UNIVERSITÉ DU QUÉBEC À MONTRÉAL

IOTFLA : UNE ARCHITECTURE DE DOMOTIQUE SÉCURISÉE RESPECTUEUSE DE LA VIE PRIVÉE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN INFORMATIQUE

PAR

ALEXANDRE MARTIN

DÉCEMBRE 2019

REMERCIEMENTS

Je voudrais tout d'abord exprimer toute ma gratitude à mon directeur de recherche, M. Sébastien Gambs pour m'avoir permis de réaliser mon projet de recherche dans le domaine de la sécurité et de la protection de la vie privée sous sa direction et au sein de l'équipe *PrivSEC*.

Je le remercie également pour son aide qu'il m'a fournit ainsi que les connaissances qu'il a su me transmettre tout au long de la maîtrise. De plus, je tiens à souligner sa disponibilité, sa patience et la qualité de ses conseils qui ont contribué à alimenter ma réflexion, ce qui m'a permis principalement de mener à bien cette maîtrise.

Je remercie toutes les personnes ayant faite partie de l'équipe *PrivSEC* dont j'ai eu l'occasion de rencontrer pendant la période de la maîtrise.

Je souhaite préciser l'importance du fait que j'ai eu l'opportunité d'effectuer la maîtrise en informatique au sein de l'Université du Québec à Montréal suite à un échange étudiant. Je remercie donc non seulement l'Université du Québec à Montréal mais également mon ancienne école, l'Exia Cesi pour l'opportunité et leurs accompagnements.

Dernièrement, cette opportunité n'aurait pas été réalisable sans le support et soutien de ma famille et je les en remercie.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

RÉSUMÉ

Le développement important du nombre d'objets connectés ainsi que le développement du nombre d'acteurs (constructeurs, entreprises, particuliers, organismes, etc.) font que l'Internet des Objets (IdO / *Internet of Things*) ainsi que sa portée sont en constante évolution et son marché ne cesse de croître. Cette croissance est significative et génère un impact de plus en plus important de l'IdO sur la sécurité ainsi que la protection de la vie privée des utilisateurs et propriétaires d'objets connectés.

La sécurité et la protection de la vie privée sont au coeur de l'IdO, nous avons donc décidé d'en faire les aspects principaux de ce sujet de maîtrise. L'IdO a un large éventail d'objets connectés et de domaines d'applications, afin de nous attaquer à un domaine plus précis, nous avons choisi la domotique ou plus communément appelé les maisons intelligentes. Les maisons intelligentes représentent peut être l'un des domaines de l'IdO le plus à risques pour la protection des données privées qui dans ce contexte sont celles d'utilisateurs et propriétaires d'objets connectés.

L'objectif de ce mémoire est de mettre en avant notre proposition dédiée spécifiquement pour le domaine de la domotique de l'IdO. Nous proposons d'explorer la combinaison de différentes technologies au sein d'une même et seule architecture nommée IOTFLA afin d'essayer de pallier aux manques de sécurité et protection de la vie privée qui sont présents dans un scénario classique d'utilisation d'objets connectés au sein de l'IdO.

Cette architecture utilise différentes technologies telles que l'apprentissage fédéré, l'agrégation sécurisée des données, une solution d'automatisation de domotique. Elle est nommée IOTFLA en raison de la convergence des différents domaines abordés ainsi que de l'utilisation de l'apprentissage fédéré, IOTFLA est d'ailleurs basée sur l'architecture originale à trois couches de l'IdO.

Ce mémoire est composé d'une première partie sur la littérature scientifique sur les différentes terminologies et technologies utilisées au sein de la proposition, telles que l'IdO, les architectures, l'apprentissage machine (et fédéré), les protocoles d'agrégation de données, les systèmes de détection d'intrusions. Puis, la seconde partie est dédiée à l'explication de la proposition et de son fonctionnement prévu, regroupant les différents composants détaillés dans la première partie.

