République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A/Mira de Béjaia
Faculté des Sciences Exactes
Département d'Informatique

Mémoire de Fin de Cycle

En vue de l'obtention du diplôme de Master Recherche en Informatique

Option :
Réseaux et Systèmes Distribués

Thème

Le Contrôle d'Accès au Big Data. Cas d'Etude : Internet des Objets

Présenté par :

Mlle ABBAS Célia
Mlle MOUZAIA Chahinas épouse REDJDAL

Devant le jury composé de :

Président	**M. AISSANI Sofiane**	**Maitre Assistant A**	**U. A/Mira Béjaia.**
Examinateur	**M. ZAMOUDJ Salah**	**Maitre Assistant B**	**U. A/Mira Béjaia.**
Examinateur	**M. BOUCHEBAH Fatah**	**Doctorant "LMD"**	**U. A/Mira Béjaia.**
Encadreur	**M. OMAR Mawloud**	**Maître de conf. A**	**U. A/Mira Béjaia.**

Promotion 2016/2017.

⁎ *Remerciements* ⁎

Le grand remerciement à Dieu le miséricordieux, Qui nous a donné le courage et la patience pour mener à bien ce travail. Nos vifs remerciements sont adressés à toutes les personnes qui ont contribuées de manière directe ou indirecte à l'aboutissement de ce travail.

En premier lieu, nos gratitude à notre encadreur Dr OMAR Mawloud qui nous a confié ce thème d'actualité. Nous apprécions son enthousiasme, sa gentillesse, ses conseils, ses orientations remarquables, son précieux suivi tout au long de la réalisation de ce travail, l'intérêt qu'il nous a apporté pour l'accomplissement de ce projet de fin de cycle et surtout pour sa grande aide et ses qualités humaines et pédagogiques.

On tient à remercier également tous les membres du jury, M. AISSANI Sofiane en tant que président, M. ZAMOUDJ Salah , M. BOUCHEBAH Fatah comme examinateurs qui ont accepté de juger ce travail.

Nos remerciements s'étendent à tous nos enseignants du département d'Informatique de l'Université Abderrahmane Mira de Béjaïa.

Un grand Merci à tous nos proches qui nous ont apporté leurs soutiens durant l'élaboration de ce travail.

⁕ *Dédicaces* ⁕

Je dédie ce modeste travail :

A mes très chers parents dont les conseilles et sacrifices ont permis la poursuite de mes études et qui m'ont élevé, formé, encouragé et soutenu durant toute ma vie, j'espère avoir été a la hauteur de tous ce qu'ils m'ont apporté

A mes frères et soeurs qui me sont très chers.

A mon très cher mari Salim, qui est mon modèle de courage et de sacrifices et qui m'apporte chaque seconde un soutien indéniable.

A toute ma famille et ma belle famille du plus grand au plus petit.

A ma binome Célia.

A mes amis.

Mlle ***MOUZAIA Chahinas***

⁜ *Dédicaces* ⁜

Je dédie ce modeste travail :
A mes très chers parents dont les conseilles et sacrifices m'ont permit la poursuite de mes études et qui m'ont élevé, formé, encouragé et soutenu durant toute ma vie, j'espère avoir été a la hauteur de tous ce qu'ils m'ont apporté

A mon frère Yanis et mes soeurs Siham et Yasmine qui m'apportent chaque jour un soutien indéniable.

A ma binome Chahinas.

A mes chers amis Lilia, Kamilia, Rabah, Nabil.

A toute ma famille du plus grand au plus petit.

M[lle] ***ABBAS Célia***

Table des matières

Table des figures

Liste des tableaux

Liste des Acronymes

AA	Attribute Authority
ABAC	Attribute-Based Access Control
ABE	Attribute Based Encryption
ABF	Attribute Bloom Filter
ACP	Access Control Policy
AES	Advanced Encryption Standard
CA	Certification Authority
CBAC	Context-Based Access Control
C-CP-ARBE	Collaborative-Ciphertext Policy-Attribute Role Based Encryption
CP-ABE	Ciphertext-Policy Attribute-Based Encryption
CT	CipherText
CTA	CipherText Authenticated
DAC	Discretionary Access Control
ECT	Encrypted CipherText
EL	Extended Leaf
GID	Global IDentifier
GRP	GroupRole Parameter
HDFS	Hadoop Distributed File System
IoT	Internet of Things
MAC	Mandatory Access Control
MAC	Message Authentication Codes
Onto-ACM	Ontology based Access Control Model
OWL	Web Ontology Language

PHR	Personal Health Record
PK	Public Key
RBAC	Role-Based Access Control
RDF	Resource Description Framework
RFID	Radio Frequence Identification
SCT	Sealed CipherText
SK	Secret Key
SS	Secret Seal
TESLA	Timed Efficient Streaming Loss-tolerant Authentication
UASK	User Attribute Secret Key
UDK	User Decryption Key
UGPK	User Global Public Key
UGSK	User Global Secret Key
UKGen	Update Key Generation

Introduction générale

L'Internet est devenu une nécessité dans la vie moderne, et pour cela les chercheurs on essayé d'améliorer la vie quotidienne où l'homme n'aura plus besoin d'avoir recours a l'intervention humaine pour fournir des services. Le terme Internet des Objets a été introduit pour la première fois en 1999 par Kevin Ashton [3]. Cette technologie consiste principalement à connecter des objets physiques du quotidien (téléphone, montre, etc.) au réseau Internet afin d'offrir des services, collecter des informations de manière autonome grâce à l'intégration de capteurs, d'actionneurs [19]. En effet, l'Internet des Objets a permis la facilité de la vie quotidienne à travers des applications dans diffèrent domaines tels que : l'aéronautique, la santé, les villes intelligentes, le transport, etc.

En effet, tous les jours, de plus en plus de périphériques sont équipés de capteurs pour recueillir des données et les personnes qui les utilisent dans différents domaines (médicale, militaire, environnement, etc.), d'où vient la notion du big data qui est un ensemble de données numériques appelé aussi données massives. A chaque seconde, des capteurs génèrent une masse importante de données hétérogènes ce qui fait que contrôler l'accès et la bonne gestion de ces données massives générées par ces appareils intelligents, aussi celles produites par l'homme est une question stimulante qui n'a pas encore de solution définitive, dû au volume et à la diversité des données sensées être confidentielles.

Dans ce contexte, de nombreux travaux de recherche portent sur des solutions basées sur la sémantique pour le contrôle d'accès en utilisant l'analyse sémantique des informations de contexte sur l'utilisateur pour prendre des décisions sur les ressources appropriées, la politique de sécurité et le contrôle d'accès. D'autres travaux se sont basés sur la cryptographie à base d'attributs. Cette dernière se repose sur le cryptage à clé publique où la clé secrète d'un utilisateur et la donnée chiffrée dépendent des attributs.

Notre contribution consiste à la proposition d'une solution pour le contrôle d'accès au big data dans l'Internet des Objets. Nous nous sommes inspirés des solutions proposées et de leurs points forts. Plus exactement, nous proposons une architecture qui assure le contrôle d'accès au big data basée sur la cryptographie à base d'attributs, ainsi l'authentification des données en se basant sur le principe de μTESLA, et tout en s'adaptant au contexte de l'Internet des objets.

Ce mémoire est organisé en quatre chapitres. Le premier chapitre est consacré à l'Internet des Objets et le big data, qui portera sur quelques définitions, ainsi que les domaines d'applications et l'architecture, puis on présentera quelques défis imposés par big data et la sécurité de l'Internet des Objets. Dans le deuxième chapitre, nous discuterons certains travaux de recherche concernant le contrôle d'accès au big data. Dans le troisième chapitre, nous présenterons en détail l'architecture que nous proposons pour le contrôle d'accès, ainsi que son principe de fonctionnement et pour terminer une analyse de sécurité est réalisée. Pour prouver l'efficacité de notre architecture, nous exposons dans le quatrième chapitre les résultats obtenus suite à l'évaluation des performances de notre proposition. Nous clôturons ce mémoire par une conclusion générale et des perspectives.

Chapitre 1

L'Internet des Objets et le big data

1.1 Introduction

L'essor des nouvelles technologies ainsi que les progrès survenus dans l'Internet ont permis l'apparition de ce qu'on appelle Internet des Objets, IoT est formé en connectant des milliards d'être humain et des milliards d'objets. Dans ce chapitre, nous présenterons l'Internet des Objets ainsi que le big data et la relation qui les relie, puis les différents domaines d'application de l'Internet des Objets et leur classification. Ensuite, nous passerons à l'architecture de l'Internet des Objets, et après nous parlerons sur les défis imposés par le big data. Par la suite, nous expliquons brièvement la sécurité dans l'IoT et on terminera par une conclusion de ce chapitre.

1.2 Définitions

Définition 1.2.1. L'Internet des Objets est "*une infrastructure dynamique d'un réseau global. Ce réseau a des capacités d'auto-configuration basée sur des standards et des protocoles de communication interopérables. Dans ce réseau, les objets ont des identités, des attributs physiques, des personnalités virtuelles et des interfaces intelligentes, et ils sont intégrés au réseau d'une façon transparente*"[12].

Définition 1.2.2. Le big data appelé aussi données massives, est un ensemble de données numériques produites par l'utilisation de nouvelles technologies qui sont caractérisées par les 5Vs et devient difficile à traiter et a analyser en utilisant les outils classiques de gestion de base de données. La notion de 5Vs vient de Volume, Variété, Véracité, Valeur et Vitesse.

1. **Volume :** Le premier V correspond au volume. Il représente la quantité massives des données générées.

2. **Variété :** Le deuxième V est la variété, le type et le format de données générées sont différents.
3. **Vitesse :** La vitesse représente la rapidité de génération et de partage des données.

4. **Véracité :** La véracité correspond à la fiabilité des données. La précision et la qualité des données sont très importantes. Donc il faut s'assurer que les données du big data sont vraies pour pouvoir prendre des décisions.

5. **Valeur :** La valeur est le V le plus important, l'objectif est de créer des valeur pour les entreprises et les clients en transformant toutes les données en valeurs exploitables.

1.3 Big data et IoT

L'IoT est intimement lié à la notion de big data dont les objets génèrent une grande quantité et une variété de données en temps réel ces données dites big data[15]. Donc, chacun a besoin de l'autre pour le rendre utile. Il n'y a pas d'IoT sans big data, et big data atteint la position la plus élevée lorsqu'il est utilisé pour IoT.

1.4 Domaines d'application de IoT

L'IoT couvre un large éventail d'applications et touche à un grand nombre de domaines dans notre vie quotidienne (comme illustré par la figure 1.1), tels que le domaine aéronautique, les soins de santé, le transport et bien d'autres encore, qui permettra l'émergence des espaces intelligents.

1.4.1 Domotique

La domotique est le domaine technologique qui traite de l'automatisation du domicile, qui consiste à mettre en place des réseaux reliant différents objets connectés. Ainsi, elle regroupe tout un ensemble de services permettant l'intégration des technologies modernes dans la maison qui permettent de centraliser le contrôle des différents systèmes de la maison afin de répondre aux besoin du confort, sécurité et de communication [2].

1.4.2 Industrie

Dans l'industrie l'IoT permettra un suivi total des produits, de la chaîne de production, jusqu'à la chaîne logistique et de distribution en supervisant les conditions d'approvisionnement. Cette traçabilité de bout en bout facilitera la lutte contre la contrefaçon, la fraude et les crimes économiques transfrontaliers [4].

1.4.3 Agriculture

Dans ce domaine, des réseaux de capteurs interconnectés à l'IoT peuvent être utilisés pour la supervision de l'environnement. Ceci permettra une meilleure aide à la décision en agriculture, notamment pour optimiser l'eau d'irrigation, l'usage des intrants, et la planification de travaux agricoles. Ces réseaux peuvent être aussi utilisés pour lutter contre la pollution de l'air, du sol et des eaux et améliorer la qualité de l'environnement en général [4].

1.4.4 Ville

Les villes intelligentes permettent d'améliorer la qualité de vie de la municipalité, en assurant l'optimisation du remplissage des parkings dans la ville grâce à l'identification des places libre en temps réel, l'éclairage publique intelligent qui varie en fonction de l'intensité lumineuse ou des déplacement également. La gestion du trafic est facilitée par une vision fine et en temps réel des flux [19]

1.4.5 Transport et mobilité

Le transport intelligent et la mobilité représentent la connexion de véhicules à Internet pour créer de nouvelles possibilités et des applications qui apportent nouvelles fonctionnalités aux individus et faisant le transport plus sûr et plus facile. Internet des Objets représente une partie inhérente du contrôle de véhicule et du système de direction, permet de nouveaux scénarios de transport et permet la régulation du trafic et le contrôle [16].

