}$ et génère un texte crypté de défi CT^* correspondant à m_γ. $\mathcal{B}$ pose implicitement $t = c$ (nous rappelons que $\mathcal{B}$ ne connait pas la valeur de c), puis calcul $C_1^* = m_\gamma e(g,g)^{c\frac{\alpha}{\beta}} = m_\gamma e(g,g)^{c(d+ab)} = m_\gamma e(g,g)^{cd+abc} = m_\gamma e(g,g)^{cd} e(g,g)^{abc}$ $= m_\gamma e(C, g^d) Z$ avec $Z = e(g,g)^{abc}$ si $\mu = 0$, sinon Z est un élément aléatoire de $\mathbb{G}_T$. $\mathcal{B}$ choisit $s' \in \mathbb{Z}_p$ de façon aléatoire pour évaluer implicitement $s = s' - c$ puis calcul $C_2'^* = C_2 h^s = h^{t+s} = (g^b)^{c+s} = g^{bc+bs} = g^{bc+b(s'-c)} = g^{bs'} = B^{s'}$ et $C_3^* = g^t = g^c = C$. $C_3'^*$ est donnée alors par $C_3^* = g^{t+s} = g^{c+s'-c} = g^{s'}$. Maintenant, pour chaque attribut i de la structure P^*, $\mathcal{B}$ choisi aléatoirement un secret s_i et construit le partage de secret à travers $P*$ et calcule $\forall i \in \mathbb{P}^* : c_i^* = g^{s_i}, c_i'^* = Hattrib(i)^{s_i}$. Enfin $\mathcal{B}$ transmet le chiffré CT^* à $\mathcal{A}$ et la valeur de $e(C, g^d)$. Ce dernier élément n'apporte aucune information à $\mathcal{A}$ mais lui permettra d'isoler la partie $m_\gamma Z$ de C_1^*.

— **Phase 2 :** Identique à la phase 1.

— **Décision :** $\mathcal{A}$ génère une estimation γ' de γ. Si $\gamma' = \gamma$ alors $\mathcal{B}$ génère $\mu' = 0$ pour indiquer qu'il croit avoir reçu $(A, B, C, Z) = (g^a, g^b, g^c, e(g,g)^{abc})$ par le challenger $\mathcal{C}$. Sinon, il produira $\mu' = 1$, ce qui signifie qu'il a reçu un élément aléatoire Z.

Pour finaliser la démonstration de sécurité du schéma proposé, nous estimons l'avantage du simulateur $\mathcal{B}$.

Lorsque $\mu = 1$, le texte chiffré est un élément aléatoire du point de vue de l'adversaire $\mathcal{A}$ et ne donne aucune information sur γ. Dans ce cas, $Pr[\gamma' = \gamma \mid \mu = 1] = \frac{1}{2}$ (l'avantage de $\mathcal{A}$ est le même s'il tire au hasard γ') et c'est le même avantage qu'aura $\mathcal{B}$. En effet, la

décision de $\mathcal{B}$ va être basée sur le résultat de $\mathcal{A}$, si $\gamma' \neq \gamma$, $\mathcal{B}$ va estimé que $\mu' = 1$ et si $\gamma' = \gamma$, $\mathcal{B}$ va estimé que $\mu' = 0$.

Lorsque $\mu = 0$, $\mathcal{A}$ a l'avantage ε. Donc, par définition, nous avons $Pr[\gamma' = \gamma \mid \mu = 0] = \varepsilon + \frac{1}{2}$. Puisque $\mathcal{B}$ estime $\mu' = 0$ quand $\gamma' = \gamma$ alors $Pr[\mu' = \mu \mid \mu = 0] = \varepsilon + \frac{1}{2}$. Enfin, l'avantage général de $\mathcal{B}$ est le suivant :

$$\begin{aligned} Adv_{\mathcal{B}} &= \frac{1}{2}[Pr[\mu' = \mu \mid \mu = 1]] + \frac{1}{2}[Pr[\mu' = \mu \mid \mu = 0]] - \frac{1}{2} \\ &= \frac{1}{2}\frac{1}{2} + \frac{1}{2}(\epsilon + \frac{1}{2}) - \frac{1}{2} \\ &= \frac{\epsilon}{2} \end{aligned}$$

Par conséquent, comme ϵ est supposé non négligeable, $\frac{\epsilon}{2}$ est également non négligeable, ce qui prouve le théorème 2.

le cnam

Youcef OULD YAHIA

Proposition d'un modèle de sécurité pour la protection de données personnelles dans les systèmes basés sur l'internet des objets

le cnam

Abstract :
Internet of Things (IoT) and IT service outsourcing technologies have led to the emergence of new threats to users' privacy. However, the implementation of traditional security measures on IoT equipment is a first challenge due to capacity limitations. On the other hand, the offloading of data processing and storage poses the problem of trust in service providers.
In this context, we have proposed an encryption solution that provides owner-centric data protection adapted to the constraining environment of IoT. This model is based on attribute-based encryption with secure offloading capability and Blockchain technology. Then, in response to the issue of trust and service selection, we explored the possibilities offered by artificial intelligence tools. To do this, we proposed a collaborative filtering model based on Kohonen maps and an efficient solution to detect the untrusted users.

Keywords :
Personal data, Internet of Things, Attributes-based encryption, Trust management, service offloading.

Résumé :
Les technologies de l'Internet des objets (IdO) et de l'externalisation des services informatiques ont conduit à l'émergence de nouvelles menace sur la vie privée des utilisateurs. Cependant, l'implémentation des moyens de sécurité traditionnels sur les équipements de l'IdO constitue un premier défi lié aux limites de capacités. D'un autre coté, la délégation du traitement et du stockage des données, nous pose le problème de confiance envers les fournisseurs de service.
Dans ce contexte, nous avons proposé une solution de chiffrement qui assure une protection de données centrée sur leurs propriétaires et adaptée à l'environnement contraignant des objets connectés. Ce modèle se base sur le chiffrement par attributs avec externalisation sécurisée et la technologie de la Blockchain . Ensuite, en réponse à la problématique de la confiance et de la sélection du service, nous avons exploré les possibilités offertes par les outils de l'intelligence artificielle. Pour ce-faire, nous avons, proposé un modèle de filtrage collaboratif basé sur les cartes de Kohonen avec une solution pour détecter les utilisateurs non fiables.

Mots clés :
Données personnelles, Internet des Objets, Chiffrement basé sur les attributs, Gestion de la confiance, Externalisation des traitements.