ACRONYMES

Les acronymes sont regroupés sur cette page :

- **2D** : **D**eux **D**imensions
- **3DES** : **3** **D**ata **E**ncrytion **S**tandard
- **AF** : **A**pprentissage **F**édéré
- **AM** : **A**pprentissage **M**achine
- **ASD** : **A**grégation **S**ecurisée des **D**onnées
- **BLE** : **B**luetooth **L**ow **E**nergy
- **CAM** : **C**ode d'**A**uthentification de **M**essage
- **CDA** : **C**oncealed **D**ata **A**ggregation
- **CoAP** : **Co**nstrained**A**pplication **P**rotocol
- **DDOS** : **D**istributed **D**enial **O**f **S**ervice
- **DES** : **D**ata **E**ncrytion **S**tandard
- **DOS** : **D**enial **O**f **S**ervice
- **DMZ** : **D**emilitarized **Z**one
- **ECC** : **E**lliptic **C**urve **C**ryptography
- **ECIPAP** : **E**fficient **C**onfidentiality and **I**ntegrity **P**reserving **A**ggregation **P**rotocol
- **EEHA** : **E**nergy **E**fficient and **H**igh **A**ccuracy secure data aggregation
- **EESSDA** : **E**nergy **E**fficient and **S**calable **S**ecure **D**ata **A**ggregation
- **ESPDA** : **E**nergy **E**fficient and **S**ecure **P**attern-based **D**ata **A**ggregation
- **GPS** : **G**lobal **P**ositioning **S**ystem
- **HA** : **H**ome **A**ssistant
- **IA** : **I**ntelligence **A**rtificielle

— **IdO** : Internet des Objets

— **IP** : Internet Protocol

— **IPHCDA** : Integrity Protecting Hierarchical Concealed Data Aggregation

— **LAN** : Local Area Network

— **MQTT** : Message Queing Telemetry Transport

— **OSI** : Open Sytems Interconnection

— **PEPPDA** : Power Efficient Privacy Data Aggregation

— **RDCA** : Recoverable Concealed Data Aggregation for Data Integrity

— **RFID** : Radio Frequency Identification

— **RGPD** : Réglement Général sur la Protection des Données

— **RSDA** : Reputation-based Secure Data Aggregation

— **SDAP** : Secure Hop-by-Hop Data Aggregation Protocol

— **SDAV** : A Secure Data Aggregation and Verification Protocol

— **SDI** : Système de Détection d'Intrusion

— **SEDAN** : Secure and Efficient procol for Data Aggregation

— **SEEDA** : Secure End to End Data Aggregation

— **SELDA** : Secure and Reliable Data Aggregation

— **SIA** : Secure Information Aggregation

— **SHA-1** Secure Hash Algorithme - 1

— **SPI** : Système de Prévention d'Intrusion

— **SRDA** : Secure Reference-based Data Aggregation Protocol

— **SSH** : Secure Shell

— **SSL** : Secure Socket Layer

— **TAG** : Tiny AGgregation

— **TCP** :Transmission Control Protocol

— **TFL** : Tensor Flow Light

— **TLS** : Transport Layer Security

— **TOR** : The Onion Router

— **VPN** : **V**irtual **P**rivate **N**etwork

— **WI-FI** : **Wi**reless **Fi**delity

CHAPITRE I

INTRODUCTION

Depuis toujours, nous (les êtres humains) essayons d'améliorer des facettes (domaines) de la vie de tous les jours, en apportant des solutions ou créant de nouveaux outils sous différentes formes. Avec la création des systèmes électroniques et informatiques, il y a de cela quelques décennies, le monde n'a cessé d'être modifié et révolutionné.

De nos jours, nous vivons dans un monde où une grande majorité de notre environnement est constitué d'éléments (objets connectés) ayant les capacités et ressources qui amènent à les qualifier d' « intelligents ». Ces objets connectés sont des parties intégrantes de ce qu'on appelle l'Internet des Objets (IdO) qui est un paradigme récent qui gagne rapidement du terrain dans le milieu des télécommunications sans-fils modernes.

L'idée générale de l'IdO réside dans l'omniprésence d'objets connectés (identification radio, capteurs, actionneurs, téléphones intelligents, etc.) pouvant interagir, communiquer et échanger des données avec d'autres objets présents dans l'environnement afin d'atteindre leurs objectifs. Les objets connectés ont une présence importante de nos jours pour un grand nombre de personnes (présenté plus en détail dans la section 2.1).

Pour le grand public, qui est composé des utilisateurs finaux des objets connectés, il est difficile de savoir ce qui peut nuire à la protection des informations privées ainsi que de la sécurité. Avec la multitude de constructeurs, d'objets connectés proposés sur le marché de l'IdO de nos jours, la confusion est compréhensible pour des personnes n'ayant pas les connaissances pour comprendre les enjeux et risques liés à la sécurité.