1.4.6 Santé

Dans le domaine de la santé, l'IoT permettra le déploiement de réseaux personnels pour le contrôle et le suivi des signes cliniques, notamment pour des personnes âgées. Ceci permettra ainsi de faciliter la télésurveillance des patients à domiciles, et apporter des solutions pour l'autonomie des personnes à mobilité réduite [4].

1.4.7 Aérospatial et aviation

Dans ce domaine l'IoT vise à porter l'innovation aéronautique pour réduire les coûts de développement et optimiser la performance des avions. De plus améliorer la sécurité des services, en assurant l'identification des produits et des éléments contrefaits grâce à l'élaboration des pièces critiques des appareils, cela par le bais d'implantation de tag RFID [19].

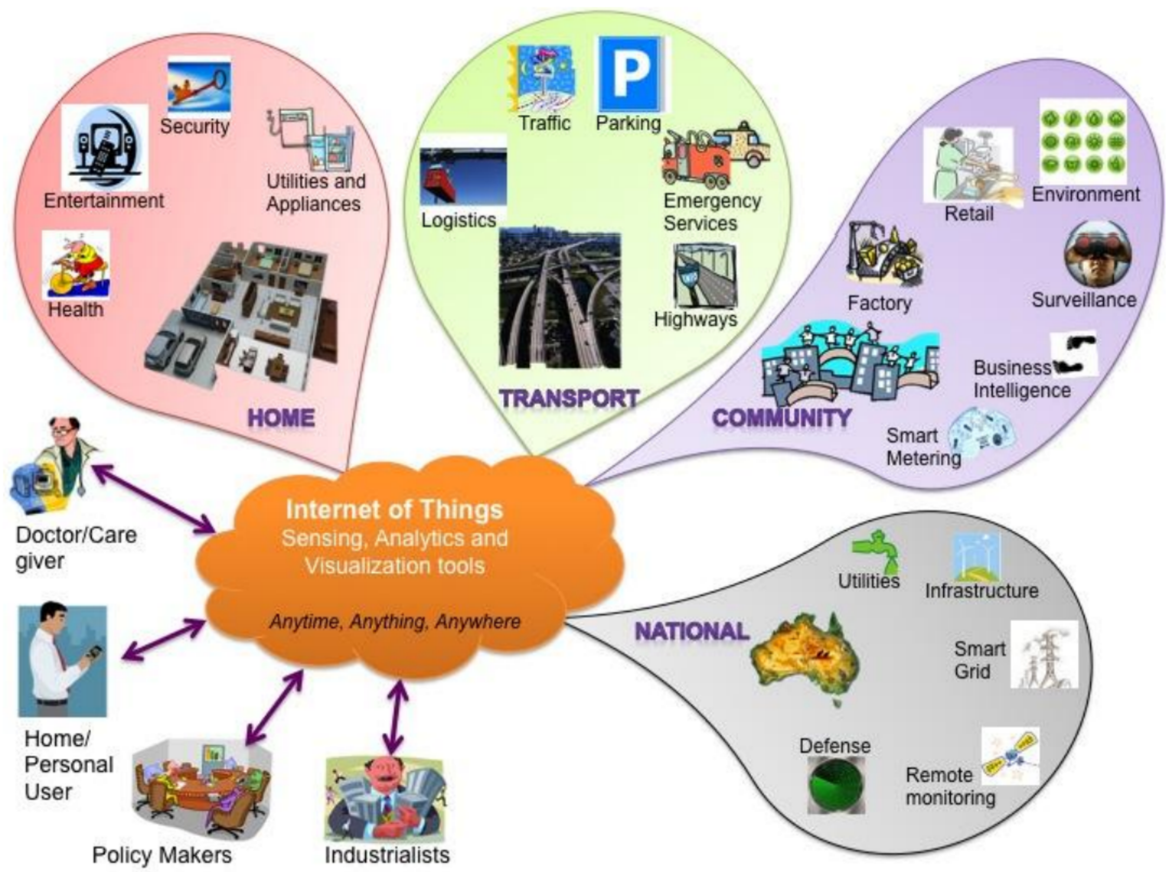

Figure 1.1 – Domaines d'application de l'Internet des Objets [8].

1.5 Classement des applications de l'IoT

K. Lasse Lueth [10] a choisit trois critères à mesurer afin de classifier les applications de l'IoT qui sont : Ce que les gens cherchent sur Google, de quoi les gens parlent sur Twiter et de quoi les gens écrivent sur LinkedIn. Le plus haut score a reçu une estimation de 100%, et les autres applications de l'Internet des objets ont été classées avec un pourcentage qui représente la relation avec le plus haut score, comme illustrer sous la figure 1.2.

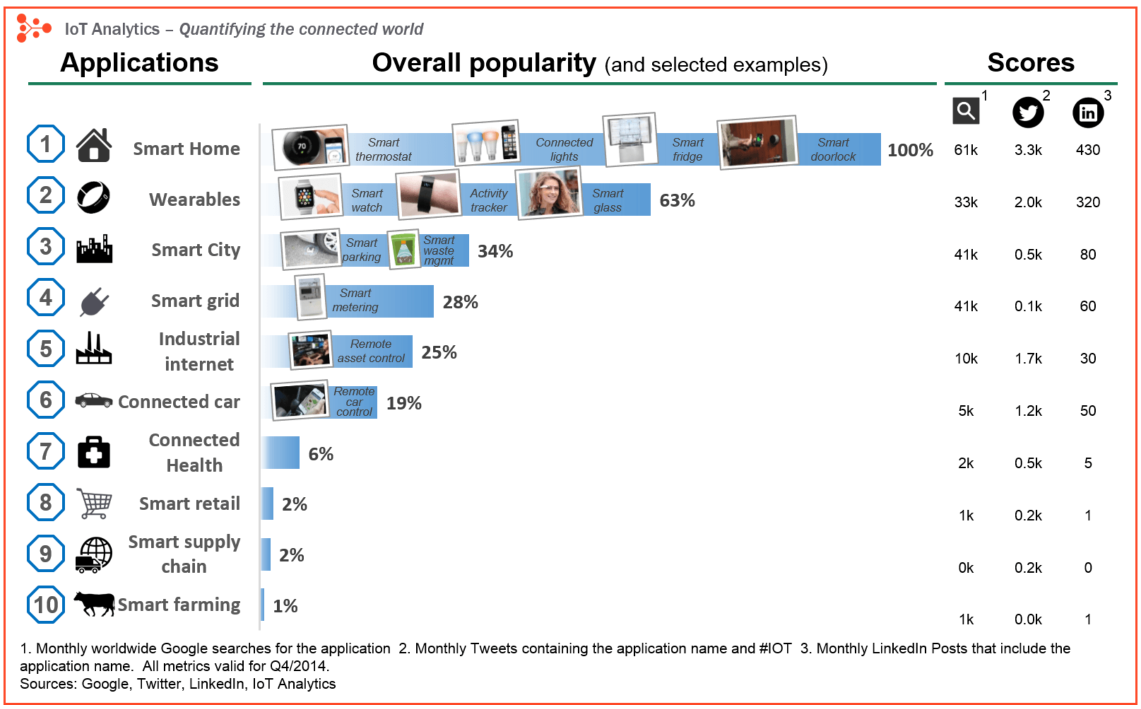

FIGURE 1.2 – Classification des applications de l'Internet des Objets [10].

1.6 Architecture générale de IoT

D'aprés la figure 1.3, l'architecture de l'IoT est composée de cinq principales couches [21] :

✠ **Couche 1 :** Couche perception ou objets, qui représente les objets physiques de l'IoT visant à collecter et traiter l'information. Elle fournit des différentes fonctionnalités telles que l'interrogation de l'emplacement, la température, le poids, le mouvement, les vibrations, l'accélération, l'humidité, etc. Ses tâches principales sont de collecter et transférer les données à la couche supérieure via des canaux sécurisés.

✠ **Couche 2 :** Couche transport appelée aussi couche réseau. Sa fonction principale est le transfert de l'information de la couche perception vers la couche traitement en utilisant des canaux de transport fiable ou peu fiable. Cette couche utilise diverses technologies pour le transport de l'information tel que la 3G, Wifi, Bluetooth, infrarouge, ZigBee, etc.

✠ **Couche 3 :** Couche traitement appellée aussi Middleware. Les objets de l'IoT offrent de différents types de services et cette couche est responsable de la gestion de ses services. La couche traitement stocke, analyse et traite les informations relatives aux éléments reçus de la couche 2 et prend une décision automatique en fonction des résultats.

✠ **Couche 4 :** Couche application fournit des services demandés par des clients en se basant sur les services et l'information d'objet traité dans la couche 3. Les applications mises en œuvre par IoT peuvent être des maisons intelligentes, transport intelligent, etc.

✠ **Couche 5 :** Couche commerce est responsable de la gestion globale des applications IoT, elle crée des modèles de gestion, des graphes, des organigrammes, etc. en fonction des données reçues de la couche application. Sur la base de l'analyse des résultats, cette couche aidera à déterminer les stratégies commerciales.

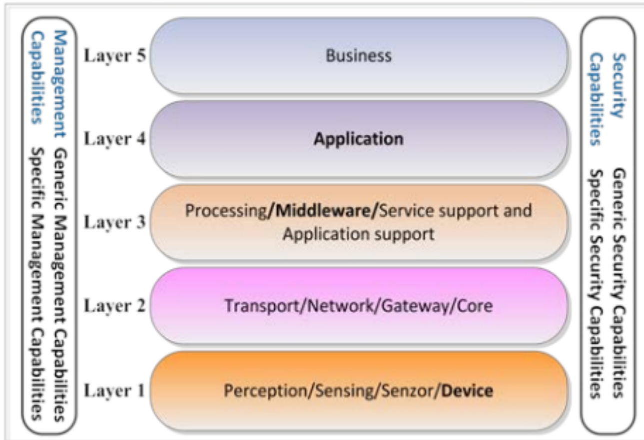

FIGURE 1.3 – Architecture de l'Internet des Objets [21].

1.7 Défis imposés par big data

Les défis de big data ont été divisés en quatre aspects principaux employés dans la division créée par le big data Working Group dans Cloud Security Alliance organisation : la sécurité de l'infrastructure, la privacy des données, la gestion des données et l'intégrité et la sécurité réactive [1]. La figure 1.4 représente les quatre catégories de défis du big data.

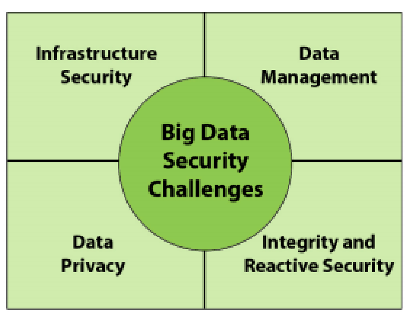

FIGURE 1.4 – Catégorie des défis du big data [11].

Afin de sécuriser l'infrastructure des systèmes big data, les calculs distribués et les serveurs de stockage de données doivent être sécurisés. Pour sécuriser les données elles-mêmes, la diffusion de l'information doit être la protection de la vie privée, et les données sensibles doivent être protégées par l'utilisation de la cryptographie et du contrôle d'accès. La gestion de l'énorme volume de données nécessite des solutions évolutives et distribuées pour sécuriser les magasins de données et permettre des audits efficaces et la provenance des données. Enfin, les données de diffusion émergeant de différents points finaux doivent être vérifiées pour l'intégrité et peuvent être utilisées pour effectuer des analyses en temps réel pour des incidents de sécurité afin d'assurer la santé de l'infrastructure [1].

Dans ce qui suit nous allons se focaliser sur l'étude d'un défi de la catégorie privacy des données, qui est le contrôle d'accès.
La privacy des données est sans doute le sujet de ce que les gens ordinaires sont les plus concernés, mais il devrait également être l'une des plus grandes préoccupations pour les organisations qui utilisent les techniques big data. Un système big data contient généralement une quantité énorme d'informations personnelles que les organisations utilisent pour obtenir un avantage de ces données. Cependant, nous devrions demander où est la limite concernant l'utilisation de cette information, pour cela des mécanismes de contrôle d'accès ont été conçus pour répondre à cette question.

* **Le contrôle d'accès** est l'une des techniques basiques utilisées pour assurer la sécurité d'un système. Son objectif est de restreindre l'accès des utilisateurs non désirés au système. Dans le cas de big data, le problème de contrôle d'accès est lié au fait qu'il n'existe que des formes basiques de contrôle d'accès. Afin de résoudre ce problème, certains auteurs proposent un cadre qui supporte l'intégration des fonctions de contrôle d'accès [11].

1.8 Sécurité de l'Internet des Objets

L'IoT est une technologie caractérisée par une forte ubiquité dans le monde physique et une omniprésence autour de ses usagers. Les diverses applications potentielles de l'IoT, l'hétérogénéité de ses technologies habilitantes et sa forte dimension humaine et socioéconomique rendent sa sécurité un sujet difficile et complexe. En plus des problèmes de sécurité des technologies qui le constitueront, l'IoT accentue les problèmes de sécurité des personnes qui l'utiliseront, et fait émerger de nouveaux problèmes liés à la sécurité des systèmes sous son contrôle. Comme illustré sur la figure 1.5, la sécurité et la confidentialité dans l'IoT peut être abordée de trois angles complémentaires qui reflètent ses dimensions technologique, humaine et systémique. La protection de la technologie concerne en premier lieu la sécurité des données, des communications et des infrastructures réseaux. Cette protection est nécessaire pour contrarier les attaques classiques et futures sur l'intégrité, l'authenticité et la confidentialité des données, ainsi que les attaques sur les infrastructures réseaux et leurs fonctionnalités. La protection des personnes concernera la protection de la vie privée des usagers qui nécessite, en plus des solutions technologiques, une régulation appropriée qui établit les responsabilités en cas de litiges. La protection des systèmes interconnectés et hébergeant les objets de l'IoT, concernera la protection des objets eux-mêmes livrés à ces systèmes et les processus qu'ils contrôleront [4].

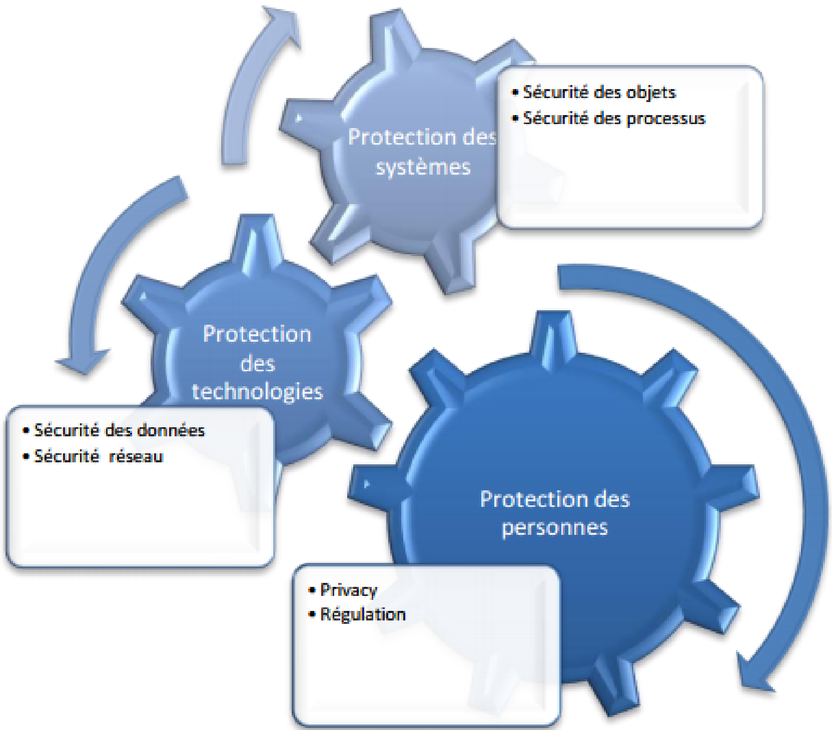

FIGURE 1.5 – Sécurité et Privacy de l'Internet des Objets [4].

1.9 Conclusion

L'IoT représente une technologie prometteuse qui permet une évolution majeure qui s'inscrit dans la continuité des développements récents de différentes technologies de l'information. Cependant l'IoT suscitera des questions stimulantes, qui concerneront directement la sécurité de big data (les données massives générées par les objets communiquant) qui contient des milliers de données sensées être confidentielles telles que des informations liées à la vie privée des personnes comme leurs déplacements et états de santé. Le chapitre suivant sera consacré à un état de l'art sur les mécanismes de contrôle d'accès au big data.

Chapitre 2

État de l'art sur les modèles de contrôle d'accès au big data

2.1 Introduction

La recherche d'un modèle de contrôle d'accès dans un domaine en pleine croissance qui est le " big data" est en émergence grâce à l'Internet des Objets, dont des milliers d'objets dans le monde se voient connecter les uns avec les autres et avec le serveur du Cloud. Donc assurer le contrôle d'accès et la bonne gestion de ces données massives générées par ces appareils intelligents, aussi celles produites par l'homme est une question stimulante qui n'est pas encore résolue, dû au volume et à la diversité des données sensées être confidentielles. Ce chapitre est consacré à l'étude critique et la présentation des modèles de contrôle d'accès existants. Pour cela nous commencerons tout d'abord par présenter les critères d'analyse, suivi par une classification des solutions étudiées, puis nous présenterons et discuterons chaque solution. Enfin, nous conclurons ce chapitre par une comparaison des travaux analysés.

2.2 Critères d'évaluation des solutions existantes

Pour une meilleure évaluation des travaux de recherche étudiés, nous avons établi certains critères jugés pertinents, en tenant compte des besoins et contraintes liés aux big data. Nous nous intéresserons à la sécurité en termes de résistance aux attaques, la scalabilité, la notion d’alèrte ainsi qu’à la cohérence de la politique de sécurité.

2.2.1 Sécurité

La sécurité informatique est l’ensemble des moyens mis en œuvre pour réduire les vulnérabilités d’un système contre les menaces. En raison de la variété des données sensibles échangées par les utilisateurs, un modèle de sécurité développé pour mieux gérer l’accèes au big data doit être résistant aux attaques et il doit répondre aux principales exigences de sécurité.

2.2.2 Scalabilité

La scalabilité dans Internet des Objets désigne la capacité de s’adapter aux nouveaux dispositifs, services et fonctions sans nuire la performance des services existants. Puisque le volume des données dans l’Internet des Objets augmente du jour en jour, nous devrons prendre en considération ce critère pour avoir un système efficace.

2.2.3 Notion d’alèrte

Une alèrte est un message utilisé pour informer le responsable de sécurité d’une attaque. On trouve les faux positifs qui sont des alèrtes générées en absence d’attaques (fausses alertes) et les faux négatifs sont des alèrtes manquantes en présence d’attaques.

2.2.4 Cohérence

Nous définissons une politique de sécurité cohérente, cette propriété est d’autant plus importante pour contrôler l’accès au big data, où les utilisateurs ne peuvent pas se trouver dans la situation dans laquelle ils auraient simultanément la permission et l’interdiction d’effectuer une action sur une donnée, ni de situation dans laquelle un utilisateur aurait simultanément l'obligation et l'interdiction d'effectuer une action sur une donnée.

2.3 Classification des travaux étudiés

Après avoir examiné les travaux récoltés, une classification est nécessaire afin d'ordonner les différentes approches suivies. La figure 2.1 représente notre classification des différentes solutions pour le problème de contrôle d'accès dans le big data.

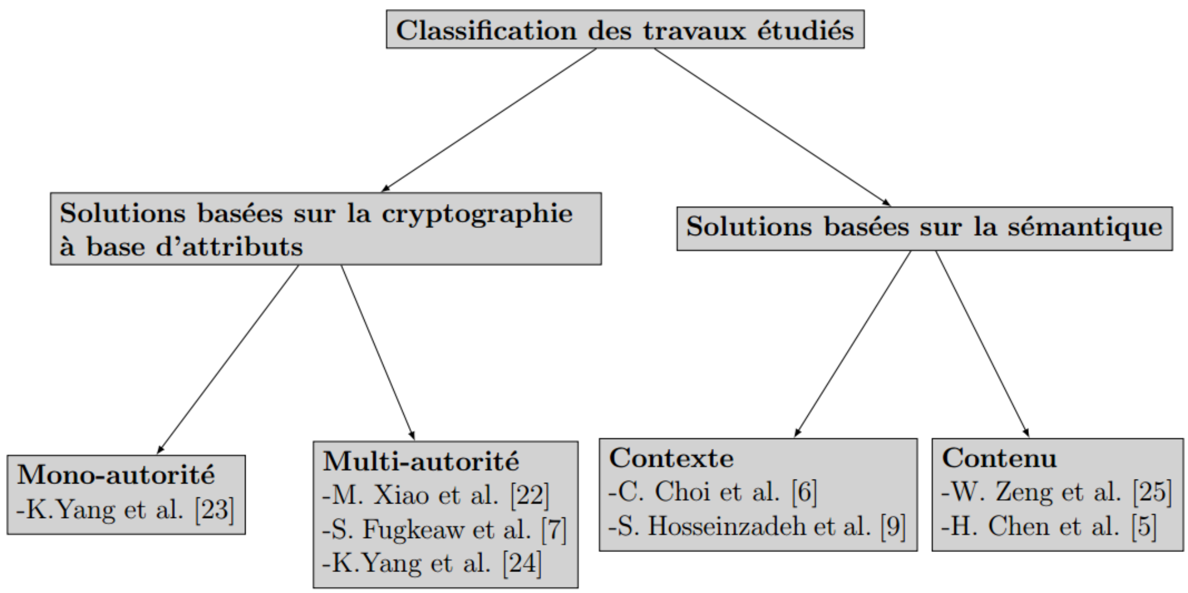

FIGURE 2.1 – Classification des travaux étudiés.

2.3.1 Solutions basées sur la sémantique

2.3.1.1 Solutions basées sur le contexte

1. Ontology-based access control model for security policy reasoning in Cloud computing

Choi et al. [6] ont proposé un modèle de contrôle d'accès à base d'ontologie (Onto-ACM), qui est un modèle d'analyse sémantique qui peut adresser la différence dans le contrôle d'accès autorisé entre les prestataires de services et les utilisateurs. Onto-ACM est un modèle intelligent qui permet d'accéder proactivement au niveau d'accès de la ressource à base du raisonnement ontologique (le raisonnement du contexte comme le contexte, le niveau d'autorisation, condition sur l'autorisation, but et chaque politique pour l'administrateur et l'utilisateur) et l'analyse sémantique.

Le modèle proposé fonctionne comme suit. L'utilisateur envoie les informations contextuelles au processus d'analyse de contexte, qui lui-même envoie une demande pour récupérer les informations de l'utilisateur, le moteur d'analyse de contexte recueille ces informations et décide de l'authentification de l'utilisateur. Après la connexion de l'utilisateur, une demande des informations d'accès est envoyée par le processus d'analyse de contexte au module de contrôle d'accès qui gère l'autorisation des utilisateurs et l'analyse du contexte d'accès, ainsi un rôle est assigné à l'utilisateur. Le processus d'analyse de contexte reçoit le rôle de l'utilisateur et envoie par la suite une demande des ressources d'information au gestionnaire d'ontologie qui se trouve au module de contrôle d'accès qui contrôle des droits d'accès par les ressources d'information par les inférences basées sur le rôle de l'utilisateur actif et celui du contexte, puis les droits d'accès sont attribués aux différentes informations contextuelles. Après l'identification des droits d'accès de l'utilisateur, le processus d'analyse de contexte envoie une requête de demande d'inférence sur le processus de sécurité, le moteur d'inférence récupère cette requête et exécute le raisonnement de la politique de contrôle d'accès après l'avoir contrôlé par le contrôleur de politique, ensuite une description de l'abstraction est effectuée pour fournir des informations contextuelles pour l'analyse sémantique à partir du capteur et de l'équipement. Le processus d'analyse de contexte reçoit les résultats de l'analyse sémantique, et envoie une réponse à l'utilisateur que le processus qu'il désire exécute est sécurisé, donc il peut accéder.