Ce projet de recherche s'inscrit dans le cadre d'une maîtrise en informatique à l'Université du Québec à Montréal au sein du laboratoire *LATECE* (*Laboratoire de Recherches sur les Technologies du Commerce Électronique*) et est dirigé par le professeur Sébastien Gambs. Il est un croisement de la sécurité et la protection de la vie privée au sein de l'IdO

La proposition présentée au sein de ce mémoire, l'architecture IOTFLA a fait l'objet d'une publication pour le *workshop SafeThings 2019 (IEEE Workshop on the Internet of Safe Things* (`https://www.ieee-security.org/TC/SPW2019/SafeThings/`). Ce workshop a pris place à San Francisco le 23 mai 2019 et notre publication porte le titre suivant : *IOTFLA : A Secured and Privacy-Preserving Smart Home Architecture Implementing Federated Learning*

1.1 Problématique de la recherche

L'IdO (Kevin, 2009) est un paradigme qui ne cesse d'être modifié, amélioré et qui continu à occuper une place de plus en plus grande dans notre monde. Augmentant avec lui le nombre d'acteurs, du constructeur de matériels, designer, concepteur, aux entreprises, vendeurs, mais également aux utilisateurs finaux (que ce soit des professionnels ou particuliers).

Chaque acteur peut modifier en fonction de leurs besoins la base de l'architecture générale de l'IdO (Wu *et al.*, 2010), aussi bien du côté de l'industrie des construc-

teurs que du côté scientifique avec diverses propositions d'architecture (présenté dans la section 2.3). En utilisant des architectures, technologies et composantes différentes, les problèmes de sécurité et la protection des données peuvent être différents de l'une à l'autre.

Le nombre important d'acteurs sur le marché de l'IdO (section 2.1) impacte directement la sécurité et la protection de la vie privée pour les utilisateurs d'objets connectés. En effet, avec la diversification de l'utilisation des objets connectés à travers différents domaines (santé, sport, domotique, militaire, automobile, agriculture, industrie, villes, environnement, etc.), l'impact grandissant des objets connectés est conséquent (illustrer dans la section 2.1).

Le marché de l'IdO (détaillé dans la section 2.1) n'est pas axé sur la sécurité et la protection de la vie privée comme pour le concept de la protection de la vie privée dès la conception, mais plutôt une course à l'innovation ayant pour but de rendre des objets du quotidien intelligents (par exemple : un frigidaire intelligent, ce qui n'était pas existant dans le passé). Dans le contexte de la domotique, la sécurité et la protection de la vie privée des utilisateurs sont deux points importants qui ne sont pas forcément couverts au maximum du potentiel, selon les différents constructeurs et fabricants d'objets connectés.

1.2 Contexte et objectifs de la recherche

En prenant en compte, les inquiétudes que peuvent représenter le manque de sécurité ainsi que de solutions pour la protection de la vie privée dans un contexte comme l'IdO qui est introduit dans bons nombres de domaines d'utilisations, nous avons donc décidé de faire ce projet de maîtrise avec les trois aspects décrits précédemment dans le contexte, à savoir, la sécurité, la protection de la vie privée et les architectures dans le domaine de l'IdO.

L'objectif de ce mémoire est de mettre en avant notre proposition dédiée spécifiquement pour le domaine de la domotique de l'IdO. Nous proposons d'explorer la combinaison de différentes technologies au sein d'une même et seule architecture nommée IOTFLA afin d'essayer de pallier aux manques de sécurité et protection de la vie privée qui sont présents dans un scénario classique d'utilisation d'objets connectés au sein de l'IdO.

En débutant notre revue de littérature sur l'IdO, il nous est apparu judicieux d'associer notre solution à un domaine particulier de l'IdO, où la sécurité et la protection de la vie privée sont importantes. L'IdO et les objets connectés interviennent dans différents domaines, mais nous avons choisi de nous orienter vers le domaine des maisons intelligentes (aussi appelé parfois domotique) car cela nous semblait être l'un des domaines sous-jacents de l'IdO comportant le plus de risques de sécurité et de protection de la vie privée pour les utilisateurs qui sont des particuliers et qui ne sont pas forcément conscients des dangers que peuvent apporter les objets connectés au sein de leurs maisons.