L'ontologie utilisée dans le modèle proposé permet de faire une analyse sémantique des droits d'accès pour détecter leurs conformités ou non avec une politique de sécurité bien défini, ce qui fait que la politique d'accès proposée est cohérente, mais la scalabilité n'est pas assurée en raison de la présence d'un seul administrateur dans le système qui contrôle l'accès de différents utilisateurs, aussi ce modèle est vulnérable aux attaques de rejeu donc la sécurité n'est pas assurée.

2. A Semantic Security Framework and Context-Aware Role-Based Access Control Ontology for Smart Spaces

Hosseinzadeh et al. [9] ont proposé un modèle de contrôle d'accès comme illustré sous la figure 2.2, qui est une combinaison entre le contrôle d'accès basé sur le rôle (RBAC) et de contrôle d'accès basé sur le contexte (CBAC) modélisé avec des techniques ontologiques et une langue d'ontologie web (OWL), qui contrôle l'accès des utilisateurs au système dans la conformité à leur rôle dans le système et l'information du contexte actuel. Les auteurs ont choisi RDF (Resource Description Framework), qui permet de créer des vocabulaires de métadonnées à travers la définition de classes et de propriétés. L'architecture proposée est composée d'un répertoire RDF pour vérifier l'authenticité, l'autorité et le droit d'accès du demandeur avant d'accéder au répertoire de données. L'ontologie utilisée est composée des composants différents, en incluant des classes, des propriétés d'objet, des propriétés de données et des individus. Les individus (représentés par les diamants) sont les objets. Les classes (représenté par les cercles) sont le concept abstrait utilisé pour grouper des individus d'un certain type. Les individus sont raccordés l'un à l'autre via les propriétés, qui sont classifiées comme les propriétés d'objet faisant la carte de deux individus ensemble et de propriétés de données faisant la carte d'un individu à une valeur de données. Les règles de contrôle d'accès incluent deux ensembles de règles : les règles conçues par l'administrateur et les règles définies par l'utilisateur dans le but de protection d'intimité. Le modèle fonctionne comme suit : quand une demande arrive le moteur d'authentification vérifie que le demandeur est authentique, si positive la demande est envoyée au moteur de contrôle d'accès là où s'exécutent les règles et il est vérifié si le demandeur a le droit d'exécuter l'action demandée. Le rondin d'accès prend note des accès récents, finalement le résultat d'accès est récupéré du répertoire de données et envoyé à l'utilisateur.

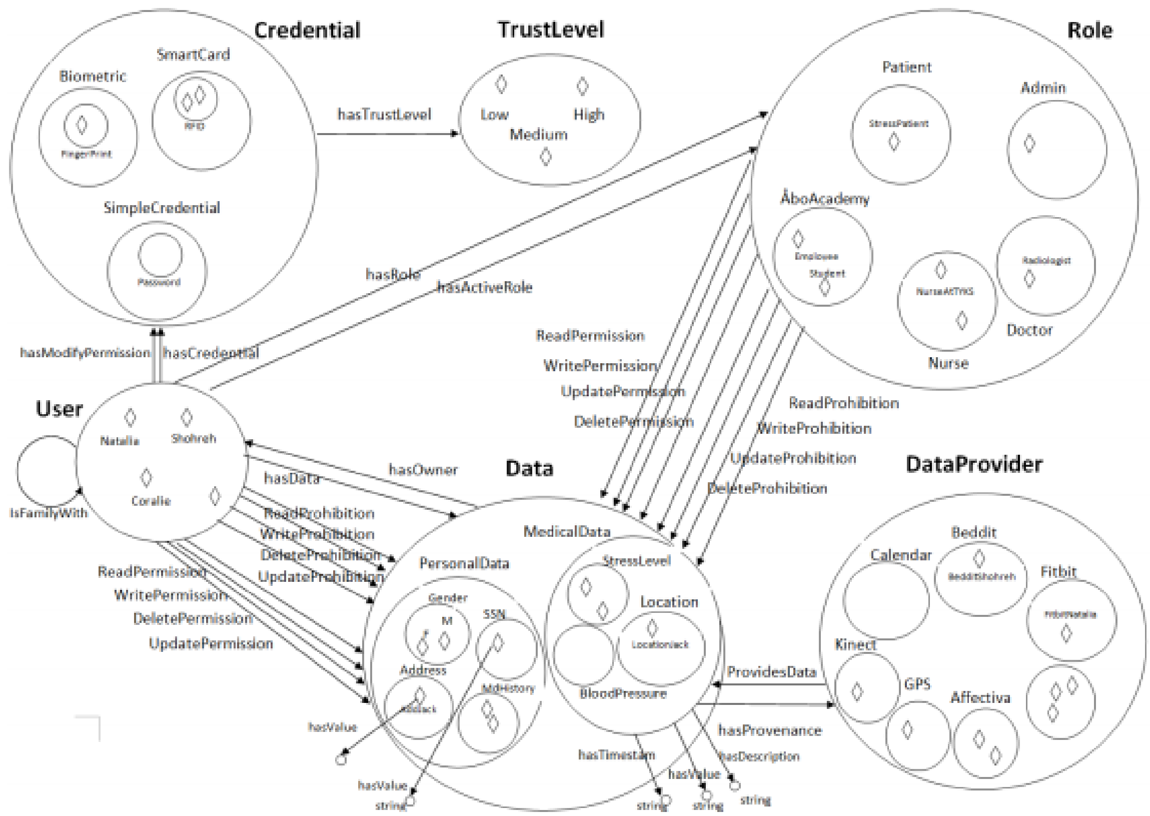

FIGURE 2.2 – Le contrôle d'accès à base d'ontologie [9].

Le modèle proposé renforce la sécurité en combinant le rôle de l'utilisateur et l'information du contexte pour accéder à certaines données, surtout en modélisant ce plan par une ontologie OWL. De plus que l'utilisateur peut déterminer des règles d'accès à leurs informations, non seulement l'administrateur qui gère le contrôle d'accès donc ce qui préserve l'intimité des utilisateurs, mais le modèle proposé est vulnérable à l'attaque par rejeu, par conséquent la notion d'alèrte n'est pas satisfaite. Un administrateur central qui gère les différents accès aux différentes ressources de données donc ce qui fait que ce système de contrôle d'accès n'est pas scalable.

2.3.1.2 Solution basées sur le contenu

1. Multilabels-Based Scalable Access Control for Big Data Applications

Chen et al. [5] ont proposé un modèle de contrôle d'accès basé sur des multilabels qui offre une protection flexible au big data dans le domaine de la santé afin de protéger la confidentialité des patients et de contrôler la granularité de l'accès à ces données. Ils ont proposé un cadre de contrôle d'accès évolutif basé sur des multilabels qui protège les données sensibles stockées dans le système de fichiers distribué Hadoop. Les multilabels comprennent le type de données, le degré de sécurité, la durée de vie, le nombre de réplications, la politique d'accès et la valeur de hachage. Ce cadre combine le faisceau actif, le contrôle d'accès basé sur les rôles (RBAC), sur les attributs (ABAC), le contrôle d'accès discrétionnaire (DAC), et le contrôle d'accès obligatoire (MAC).

Lorsque les propriétaires de données veulent écrire des données dans HDFS, ils doivent être authentifiés par le protocole Kerberos puis ils créent des multilabels en sélectionnant et définissant des étiquettes. Une fois que le propriétaire des données définit les multilabels, le système de contrôle d'accès sécurisé les protégera en empêchant les attaquants de les manipuler ou de les effacer. Ensuite, Ils écrivent les données dans les métadonnées HDFS associées aux données PHR (les dossiers médicaux personnels). Chaque entité souhaitant lire ou écrire des données dans HDFS doit être authentifiée par le protocole Kerberos, elle se voit accorder un niveau de privilège de sécurité en fonction de son rôle dans le système d'informations sur les soins de santé. Le client Hadoop vérifie les multilabels dans les métadonnées HDFS auxquelles il accède, si les attributs de l'entité et les multilabels PHR répondent aux règles de contrôle d'accès, l'entité a accès aux données PHR, sinon l'accès est refusé.

Dans ce modèle, le contrôle d'accès est basé sur des étiquettes, qui offre une protection flexible au big data. Ce modèle est cohérent, mais il n'assure pas la scalabilité. La notion d'alerte n'est pas satisfaite, c'est pourquoi le modèle est vulnérable aux attaques.

2. Access control for big data using data content

Zeng et al. [25] ont proposé un modèle de contrôle d'accès qui prend des décisions basées sur le contenu sémantique des données. La similitude de contenu entre les enregistrements est évaluée et utilisée pour prendre automatiquement les décisions de contrôle d'accès. L'utilisateur est autorisé à accéder à toutes les données en fonction de la règle de contenu définie dans la politique, c'est-à-dire que le contrôle d'accès est basé sur la similarité du contenu entre les informations d'identification utilisateur et le contenu des données dynamiquement.

La politique CBAC pourrait être représentée comme suit :

ACR = $\{$*subject, object, action,* $f(u,d_i)\}$

Où u représente le sujet et d représente l'objet de données. Dans le contrôle d'accès basé sur le contenu, il est supposé que tous les utilisateurs CBAC appartiennent à un rôle spécial qui est autorisé à accéder à "certaines données". En outre, ils supposent un contrôle d'accès choisit avec soin sur les données relationnelles, de sorte que chaque tuple est considéré comme un objet de données. Pour une requête utilisateur, la fonction de contrôle d'accès basée sur le contenu évalue à $f(\mathrm{u}, d_i) = \{true, false\}$ pour chaque d_i dans l'ensemble candidat. L'accès à l'enregistrement est accordé lorsque $func(\cdot\ ,\ \cdot) = true$, et donc d_i est inclus dans le jeu de résultats. Dans le modèle de base, un sujet u est représenté par un ensemble d'objets de données qui lui appartiennent. Les objets de données sont représentés par le modèle bag-of-words (vector-space). La fonction de décision obtiendra la similarité maximale entre l'objet de données candidat et tous les enregistrements du propriétaire et le comparera avec un seuil prédéfini : $f\ (u,\ d_i) = \max_j\ (SIM_d\ (d_{u,j}\ ,\ d_i)) \geq T$

Le modèle de contrôle d'accès prend des décisions en se basant sur la similarité du contenu entre les informations d'identification de l'utilisateur et le contenu des données. Ce modèle assure le contrôle d'accès en ignorant les problèmes de scalabilité et de vulnérabilités aux attaques, par conséquent il n'est pas cohérent et la notion d'alerte n'est pas assurée.

2.3.2 Solutions basées sur la cryptographie à base d'attributs

2.3.2.1 Multi-autorité

1. Enabling Efficient Access Control with Dynamic Policy Updating for Big Data in the Cloud

Yang et al. [24] ont comme objectif, la mise à jour d'une politique dynamique pour assurer un contrôle d'accès efficace pour le big data. Ils ont focalisé sur une méthode de CP-ABE adaptée et ils ont discuté la façon de mettre à jour les politiques en reposant sur les clés comme représenter dans la figure 2.3.

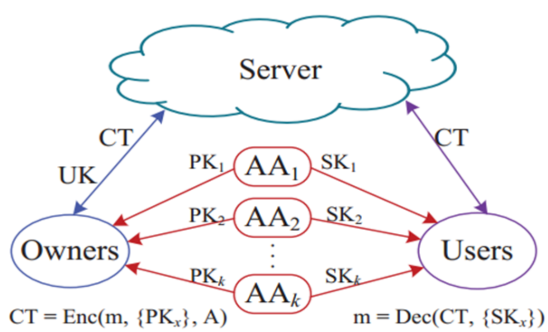

FIGURE 2.3 – Modèle de contrôle d'accès [24].

Le protocole se déroule en cinq (05) phases comme décrit ci-après.

- **Phase d'initialisation**

L'initialisation du système comprend deux configuration : la configuration globale dont chaque utilisateur attribuer un identifiant globale (GID), et la configuration de l'autorité dont chaque autorité d'attributs exécute l'algorithme de configuration d'autorité pour produire sa biclé secrète/publique d'attribut. Ainsi que attribuer l'ensemble de clés publiques d'attributs aux propriétaires de données.

- **Phase de génération clé**

Pour produire la clé secrète pour l'utilisateur GID, chaque autorité lui assignera d'abord un ensemble d'attributs. Ensuite, il execute l'algorithme de génération clés secrète pour lui produire un ensemble de clés secrètes.

- **Phase de cryptage de données**

En premier lieu, le propriétaire de données chiffre les données m en exécutant l'algorithme de cryptage. L'algorithme prend comme les contributions un ensemble de clés publiques PK pour les autorités compétentes, les paramètres globaux, et les données m.

- **Phase de decryptage de données**

L'utilisateur peut décrypter le ciphertext, seulement quand ses attributs satisfont la politique d'accès définie dans le ciphertext.

- **Phase de mise à jour de la politique de sécurité**

Pour mettre à jour la politique d'accès des données cryptées dans le Cloud, il faut déléguer la mise à jour du texte chiffré du propriétaire de données au serveur de cloud, lorsque le propriétaire de données souhaite mettre à jour le texte chiffré de la politique d'accès précédente A à la nouvelle politique d'accès A', il génère d'abord une clé de mise à jour UK_m en exécutant l'algorithme de génération de clés de mise à jour UKGen, puis le propriétaire envoie la clé de mise à jour UK_m au serveur en Cloud. Après avoir reçu la clé de mise à jour du propriétaire de données, le serveur de Cloud exécutera l'algorithme de mise à jour de texte chiffré pour mettre à jour le texte chiffré de la politique d'accès précédente A à la nouvelle A'.

Le modèle proposé s'est basé sur CP-ABE multi-autorités qui est une méthode prometteuse pour contrôler l'accès au big data dans le Cloud, le protocole proposé permet aux propriétaires de données de minimiser leur charge en matière de calcul, au lieu qu'ils téléchargent les données chiffrées à partir du Cloud et les déchiffrer puis rechiffrer ces données sous la nouvelle politique d'accès, c'est juste la clé de mise à jour calculée par les propriétaires de données et cette clé est dirigée vers le Cloud et c'est là où la mise à jour de politique d'accès est faite. En effet la scalabilité est assurée par ce modèle, mais il est vulnérable aux attaques de collusion, donc la sécurité n'est pas assuré, non plus la cohérence et la notion d'alèrte.

2. Efficient Distributed Access Control for Big Data in Clouds

Xiao et al. [22] ont proposé un système de contrôle d'accès distribué pour le big data dans le Cloud illustré sous la figure 2.4, basé sur le CP-ABE multi-autorité dans lequel il n'est pas nécessaire d'avoir une autorité centrale et aucune entité n'est capable de décrypter individuellement les textes chiffrés. Dans ce système, le serveur stocke les chiffrés et la clé privée d'attribut de chaque utilisateur, et chaque utilisateur doit uniquement stocker la clé privée globale et le certificat. Ils ont aussi proposé une nouvelle méthode de révocation des utilisateurs, le processus de révocation des utilisateurs n'est associé qu'aux utilisateurs révoqués.

Le modèle de système est composé de cinq entités : le serveur d'authentification, les autorités à attributs multiples AA, les propriétaires de données, le serveur Cloud et les utilisateurs.

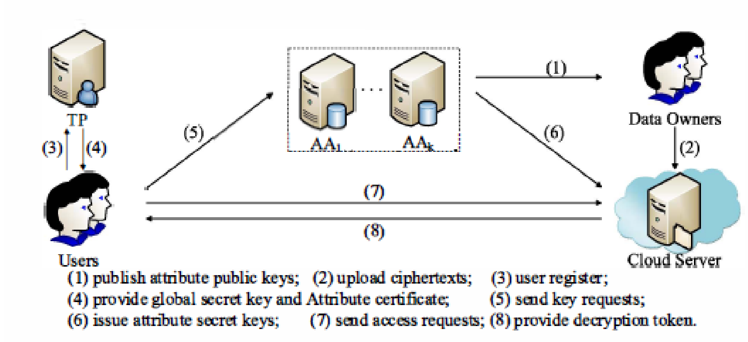

FIGURE 2.4 – Système de contrôle d'accès distribué pour les grandes données [22].

✠ Les phases de construction du modèle [22]

- **Phase 1 : Configuration**

La phase configuration comprend la configuration du serveur d'authentification, la configuration des autorités et la configuration de l'utilisateur.

1. **Configuration du serveur d'authentification :**

   Le serveur d'authentification génère une clé privée et une paire de clés publiques (sk_{TP}, pk_{TP}) pour lui-même.

2. **Configuration AA :**

   Chaque AA_K exécute l'algorithme AAsetup pour génèrer une clé secrète d'attribut x et les clés publiques pour chaque attribut x.

3. **Configuration de l'utilisateur :**

   Lorsqu'un utilisateur rejoint le système, il présente des preuves de son identité au serveur d'authentification. Après que ce dernier authentifie l'utilisateur, il attribue un identificateur global GID unique et une liste d'attributs AL_{GID} à l'utilisateur. Le serveur d'authentification choisit également de façon aléatoire une clés secrètes globales $UGSK_{GID}$ de l'utilisateur. Ensuite, il génère les clés publiques globales utilisateurs $UGPK_{GID}$ et utilise sa clé privée sk_{TP} pour générer un certificat $ACert_{GID}$. finalement il envoie le certificat $ACert_{GID}$ et la clé secrète globale $UGSK_{GID}$ à l'utilisateur en toute sécurité.

- **Phase2 : Chiffrer**

L'algorithme Encrypt prend comme entrées le message m, les paramètres publics, un ensemble d'attributs de clés publiques et une matrice de structure d'accès et sort le CT de texte chiffré.

- **Phase 3 : Génération de clés**

Lorsqu'un utilisateur obtient le $ACert_{GID}$ du serveur d'authentification, il le soumet aux AAs concernées. Chaque AA_k génère la clé secrète d'attribut de l'utilisateur $UASK_{GID,x}$ pour chaque attribut x et cette dernière sera envoyé au serveur Cloud.

- **Phase 4 : Décrypter**

1. **Génération de jeton de décryptage :**

   Lorsque le serveur reçoit une requête d'accès pour un texte chiffré à partir d'un utilisateur, il exécute l'algorithme TKGen. Le serveur utilise $UASK_{GID}$ stocké pour calculer correctement le jeton de décryptage TK si et seulement si les attributs de l'utilisateur satisfont la politique d'accès incorporée dans le texte chiffré.

2. **Décryptage des données**

   Lorsque l'utilisateur reçoit le TK du serveur, il combine sa clé secrète globale avec le TK pour décrypter le texte chiffré et obtient le message m.

✠ **Révocation des utilisateurs :**

La révocation des utilisateurs est considérée comme une mise à jour de l'identité de l'utilisateur et pour cela les auteurs ont rajouté deux nouvelles opérations. L'utilisateur révoqué reçoit un nouvel identifiant GID', une nouvelle clé secrète globale $UASK_{GID'}$, un nouveau certificat $ACert_{GID'}$ et une nouvelle clé secrète d'attribut d'utilisateur $UASK_{GID'}$. Le serveur d'authentification ajoute le GID révoqué à la liste d'utilisateurs révoqués et le serveur supprime $UASK_{GID}$.

Cette solution permet d'assurer le contrôle d'accès au big data dans le Cloud mais aussi d'assurer la sécurité des données en stockant la clé privée d'attribut de l'utilisateur dans le serveur Cloud au lieu de l'envoyer directement à l'utilisateur. Pour pouvoir décrypter un texte chiffré, l'utilisateur aura besoin d'un jeton généré par le serveur mais cette solution est critique car elle ne garantit pas la scalabilité à cause de l'existence d'un seul serveur d'authentification.

3. Privacy-Preserving Access Control Model for Big Data Cloud

Fugkeaw et al. [7] ont proposé une amélioration au système de contrôle d'accès appelé chiffrement basé sur le rôle d'attribut de politique de chiffrement collaboratif (C-CP-ARBE). Pour rendre le processus de cryptage et de décryptage des données plus pratique, ils ont incorporé un cryptage symétrique pour chiffrer les gros fichiers de données compressés. Le C-CP-ARBE est composé de l'intégration d'un chiffrement basé sur les rôles (RBAC) dans un cryptage basé sur les attributs de la politique de chiffrement (CP-ABE).

- **Politique de contrôle d'accès :**

La structure d'accès dans C-CP-ARBE est une structure à base d'arbres ACP. Chaque noeud non feuille de l'arbre ACP représente le noeud de rôle et la porte de seuil où le noeud de rôle est un noeud parent du noeud de seuil. Ainsi, le parent du nœud feuille x est la paire de {Role node, threshold gate}. La fonction attr (x) définie que x dans un nœud feuille de l'arbre. Un attribut spécial "privilège" est introduit en tant que nœud de feuille étendue (EL) de l’arbre ACP afin d’identifier la lecture ou l’écriture Privilège du rôle.

- **Gestion des clés :**

Le système utilise 3 types de clés utilisateurs. Une paire de clés PKI générée par autorité de certification (CA). Une clé de décryptage utilisateur (UDK) qui est générée par AA, UDK est construite à partir des attributs utilisateur et elle est utilisée pour décrypter le texte chiffré. La troisième clé est la clé de cryptage symétrique, il ya deux clés symétriques comprenant SymKey une clé de session générée par temps de cryptage des données et elle est utilisée pour chiffrer les fichiers de données et elle sera concaténé avec le texte chiffré à chiffrer par l’ACP et l’autre clé est un joint secret (SS) utilisé pour chiffrer le texte chiffré comme une autre couche supplémentaire de cryptage.

- **Chiffrement :**

Avant de crypter, le fichier est compressé. Ensuite, l’algorithme de cryptage effectue trois étapes, il prend comme entrées SymKey et la donnée M et il retourne un CT de texte chiffré. Puis il prend ACP pour chiffrer le CT et SymKey et il retourne le chiffré du texte chiffré ECT. Enfin il prend GRP (GroupRole parameter) pour chiffrer ECT et retourne CT scellé (SCT).

- **Déchiffrement :**

La clé secrète globale de l’utilisateur GSK_{uid} sera utilisée pour déchiffrer la clé de session , et EDK_{uid}. Ensuite, l’algorithme retourne ECT et UDK_{id}. Puis, on obtient SymKey et CT en déchiffrant le ECT avec UDK_{id}. Enfin un CT est déchiffré par le SymKey et le M compressé est obtenu et il sera alors décompressé pour retourner le M original.

Le modèle proposé est basé sur une combinaison de contrôle d’accès basé sur le rôle, le cryptage symétrique et le cryptage basé sur les attributs chiffrés ce qui garantit la résistant aux attaques. En effet ce modèle est cohérent et il assure la notion d'alerte, mais malheureusement il n'assure pas la scalabilité a cause du volume de données, quand le volume augmente la performance d'accès diminue.

2.3.2.2 Mono-autorité

1. An Efficient and Fine-grained Big Data Access Control Scheme with Privacy-preserving Policy

Yang et al. [23] ont proposé un plan de contrôle d'accès au big data, avec une politique préservant l'intimité, qui se focalise sur la technique Ciphertext-Policy Attribute based Encryption (CP-ABE) en exerçant une influence au niveau l'attribut et l'idée fondamentale est d'utiliser la formule suivante (M, ρ), où M est une matrice de politique et ρ correspond à chaque ligne M_i de la matrice M à un attribut et cachez les attributs dans la politique d'accès en enlevant simplement la fonction ρ correspondante aux attributs et pour aider le décodage de données ils ont pu concevoir un Bloom Filter d'attribut original pour évaluer si un attribut est dans la politique d'accès et localiser sa position dans cette dernière, s'il est dans la politique d'accès, et chaque case de ce Bloom Filter correspond à $rownumber/attribute$

Le modèle proposé fonctionne comme suit. Chaque propriétaire de données se voit attribuer un ensemble de clés publiques. Chaque consommateur de données devrait s'enregistrer et s'authentifier à l'autorité d'attribut, si le consommateur de données n'est pas légal, il avorte. Autrement, l'autorité d'attribut évaluera le rôle du consommateur de données dans le système et assignera un ensemble d'attributs S choisi d'un espace d'attribut à ce consommateur de données. Avec l'ensemble de ces attributs, l'autorité produit aussi une clé secrète correspondante pour ce consommateur de données. Les propriétaires de données chiffrent leurs données en utilisant l'ensemble des clés publiques d'attribut et une politique d'accès puis il envoie les données chiffrées au serveur du Cloud dans une structure de donnée dite ABF. Les consommateurs de données veulent accéder aux données mises en mémoire dans le Cloud, ils peuvent télécharger les données cryptées selon leurs intérêts mais ils devraient vérifier d'abord que les attributs qu'ils ont possédés sont dans la matrice d'accès, si l'attribut est dans la matrice d'accès, le consommateur de donnée utilise sa secrète et calcule puis récupérer les données en clair. Autrement, l'attribut n'est pas dans la matrice d'accès le décodage échoue. Le contrôle d'accès dans ce modèle proposé arrive pendant le décodage, ce qui signifie que les consommateurs de données peuvent décrypter les données seulement quand leurs attributs peuvent être satisfaisants les politiques d'accès avaient crypté les données.