Il nous est apparu qu'il y avait d'ores et déjà plusieurs propositions et solutions d'architecture dans le domaine de la domotique et nous avons décidé de définir notre solution sous la forme d'une architecture de domotique sécurisée et respectueuse de la vie privée, ce qui permet ainsi d'englober l'ensemble des participants au sein d'une maison intelligente (objets connectés, utilisateurs, réseau, protocoles, technologies, réseau externe, serveur d'infonuagique, etc.), au lieu de nous concentrer uniquement sur un seul aspect ou un objet en particulier.

Malgré l'environnement hétérogène proposé par l'IdO, un certain nombre de technologies sont utilisées par les objets connectés, comme les protocoles de communications, identifications, transports, données, sémantique, mécanismes de sécurité, etc. Dans ce cas, nous avons étudié les différentes technologies les plus utilisées et

adaptées pour la domotique. Ces technologies peuvent être mises en place dans une architecture dédiée à la domotique permettant d'apporter des solutions ayant pour but d'améliorer la sécurité globale du réseau privé, mais également de protéger les données des utilisateurs.

1.3 Méthodologie

La problématique et les objectifs du sujet de maîtrise ayant été posés précédemment, nous allons pouvoir décrire comment nous avons procédé pour répondre à ceux-ci.

Afin de pouvoir arriver jusqu'à la proposition de notre architecture IOTFLA pour répondre à la problématique ainsi qu'à nos objectifs, nous avons tout d'abord procéder par effectuer une étude bibliographique, permettant de constituer une base de connaissance sur la littérature scientifique dans le domaine de l'IdO. Les informations regroupées lors de ces recherches nous ont permis de créer et d'écrire un état de l'art sur le sujet de la sécurité et vie privée dans le monde des maisons intelligentes ainsi que plus généralement dans l'IdO et le réseau sans-fils.

En poursuivant les recherches il est devenu important pour nous d'orienter la recherche de la littérature sur les attaques, failles et différentes vulnérabilités liées aux réseaux sans-fils, à l'IdO et aux objets connectés utilisés dans les réseaux de domotique.

Nous nous sommes également intéressés à ce qui se faisait du côté des constructeurs, car ce sont eux qui produisent et mettent en services les objets connectés. En continuant nos recherches plus élargies sur le domaine de l'IdO, nous avons pu synthétiser nos idées afin de proposer une solution innovante. En explorant différentes possibilités et scénarios réalisables. Nous avons fait émerger l'idée de créer une architecture qui regrouperait différents composants dans le but de répondre

à notre problématique et d'atteindre nos objectifs de sécurité et de protection de la vie privée au sein de la domotique.

Plusieurs technologies nous sont apparues plus appropriées à notre vision de développement d'architecture visant à mettre en avant la sécurité et la protection de la vie privée. Notamment, l'AF (présenter dans la section 2.6.3), les SDI (détailler dans la section 2.6.1), les solutions d'automatisation de domotique (la section 3.3 montre la solution choisie) ainsi que l'ASD (illustrer dans la section 2.6.4).

Nous nous sommes dirigés vers l'utilisation de l'AF qui a pour but de mettre à jour des modèles (dans notre cas d'objets connectés) de manière à promouvoir la protection des données, mais également de garder au maximum en local les données. Les protocoles d'ASD sont bien étudiés dans les réseaux de capteurs sans fil et l'IdO avec de nombreux protocoles légers qui peuvent être adaptés au sein de notre proposition, de plus ils sont un élément de construction pour l'utilisation de l'AF. Les SDI ont pour but de détection tout comportement suspect à l'intérieur d'un réseau (dans notre contexte, un réseau de domotique), ce qui permet de pallier à certains types d'attaques.

La solution d'automatisation quant à elle permet non seulement du côté de l'utilisateur de la maison d'avoir une meilleure compréhension et gestion du réseau, mais également l'interaction des objets connectés. Cela permet également de pouvoir gérer les flux de données entre les objets connectés et l'extérieur ajoutant un niveau de contrôle supplémentaire. Dans notre cas, la solution d'automatisation devient le point central de l'architecture mise en avant dans ce mémoire.