Le modèle proposé est résistant aux attaques, et pour remédier au problème de faux positifs et faux négatifs, les auteurs ont conçu un ABF, donc les fonctionnalités de sécurité sont assurées ainsi que ce modèle est bien cohérent. Un modèle de contrôle d'accès devrait représenter le support pour des changements, manipulation, pour cela ce modèle est critique en raison de la scalabilité qui n'est pas assurée par le fait qu'il y a une seule autorité d'attribut dans ce système.

2.4 Étude comparative

La table 2.1 illustre une étude comparative que nous avons menée sur les différentes solutions analysées précédemment selon les critères d'évaluation discutés en section 2.2.

Solution	Sécurité	Scalabilité	Notion d'alèrte	Cohérence
Choi et al. [6]	×	×	×	✓
Hosseinzadeh et al. [9]	×	×	×	✓
Zeng et al. [25]	×	×	×	×
Chen et al. [5]	×	×	×	✓
Yang et al.[24]	×	✓	×	✓
Xiao ar al. [22]	✓	×	✓	✓
Fugkeaw et al. [7]	✓	×	✓	✓
Yang et al. [23]	✓	×	✓	✓

TABLE 2.1 – Comparaison des solutions sur contrôle d'accès aux big data

Notre étude des travaux basés sur la sémantique montre que les solutions proposées présentent l'avantage du contrôle d'accès mais s'avèrent de faible efficacité du point de vue de la résistance aux attaques et la notion d'alerte, ce qui influence négativement sur les décisions d'accès qu'elles génèrent et fait que cette approche n'est pas la mieux adaptée pour le contrôle d'accès au big data dans l'Internet des Objets. Bien que le reste des solutions analysées soient résistantes aux attaques et garantissant la notion d'alerte en utilisant les cryptographie à base d'attributs mais elle souffre du problème de scalabilité. À l'issue de cette comparaison il nous a apparu qu'une solution basé sur la cryptographie à base d'attribut sera la mieux adaptée pour le problème posé.

2.5 Conclusion

La sécurité des données est d'une importance primordiale et l'apparition de l'Internet des Objets où des milliers d'objets hétérogènes se voient communiquer les uns avec les autres ce qui génère du big data connu par 5Vs. Donc, contrôler l'accès à ces données est une question stimulante, et il reste encore un certain nombre considérable de défis à surmonter dans cette problématique posée. La recherche des modèles de contrôle d'accès au big data a marqué un accroissement considérable. Dans ce chapitre, nous avons établi un état de l'art sur les travaux de recherche concernant les mécanismes de contrôle d'accès au big data. Pour ce faire, nous avons proposé une classification des solutions selon l'approche suivie. Ensuite, nous avons brièvement décrit chaque solution étudiée suivie d'une discussion des points forts et des points faibles. Enfin, nous les avons comparées selon les différents critères retenus. Le chapitre suivant sera consacré à la description détaillée de notre solution.

Chapitre 3

Proposition d'un modèle de contrôle d'accès au big data dans l'Internet des objets

3.1 Introduction

Le contrôle d'accès est un moyen important pour sécuriser des données et d'empêcher des fuites d'informations, cela peux assurer aux utilisateurs légaux d'accéder et d'utiliser des ressources qu'ils désirent avec autorisation et d'empêcher des utilisateurs illégaux d'une invasion ou d'une éventuelle destruction. En effet, plusieurs domaines (médicale, militaire, environnement, etc.) en tendance à profiter de plus en plus des technologies de collecte de données, ce qui explique la croissance rapide de l'utilisation de ces dernières, d'où vient la notion du big data. À chaque milliseconde des capteurs génèrent une masse importante de données hétérogènes ce qui motive l'importance de contrôler l'accès à ces données, ce problème reste a nos jour un challenge du fait qu'aucune sollution définitive n'ai proposé. Dans ce chapitre, nous parlerons en premier lieu des motivations de réaliser notre travail, puis nous décrirons une architecture détaillée de notre modèle de contrôle d'accès. Nous présenterons, par la suite, le principe de notre solution. Nous terminerons par une analyse de sécurité de la solution proposée et une conclusion.

3.2 Motivation

Le monde se voit de plus en plus connecté, des individus, des données et des objets se lient à travers un écosystème numérique global. Au cours de la dernière décennie, des milliards de périphériques se sont connectés à Internet : 10 à 20 milliards aujourd'hui, contre 500 millions il y a 10 ans. D'ici 2020, il se peut atteindre 30 à 210 milliards d'objets connectés dans le monde d'après les prévisions de Cisco Systems [18]. Ces objets connectés produisent et reçoivent des données massives, hétérogènes dans chaque milliseconde, ce qui est appelé aujourd'hui big data, qui sont difficiles à traiter avec les outils classiques de gestion de base de données. Ces données sont collectées, stockées, et analysées grâce aux serveurs du Cloud qui les traitent pour leur donner du sens. Ces données massives doivent être sécurisées d'une manière efficace, pour cela un modèle de contrôle d'accès choisit avec prudence est nécessaire pour empêcher l'accès des intrus et préserver l'intègrité des données. En effet nous nous somme intéressé aux protocoles basées sur la cryptographie à base d'attributs et μTESLA (timed, efficient, streaming, loss-tolerant, authentication).
Nous jegeons que la technique basée sur la cryptographie à base d'attributs est meilleure en termes de sécurité que les autres techniques de contrôle d'accès aux données (ABAC, CABAC, etc.). ABE est une technique de cryptage prometteuse qui permet de crypter leur données générées par les objets sous les politiques d'accès définies sur quelques attributs de consommateurs de données et permet seulement aux consommateurs de données dont les attributs satisfont les politiques d'accès de décrypter ces données. D'autre part nous somme focalisées sur le principe de μTESLA [14] pour authentifier les grosses données. Le modèle proposé vise à contrôler l'accès au big data ainsi qu'au déchiffrement de données en temps réel.

3.3 Architecture du modèle proposé

Dans cette section, Nous considérons le système de contrôle d'accès de données, comme montré dans la figure 3.1. Le système se compose de cinq entités, à savoir les objets communicants, les Fogs, les serveurs du Cloud, les autorités d'attribut (AAs) et les utilisateurs.

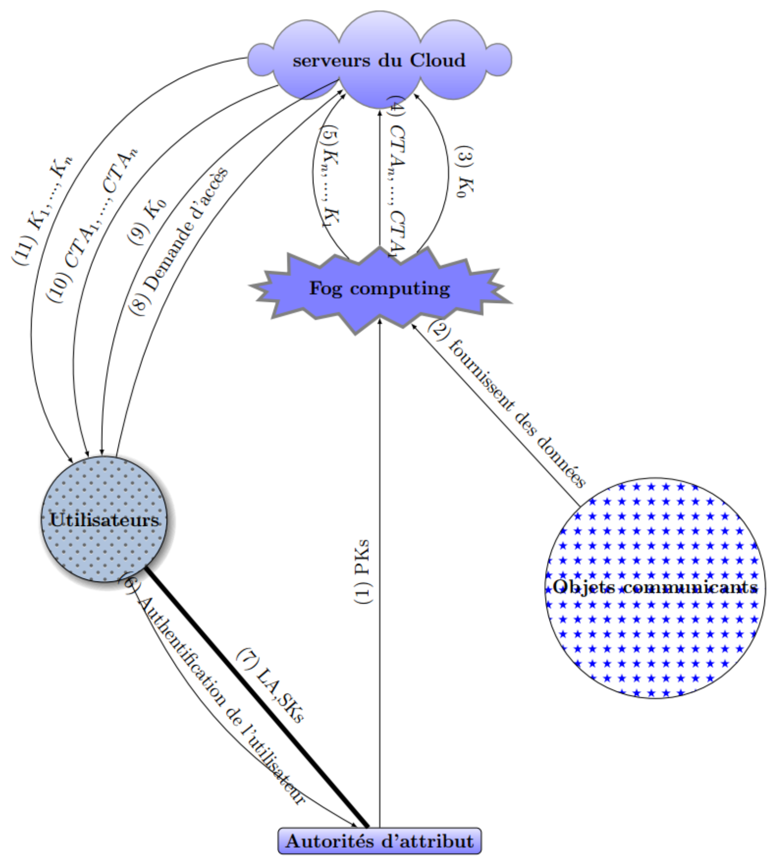

FIGURE 3.1 – Approche proposée.

objets communicants : C'est les objets communicants appartenant au réseau d'Internet des Objets. Ils génèrent des données massives et hétérogènes avec une vitesse importante de génération de données en temps réel. Ces données sont dirigées au Fog d'une manière sécurisée qui est chargé de les envoyer lui-même au serveur du Cloud afin d'être stockées.

Fogs : Le Fog est connu pour sa flexibilité, il offre des services de stockage, de calcul et d'analyse. Dans notre architecture le Fog est utilisé comme intermédiaire entre les objets et le serveur du Cloud où les objets de l'Internet des Objets vont envoyer leurs données au Fog au lieu de les envoyer au serveur Cloud. Donc le Fog va réduire la charge de calcul de ces objets vu que notre modèle s'appuie sur la cryptographie à base d'attributs qui est coûteux de plus que ces objets ont une puissance d'énergie critique, pour cela c'est le Fog qui va chiffrer les données qui seront stockées dans le Cloud.

Serveurs du Cloud : Le serveur Cloud Sest une infrastructure utilisée pour stocker, partager et traiter le big data dans le système. Dans notre architecture, il est chargé de stocker les données chiffrés envoyés par le Fog et de fournir un service d'accès aux données pour les utilisateurs légaux.

Autorité d'attributs : L'autorité d'attribut dirige tous les attributs dans le système et assigne des attributs choisis de l'espace d'attributs aux utilisateurs finaux. C'est aussi un centre de génération des clés, où les paramètres publics sont produits. Il accorde aussi de différents privilèges d'accès aux utilisateurs finaux en publiant des clés secrètes selon leurs attributs et chaque autorité d'attributs (AAs) est indépendante l'une des autres c'est-à-dire pas de communication entre elles.

Utilisateurs : Les utilisateurs sont des consommateurs de données qui demandent des données auprès du serveur. Seulement quand leurs attributs peuvent satisfaire les politiques d'accès des données, les consommateurs de données peuvent décrypter les données.

Les notations utilisées pour l'architecture, ainsi que leurs significations sont données dans la table 3.1.

Notation	Signification
PKs	Ensemble de Clés publiques d'attributs
SKs	Ensemble de Clés secretes d'attributs
LA	Liste d'attributs d'utilisateur
CTA_i	Texte chiffré concaténé avec le code d'authentification
K_i	Clé d'authentification

TABLE 3.1 – Notations.

3.4 Principe du fonctionnement de notre modèle

Afin d'assurer un contrôle d'accès efficace et scalable, nous proposons un nouveau système basé sur la cryptographie à base d'attribut et le principe de μTESLA pour l'authentification des données. Notre modèle fonctionne comme suit :

*** Génération de clés PKs**

Les autorités d'attributs génèrent un ensemble de clés publiques d'attributs PKs et les envoient au Fog pour les stocker et les utiliser quand c'est nécessaire.

*** Génération et envoi des données au Fog**

L'ensemble des objets se mettent d'accord sur une clé de chiffrement (on suppose qu'on n'a pas d'objets malicieux), et l'échanger avec le Fog en utilisant le protocole AES. À chaque fois que les objets génèrent des données, ils les chiffrent puis les diriger au Fog.

*** Déchiffrement et chiffrement au niveau du Fog**

En premier lieu, le Fog reçoit les données chiffrées de différents objets. Il déchiffre ces dernières et récupére les données initiales. Ensuite, il les chiffre en utilisant une clés publique d'attribut PKs sous une politique d'accès, ce qui donne ciphertext. Pour authentifier les données, le principe de micro TESLA est appliqué comme illustrer sous la figure 3.2. Il utilise une fonction de hachage à sens unique pour générer une liste de clés $K_i = F(K_{i+1})$ et calculer un MAC (Message Authenticated Codes) comme suit. Le Fog calcule K_0 et l'envoie au serveur du Cloud comme engagement pour identifier la source de la donnée. La liste des clés K_i est divisée en intervalles, dont tous les paquets envoyés dans un même intervalle de temps sont authentifiés avec la même clé K_i. Ensuite, attacher le MAC au chiffré ce qui donne CTA ; et l'envoyer au serveur du Cloud, un ensemble de clés K_i sera divulgé et stocké ultérieurement au serveur dans un lampse de temps.