1.4 Plan du mémoire

Cette section a pour but de présenter la structure et le découpage de ce mémoire. Le prochain chapitre (Revue de littérature) permet d'entrer plus réellement dans le sujet du mémoire en faisait un état de l'art de l'IdO et des différentes architectures ainsi que son marché et les acteurs majeurs.

Par la suite, nous nous intéresserons à la sécurité et la protection de la vie privée de manière approfondie dans la deuxième partie du chapitre (Enjeux en termes de sécurité et respect de la vie privée au niveau de l'IdO). Puis, la troisième partie du chapitre (Briques de constructions) met en avant trois technologies utilisées dans notre proposition, à savoir : les SDI, l'apprentissage machine (AM) et AF ainsi que les protocoles d'ASD.

La deuxième partie du mémoire consiste à mettre en avant notre proposition. Le chapitre 5 (Conception d'une nouvelle architecture - IoTFLA) présente le processus de mise en avant d'une proposition sous la forme d'une nouvelle architecture pour les maisons intelligentes. Par la suite, le chapitre 6 (IOTFLA - Principe de fonctionnement et proposition d'implémentation) présente l'ajout de l'AF ainsi que l'ASD à l'architecture présentée. Puis la conclusion vient clôturer ce mémoire.

CHAPITRE II

REVUE DE LA LITTÉRATURE

2.1 Présentation de l'internet des objets

C'est en 1999 que le paradigme d'Internet des Objets est apparu et utilisé pour la première fois par l'entrepreneur anglais Kevin Ashton employé de *Procter and Gamble*, ce dont il fait part dans sa publication dans le *RFID Journal*, une décennie plus tard (Kevin, 2009).

Il y décrit l'IdO, comme étant une *architecture émergente mondiale d'informations basée sur Internet, facilitant l'échange de biens, d'informations, mais également de services.* Elle a pour but de fournir une infrastructure informatique, qui permettra de surmonter l'écart entre les objets dans le monde physique et leurs représentations dans les systèmes informatiques.

Le potentiel offert par l'IdO rend possible le développement d'un très grand nombre d'applications. En donnant la possibilité de communiquer à des objets de la vie de tous les jours, l'IdO permet de déployer une pluralité d'applications dans différents environnements.

Ces environnements peuvent être regroupés sous forme de domaines (transport et logistique, santé, sports, environnement, social, personnel, sécurité ainsi que la domotique). Le domaine ciblé dans ce mémoire est la domotique.

Il existe beaucoup de définitions et de points de vue sur ce qu'est l'IdO, que ce soit du côté industriel ou du côté scientifique (Gubbi *et al.*, 2013). Néanmoins, je vais citer un seul article dont j'apprécie la clarté de leur définition, avant de vous donner ma propre définition de ce paradigme.

Selon (Atzori *et al.*, 2010), l'IdO, est un nouveau paradigme qui gagne rapidement du terrain dans le milieu des télécommunications sans-fils modernes. L'idée du concept réside dans l'omniprésence d'objets connectés (identification radio, capteurs, actionneurs, téléphones intelligents, etc.) qui peuvent interagir, communiquer et échanger des données avec d'autres objets présents dans l'environnement afin d'atteindre leurs objectifs.

La force de ce nouveau paradigme réside dans l'impact fort que les objets ont et qu'ils pourront avoir dans le monde moderne que nous connaissons. Les objets interviennent dans de différents et nombreux aspects de la vie de tous les jours. Les objets ont et continueront d'avoir un impact majeur dans de nombreux domaines.

Les auteurs (Atzori *et al.*, 2010) décrivent que le paradigme serait la convergence de trois principales visions, la première est une vision orientée objets ou choses, la seconde est une vision orientée Internet et enfin la dernière est une vision orientée sémantique (représenté dans la Figure 2.1).

Si je devais définir ce qu'est l'IdO, cela serait de la manière suivante : « *Il s'agit d'une infrastructure ou un écosystème mondial composé d'objets physiques au sein de notre environnement ayant la capacité de générer, d'utiliser et de transmettre des données sur celui-ci ainsi que d'agir et d'interagir dans notre quotidien.* ».