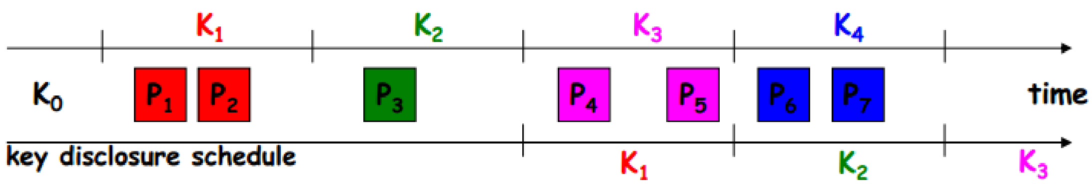

FIGURE 3.2 – Principe de μTESLA [14].

*** Contrôler l'accès aux utilisateurs**

Lorsqu'un utilisateur désire accéder aux données stockées dans le Cloud, il doit s'authentifier auprès des autorités d'attributs. Ces dernières lui envoient une liste d'attributs et un ensemble de clés secrètes d'attributs à cet utilisateur via un canal sécurisé. Puis, l'utilisateur envoie une requête d'accès au serveur, qui lui répond par l'envoi d'une clé initial K_0, ainsi l'utilisateur peut télécharger les CTA_i dont leurs attributs satisfont la politique d'accès définie dans le texte chiffré. Par la suite il vérifie l'authenticité des ciphertexts reçus, après avoir reçu les clés K_i. Enfin, il utilise l'ensemble des SK_s pour déchiffrer les données paquet par paquet.

3.5 Analyse de sécurité

Dans cette section, nous passons à l'analyse des propriétés de sécurité de la solution proposée afin de montrer qu'elle est résistante aux types d'attaques suivants :

3.5.1 Attaque par rejeu

Les attaques par rejeu sont des attaques de type « Man in the middle » consistant à intercepter des paquets de données et à les rejouer, c'est-à-dire les retransmettre tels quel (sans aucun déchiffrement) au serveur destinataire. Ainsi, selon le contexte, l'attaquant peut bénéficier des droits de l'utilisateur. Dans notre modèle si jamais un intrus essaye d'avoir la liste d'attributs d'utilisateur ainsi que l'ensemble de clés secrètes d'attributs, il n'arrive pas puisque un canal sécurisé est utilisé entre les autorités d'attributs et les utilisateurs. De même quand l'utilisateur accède au serveur Cloud, il télécharge les données chiffrées donc cette attaque ne peut pas y arriver dans notre modèle.

3.5.2 Attaque de collusion

Dans notre système, il n'est pas possible pour les différents utilisateurs de combiner leurs ensembles d'attributs afin d'accéder aux ressources. Même si les adversaires utilisent un attribut ayant le même nom et la même valeur (par exemple position : médecin) de différentes autorités d'attributs, ils ne peuvent pas collaborer avec l'utilisateur légal dans l'autorité partagée de la ressource cible. Ceci est dû au fait que chaque attribut possède sa propre clé publique, son identifiant unique et les clés privées d'attributs de chaque utilisateur sont randomisées, c'est à dire, un nombre aléatoire r constitué dans l'algorithme UserKeyGen permet à la clé privée d'attribut construite à partir d'un ensemble d'attributs est techniquement distinguable et non substituable. En outre, la liste d'attributs et l'ensemble de clés secrètes d'attributs sont échangées via un canal sécurisé. Cela empêche l'attaque collusion des adversaires.

3.6 Conclusion

La gestion de contrôle d'accès au big data est un problème difficile à résoudre dans l'Internet des Objets, dû a la vitesse d'envoie, le volume et la variété des données. Toute solution a ce problème doit être résistante aux attaques et assure la scalabité. Ce chapitre a été consacré à la presentation de notre proposition dans le contrôle d'accès au big data. Tous d'abord nous avons décrit brièvement ce qui nous a motivé à réaliser ce travail. Puis, nous avons présenté l'architecture du notre modèle ainsi que le principe de son fonctionnement. Enfin, nous avons conclu par une analyse de sécurité de la solution proposée. Le prochain chapitre sera consacré à l'evaluation des performances de notre proposition.

Chapitre 4

Simulation et évaluation de performances

4.1 Introduction

Ce chapitre est consacré à l'évaluation des performances de notre protocole de contrôle d'accès. Nous présenterons en premier lieu l'environnement et les paramètres de simulation considérés pour l'évaluation. Nous décrirons par la suite les critères et métriques de simulation utilisés. Les résultats obtenus à l'issue de ces simulations seront finalement interprétés et comparés avec un protocole récent étudié dans le chapitre de l'état de l'art.

4.2 Environnement de simulation

Dans cette section, nous présentons au préalable les paramètres de simulation, puis nous décrivons les critères et métriques de simulation utilisés.

4.2.1 Paramètres de simulation

Notre protocole de contrôle d'accès a été simulé sous l'environnement Java. La simulation a débuté par le chiffrement AES, ou les données générées par les objets ont était chiffrées. Ensuite, elle vient l'étape de déchiffrement AES afin de récupérer

Paramètre	Valeur
Vitesse de transmission des données	5 bit/s
Vitesse de transmission de μTESLA	240 bit/s
Taille de la clé μTESLA	64 bit
Taille du paquet μTESLA	80 bit
Taille du jeton	2048 bit

TABLE 4.1 – Paramètres de simulation

les données initiales pour pouvoir les chiffrer en utilisant la cryptographie a base d'attributs CP-ABE. Après cette dernière étape, elle vient l'étape de μTESLA qui consiste à construire des paquets de petite taille en fragmentant la donnée chiffrée en petites parties et le tout sera transféré vers le Cloud pour le stockage. L'utilisateur récupère les paquets, ensuite il déchiffre avec les clés de μTESLA pour récupérer les données chiffrées et enfin il aura la donnée en claire en exécutant la procédure de déchiffrement CP-ABE. Les paramètres fixés pour la réalisation des simulations sont définis dans la Table 4.1.

4.2.2 Critère et métriques de simulation

Dans cette section, nous présentons le critère et métriques de simulation que nous avons utilisés pour l'évaluation de performances de notre protocole.

4.2.2.1 Critère de simulation

Le critère de simulation utilisé pour l'évaluation de performances de notre protocole est la fréquence de génération de données. Dans l'Internet des Objets, les objets fournissent des masses de données importantes ce qui est dit aujourd'hui le big data. Dans de tel domaine, la fréquence de génération de données se trouve être un paramètre important à prendre en compte, elle représente la quantité de données transmise par unité de temps.

4.2.2.2 Métriques de simulation

Afin d'évaluer les performances de notre protocole, nous utilisons les métriques de simulation suivantes :

Temps d'execution
Le temps d'exécution est une métrique importante à mesurer dans notre architecture. En effet, elle représente la durée nécessaire pour que les données générées par les objets arrivent à l'utilisateur d'une manière sécurisée en effectuant les différents chiffrements possibles.

Temps de récupération des données en clair
Le temps de récupération les données en clair représente le temps nécessaire pour que la requête d'accès de l'utilisateur soit satisfaite auprès du serveur du Cloud, c'est à dire le temps qu'il faut pour télécharger les données chiffrées et les déchiffrer.

4.3 Résultats et discussion

Dans cette section, nous nous sommes intéressés à comparer les performances d'un protocole étudié dans le chapitre de l'état de l'art avec celles de notre proposition. Dans ce qui suit, nous présentons les résultats sous forme de graphiques puis nous les interprétons.

La figure 4.1 illustre la variation de temps d'exécution total en fonction de la fréquence de génération de données.

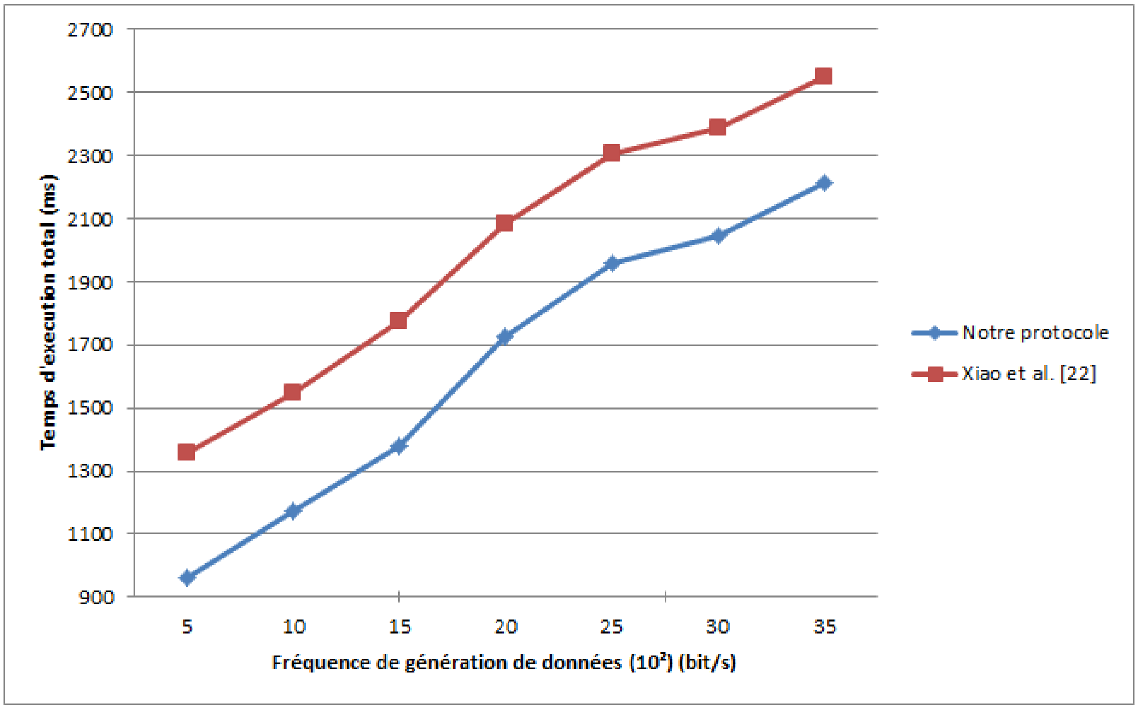

FIGURE 4.1 – Temps d'exécution total en fonction de la fréquence de génération de données.

La figure 4.2 illustre la variation de temps de récupération de données en clair en fonction de la fréquence de génération de données.

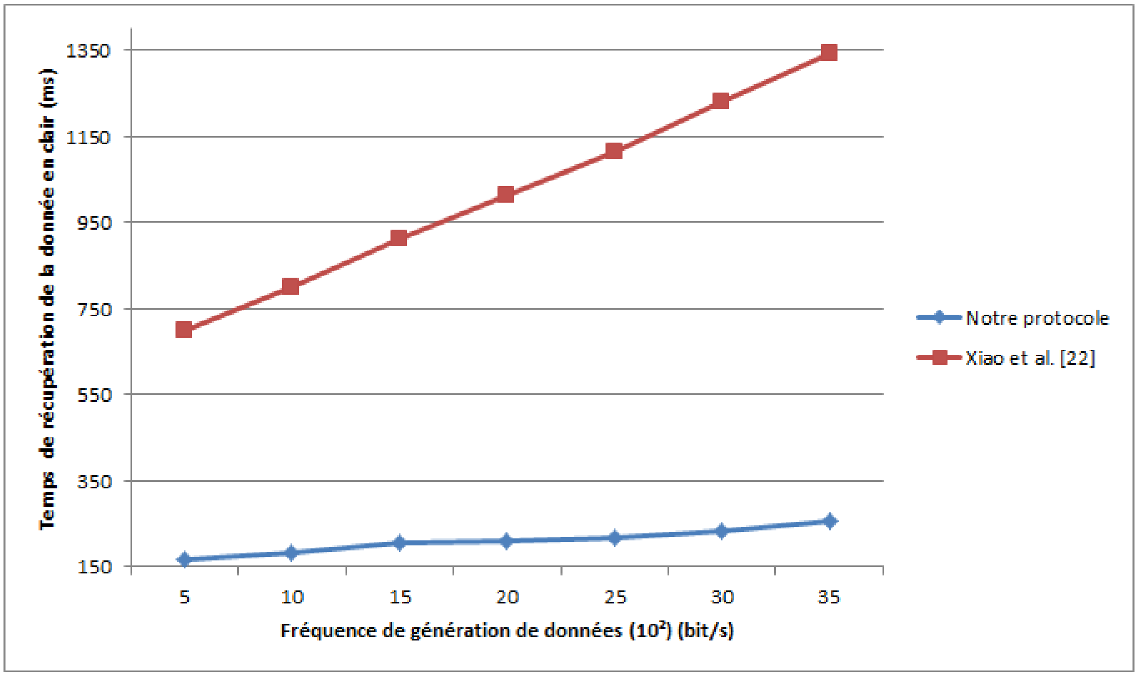

FIGURE 4.2 – Temps de récupération des données en clair en fonction de la fréquence de génération de données.