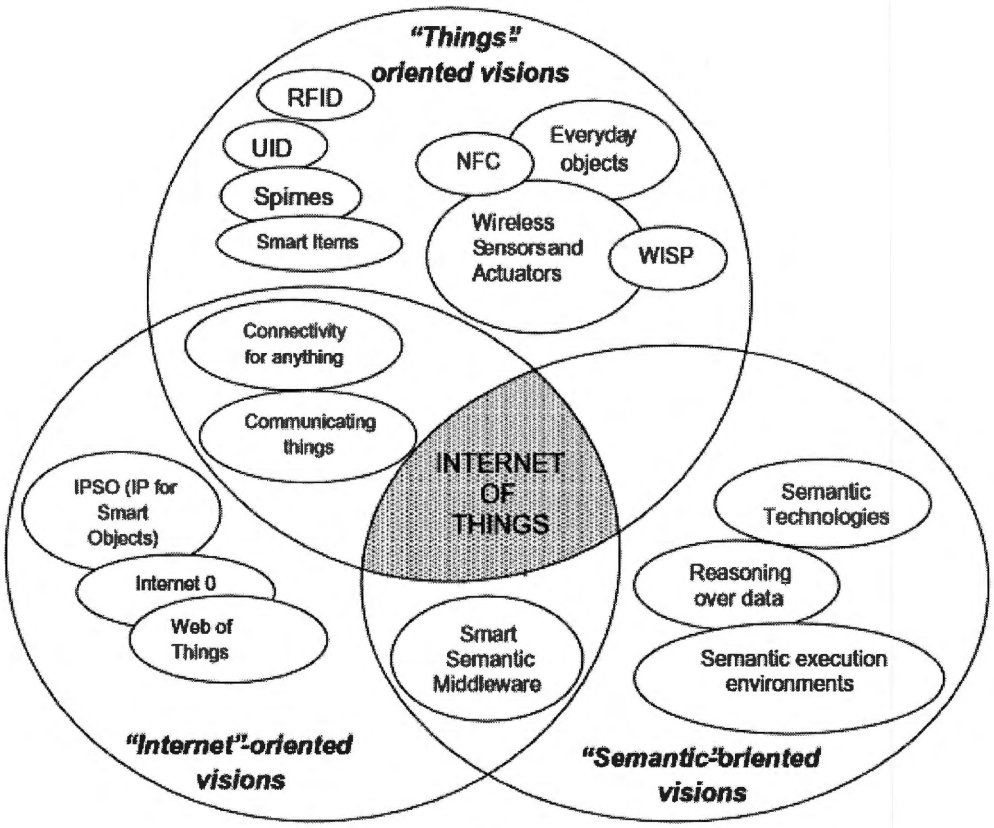

Figure 2.1 Internet des Objets : un paradigme, plusieurs visions (Atzori *et al.*, 2010)

2.2 Marché de l'IdO

Même si l'IdO existe depuis un bon nombre d'années, il continue et continuera à occuper de plus en plus de place dans notre quotidien. Différents groupes, entreprises, organismes ont décrit et estimé la place qu'occupera l'IdO dans les prochaines années et décennies. Afin de souligner l'importance que représente l'IdO de nos jours, mais également la place qu'ils prendront à l'avenir, voici quelques chiffres provenant de différentes sources (Cisco, ETH de Zurich, IDATE, Forbes

(For, 2017), Gartner, etc.).

L'entreprise *Cisco* estime que le nombre d'objets connectés sera bien au-delà des 50 milliards, ce qui représente environ 6,5 objets connectés par être humain pour l'année 2020. L'*IDATE* (IDA, 2016) quant à elle estime à environ 80 milliards pour 2020, ce qui est approximativement 10,5 objets par être humain. Selon une équipe de l'*ETH* de Zurich, le nombre d'objets connectés en 2025 serait d'environ 150 milliards, ce qui avoisine 20 objets par être humain. De plus, *Gartner* (Gar, 2017) a également donné une estimation de l'évolution des objets connectés au cours des dernières années avec une augmentation de 31 % de 2016 à 2017 et prévoit plus du double d'objets connectés pour 2020.

Ces estimations reflètent l'impact important que peut avoir l'IdO et des objets connectés sur la vie de tous les jours ainsi que les potentielles conséquences qu'il peut y avoir sur la sécurité et la protection de la vie privée des utilisateurs.

Afin de donner un peu plus de contexte aux chiffres sur l'IdO présentés dans la section précédente, nous vous présentons en amont des parties plus spécifiques de mon sujet, les différents acteurs majeurs, des concentrateurs ainsi que des solutions de sécurité au sein de la domotique en 2018 et 2019.