L'analyse des graphiques précédents montre que l'augmentation de la fréquence de génération de données entraîne l'accroissement des deux métriques mesurées.

Concernant le temps d'exécution, notre protocole montre de meilleurs résultats comparés à l'autre protocole en raison de fragmentation de la taille des données transmises par les différents objets en différents paquets en appliquant le principe de μTESLA qui a mis en faveur notre architecture. Le protocole proposé par Xiao et al. [22] montre une augmentation de temps d'exécution total due à la quantité de données générées par les objets communicants. La méthode CP-ABE utilisée nécessite un temps important de chiffrement et de déchiffrement et la taille du jeton à transmettre est importante.

Concernant le temps de récupération des données en clair, les résultats relatifs à notre protocole sont très satisfaisants et meilleurs que ceux présentés par l'autres solution. Ceci est dû à la rapidité de déchiffrement des différents paquets (64bits). En revanche, dans le protocole proposé par Xiao et al. [22], le temps de récupération des données en clair est très élevé, en raison de la génération d'une masse importante de données de taille grande par les objets. Ce qui provoque un temps de déchiffrement CP-ABE qui est assez grand.

4.4 Conclusion

Dans ce chapitre, nous avons évalué les performances de notre protocole en le comparant avec un autre protocole de contrôle d'accès. Pour ce faire, nous avons varié la fréquence de génération de données afin d'étudier son impact sur le temps d'exécution total et le temps de récupération des données en clair. Les résultats obtenus montrent que notre protocole présente de bonnes performances en termes de temps d'exécution et de temps de récupération des données.

Conclusion générale et perspectives

Le “big data” est en pleine croissance grâce à l’Internet des objets, dont des milliers d’objets dans le monde se voient connecter les uns avec les autres en partageant des données avec les serveurs du Cloud. l’Internet des objets est apparus comme une technologie ayant le potentiel pour révolutionner divers domaines de notre vie quotidienne. Toutefois, le big data et l’Internet des objets sont encore au stade précoce de leur développement, et plusieurs défis de recherche doivent être surmontés afin qu’ils puissent être largement déployés. Le contrôle d’accès est l’un des principaux défis du big data à relever tant que les données collectées sont sensibles et doivent être confidentielles. Un intrus mal intentionné peut tenter d’usurper l’identité d’un utilisateur légitime afin d’altérer des données, ou bien accéder à certaines données qui ne lui sont pas tolérées pouvant mettre en danger la vie privée de leurs propriétaires. De plus, la conception d’un mécanisme de contrôle d’accès au big data doit faire face à certaines contraintes en raison des objets qui possèdent des ressources limitées en termes de puissance de calcul, énergie, espace de stockage, etc. Donc, assurer le contrôle d’accès et la bonne gestion des données massives générées par les appareils intelligents, aussi celles produites par l’homme est une question stimulante.

Nous avons étudié à travers ce mémoire le problème de contrôle d’accès au big data générées par les objets communicants. Nous avons présenté en premier lieu la définition, l’architecture et les domaines d’application de l’Internet des objets et leur classification, puis le big data et ses caractéristiques. Nous avons ensuite expliquer la relation entre le big data et l’Internet des objets. Enfin, nous avons clôturé cette étude en décrivant les principaux défis imposés par le big data et la sécurité dans l'Internet des objets. Par la suite, nous avons cité quelques travaux liés au problème étudié, où nous avons présenté une étude critique, selon plusieurs critères. Nous avons ainsi constaté que dans la plupart de ces travaux, la sécurité n'est pas assurée. La deuxième partie de ce mémoire concerne notre contribution dans la sécurité du big data. Nous avons pris de certains travaux de recherche des idées que nous avons utilisées par la suite pour élaborer l'approche que nous avons proposée pour le problème de contrôle d'accès. Notre protocole se base sur le chiffrement à base d'attributs qui est une technique prometteuse pour contrôler l'accès aux données massives ainsi que adapter le principe de μTESLA pour authentifier les différentes données. Nous avons simulé notre proposition en la comparant avec un protocole récent parmi ceux étudiés. Les résultats obtenus montrent l'influence des paramètres choisis sur le comportement de l'approche concernant le délai de récupération des données en clair ainsi que le temps d'exécution total en fonction de la fréquence de génération de données. Les résultats obtenus sont satisfaisants et montrent l'avantage de notre approche en ce qui concerne le contrôle d'accès par la réduction du délai d'attente des requêtes et minimiser le temps d'execution total de notre protocole.

Comme perspectives pour les futurs travaux, évaluer les performances de notre proposition sur la base des critères : la scalabitilé et la notion d'alerte.

Bibliographie

[1] Expanded top ten big data security and privacy challenges, April 2013. https ://downloads.cloudsecurityalliance.org/initiatives/bdwg/(accédé le 01/06/2017.

[2] ALPWISE. Domotique, 2016. http ://www.alpwise.com/fr/conception-objets-connectes/solution-iot/objets-connectes/domotique/(accédé le 31/05/2017).

[3] K. Ashton. That 'internet of things' things. *The Real World Things Matter More than Ideas. RFID Journal*, 2009.

[4] Y. Challal. *Sécurité de l'Internet des Objets : vers une approche cognitive et systémique*. HAL, September 2013.

[5] H. Chen, B. Bhargava, and F. Zhongchuan. Multilabels-based scalable access control for big data applications. *IEEE Cloud Computing*, 01 :65–71, 2014.

[6] C. Choi, J. Choi, and P. Kim. Ontology-based access control model for security policy reasoning in cloud computing. *The Journal of Supercomputing*, 67(3) :711–722, March 2014.

[7] S. Fugkeaw and H. Sato. Privacy-preserving access control model for big data cloud. *International Computer Science and Engineering Conference (ICSEC)*, DOI : 10.1109/ICSEC.2015.7401416, 2015.

[8] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of things (iot) : A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7) :1645–1660, September 2013.

[9] S. Hosseinzadeh, N. Díaz-Rodríguez, S. Virtanen, and L. Johan. A semantic security framework and context-aware role-based access control ontology for smart spaces. In G. Sven and G. Lee, editors, *International Workshop on Semantic Big Data*, page 81–86. ACM, 2016.

[10] K. L. Lueth. The 10 most popular internet of things applications right now, February 2015. https ://iot-analytics.com/10-internet-of-things-applications/(accédé le 31/05/2017.

[11] J. Moreno, A. Manuel-Serrano, and E. Fernández-Medina. Main issues in big data security. *Journal of Future Internet*, 44(8), 2016.

[12] Cluster of Europrean Research Projects on the Internet of Things. *Vision and Challenges for the Internet of Things*. March, 2010.

[13] T. Pasquier, J. Bacon, J. Singh, and David Eyers. Data-centric access control for cloud computing. *SACMAT '16 Proceedings of the 21st ACM on Symposium on Access Control Models and Technologies*, 21 :81–88, 2016.

[14] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J. D. Tygar. Spins : Security protocols for sensor networks. *Kluwer Academic Publishers*, 8 :521–534, 2002.

[15] S. Poudel. Internet of things : Underlying technologies, interoperability, and threats to privacy and security. *Berkeley Technology Law Journal*, 31(2), September 2016.

[16] H. Roxana and G. Mircea. Small steps or big changes in actual society : The impact of internet of things. *Journal of Public Administration, Finance and Law*, (10), 2016.

[17] V.C. Ru, T. Grance, D .F. Ferraiolo, and D. Rick Kuhn. An access control scheme for big data processing. *10th IEEE International Conference on Collaborative Computing : Networking, Applications and Worksharing*, DOI : 10.4108/icst.collaboratecom.2014.257649, 2014.

[18] C. Schmiedt. L'internet des objets et le big data, qu'est-ce que c'est ?, Juin 2016. https ://www.entreprisedufutur.com/actualites/clementine-schmiedt/l-internet-des-objets-et-le-big-data-qu-est-ce-que-c-est/(accédé le 31/05/2017).

[19] H. Syphax. Vers un modèle de confiance pour l'internet des objets. Master's thesis, Univers ité ABDERRAHMANE MIRA DE BEJAIA, 2016.

[20] Z. Wassim. *Quelques propositions de solutions pour la sécurité des réseaux de capteurs sans fil*. PhD thesis, Institut National des Sciences Appliquées de Lyon, 2010.

[21] M. Wu, T. Lu, F. Ling, J. Sun, and H. Du. Research on the architecture of internet of things. *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on*, 5 :V5–484,V5–487, Aug 2010.

[22] M. Xiao, M. Wang, and J. Sun X. Liu. Efficient distributed access control for big data in clouds. *The Third International Workshop on Security and Privacy in Big Data*, DOI :10.1109/INFCOMW.2015.7179385, 2015.

[23] K. Yang, Q. Han, H. Li, X. Sherman-Shen, K. Zheng, and Z. Su. An efficient and fine-grained big data access control scheme with privacy-preserving policy. *IEEE*, 2016.

[24] K. Yang, X. Jia, K. Ren†, R. Xie, and L. Huang. Enabling efficient access control with dynamic policy updating for big data in the cloud. *IEEE Conference on Computer Communications*, DOI : 10,1109 / INFOCOM.2014.6848142, 2014.

[25] W. Zeng, Y. Yang, and B. Luo. Access control for big data using data content. *IEEE International Conference*, DOI : 10,1109 / BigData.2013.6691798, 2013.

RÉSUMÉ

Chaque jours, de plus en plus de d'objets sont équipés de capteurs pour recueillir des données et les personnes qui les utilisent dans différents domaines (médicale, militaire, environnement, etc.), d'où vient la notion du Big data. A chaque seconde les objets génèrent une masse importante de données hétérogènes ce qui fait que contrôler l'accès à ces données est un sujet de préoccupation majeure non encore résolu. L"objectif de ce mémoire est de répondre au problème de contrôle d'accès au big data dans l'Internet des objets. Pour atteindre à cet objectif, nous avons proposé un modèle de contrôle d'accès basé sur la méthode de chiffrement à base d'attribut ABE, qui est une technique prometteuse pour contrôler le big data, combinée avec le protocole de micro TESLA pour authentifier les données en temps réel. L"analyse de sécurité a permis de démontrer la robustesse de notre protocole contre certaines attaques connues. De plus, les résultats des simulations de notre protocole et une comparaison avec une autre solution existante dans la littérature ont mis en relief les avantages de notre protocole, en termes de temps d'execution et de temps de récupération des données en clair.

Mots clés : ABE, Big data, Sécurité, Contrôle d'accès, Internet des objets, Micro TESLA.

ABSTRACT

Today, the world is experiencing an exponential increase of sensors integration into different devices in order to capture data and their users in various fields such as medical, military and environmental applications. This lead to the new notion of Big Data. The amount of heterogeneous data generated at any given moment is tremendous and its access control is still a major challenge encountered today that needs further focus to be resolved. Hence, the objective of this thesis is to address Big Data access control in the Internet of Things (IoT). A control access model based on a coding methodology using ABE attribute is chosen and proved to be a very promising technique for the Big Data control, this is combined in line with the micro TESLA protocol to help in real time data authentication. Security analysis against some knows threats has shown the reliability of our model, in addition; simulation results and comparison to an already. Established solution in the academy demonstrated clearly the advantages attained by our protocol in terms of execution and recuperation times.

Key words : ABE, Big data, Security, Access control, Internet of Things, Micro TESLA.