Au sein de l'IdO nous retrouvons des acteurs bien connus dans le domaine des technologies et du multimédia, tels que Google avec également la sous-branche Nest, incluant une panoplie d'objets pour les maisons intelligentes tels que des caméras, des thermostats et bien d'autres encore, Amazon avec comme produit phare Echo (et d'autres variations), Apple avec HomeKit composé également d'une grande gamme de produits allant de lumières (ampoules), aux caméras ou encore aux thermostats, Samsung avec la suite SmartThings (composé d'une gamme d'objets tels que : des concentrateurs, capteurs, boutons, etc.), Phillips qui est plus spécialisée dans l'éclairage et les accessoires et bien d'autres encore.

L'origine des concentrateurs provient de sa forme originelle qui est à la base un équipement de réseautique permettant d'inter-connecter physiquement plusieurs appareils, typiquement des ordinateurs (connexions réseau Ethernet) ou encore des périphériques, mais aussi parfois un commutateur ou un routeur. Dans le cadre de l'IdO, les concentrateurs sont également présents, en revanche ils ont comme but d'inter-connecter des objets connectés plutôt que des équipements de réseautique plus classiques.

Afin d'exprimer mieux le contexte du projet de ma maîtrise, il est important de situer le marché des concentrateurs (HUB) sur le marché professionnel, par exemple : Samsung SmartThings Hub, Wink Hub 2, Amazon Echo, Brillant, Control, Logitech harmony Home Hub, Google Home ou encore Vera Smart Home.

Il existe également des solutions de concentrateurs libres, comme cité ci-dessous :

- Home Assistant [1] est la solution utilisée dans notre proposition qui sera décrite plus en détail dans la section 3.3.
- Open HAB [2] est une plate-forme domotique libre, qui a pour but de fonctionner comme le centre d'une maison intelligente.
- Calaos [3] est un projet libre sous (GPLv3) et gratuit de domotique permettant de gérer et contrôler intelligemment sa maison.
- Domoticz [4] est un système d'automatisation de domotique permettant de surveiller et de configurer divers objets connectés tels que des lumières, des thermostats, des capteurs, etc.

---

1. https ://www.home-assistant.io/
2. https ://www.openhab.org/
3. https ://calaos.fr/fr/
4. http ://www.domoticz.com/

— Bitdefender Box [5] est probablement la solution qui s'approche le plus de ce que nous proposons dans ce mémoire. Il s'agit d'un concentrateur qui est concentré principalement sur la sécurité des objets connectés ainsi que de la protection des données.

Afin de continuer de faire le périmètre du mémoire, il est important de présenter l'IdO bien évidemment mais également les différentes architectures proposées ainsi que l'architecture générique.

2.3 Architecture de l'internet des objets

Même si la définition de l'IdO n'est pas vraiment universelle, son architecture en revanche est quant à elle est généralement acceptée et est basée sur les modèles de base d'Internet (TCP/IP et OSI) (Wu *et al.*, 2010).

L'architecture de base de l'IdO (représentée dans la figure 2.2) est composée de trois couches : une couche de perception (objets), une couche réseau (transport et traitement) et la couche application (services et applications).

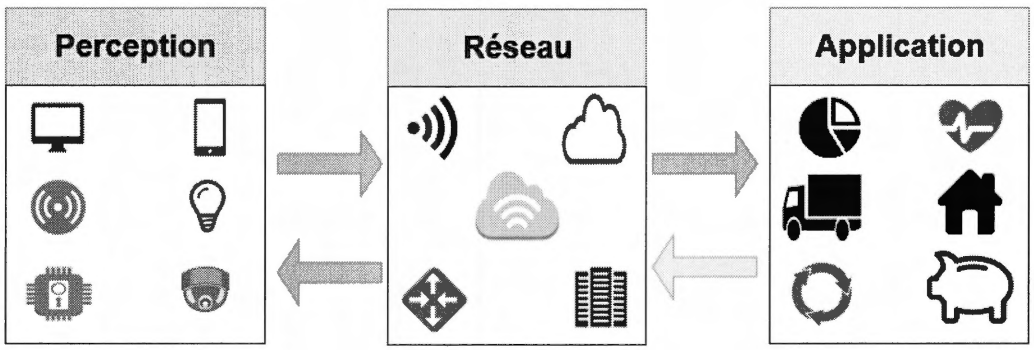

Figure 2.2 Architecture de l'IdO

5. https ://www.bitdefender.fr/box/

1. La *couche perception* est essentiellement composée des objets connectés et de la récolte d'informations. Elle comprend, des étiquettes et des lecteurs de code-barres à deux dimensions (2D), des étiquettes de radio-identification (RFID), des systèmes mondiaux de positionnement (GPS), des capteurs, des terminaux, des objets pouvant être qualifiés d'actionneurs ainsi que de capteurs et un réseau de capteurs. Cette couche de perception est chargée de convertir l'information en signaux numériques pour leurs transmissions sur le réseau.
2. La *couche réseau* est comme le système nerveux et le cerveau de l'IdO, sa fonction principale est de traiter les informations provenant de la couche perception. Elle est également chargée de la transmission des informations traitées à la couche application par le biais de diverses technologies de réseaux (WiFi, Bluetooth, Bluetooth Low Energy, Constrained Application Protocol, protocole de routage RPL, Message Queing Telemetry Transport, etc.).
3. La *couche application* utilise les données traitées par la couche précédente, elle constitue la partie de front de l'architecture, qui permet d'exploiter tout le potentiel des données. De plus, elle a pour rôle de fournir les outils nécessaires aux développeurs pour réaliser la vision et toutes les applications possibles de l'IdO.

2.4 Architectures proposées pour l'IdO

Dans la section précédente, nous avons vu que l'architecture de base de l'IdO était à la fois bien définie, mais aussi sujette à des modifications de la part des différentes parties prenantes du monde professionnel et du monde scientifique. En voici quelques exemples concrets :

Dans leur article (Wu *et al.*, 2010), les auteurs proposent une nouvelle architecture (illustré dans la figure 2.3) pour l'IdO nouvelle composée de cinq couches (Perception, Transport, Traitement, Application et Business). Le point clé de cette architecture réside dans le fait que les auteurs estiment que l'architecture à trois couches de base de l'IdO ne peut pas exprimer correctement et complètement toutes les caractéristiques et connotations de l'IdO (protocoles, technologies, etc.) et que de diviser les trois couches en cinq permet d'offrir une plus grande clarté et simplicité.

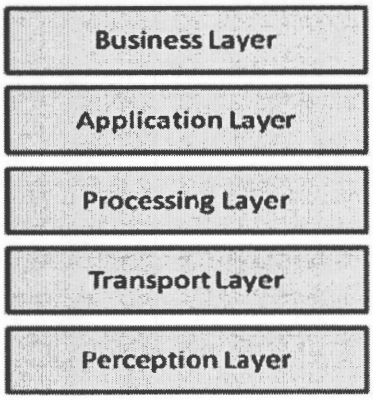

Figure 2.3 Architecture de (Wu *et al.*, 2010)

La couche de perception reste la même que celle de l'architecture générique, elle a donc les mêmes propriétés que celle décrite dans précédemment. En revanche, la couche réseau est divisée en deux, dans cette architecture la couche transport occupe la fonction de transmettre les informations reçues de la couche perception à la couche de traitement à travers différents réseaux, tels que sans-fil, câble ou encore réseau locaux (LAN). Les informations sont reçues par la suite par la couche de traitement qui a pour but principal de stocker, analyser et traiter celles-ci.

Nous retrouvons également dans cette architecture la couche application qui a un fonctionnement similaire à celle décrite dans la section précédente. Sa tâche est

basée sur les données provenant de la couche de traitement et permet de mettre en place diverses applications, telles que le transport intelligent, la gestion logistique, l'authentification d'identité, la sécurité, etc. En plus de ces quatre couches, la couche business au dernier niveau agit comme un gestionnaire de l'IdO. Ainsi, elle gère non seulement la mise à disposition et le chargement de diverses applications, mais aussi la recherche sur le modèle économique et le modèle de profit.

Dans leur article (kha,), les auteurs présentent également une nouvelle architecture (illustré dans la figure 2.4) qui se compose de cinq couches (Perception, Réseau, Intergiciel, Application et Business).

Dans ce cas, le point clé réside dans une explication détaillée des différentes couches de l'architecture et de leurs principaux composants (protocoles, technologies, etc.). De plus, les auteurs présentent les défis présentés par l'IdO ainsi que ses applications potentielles.

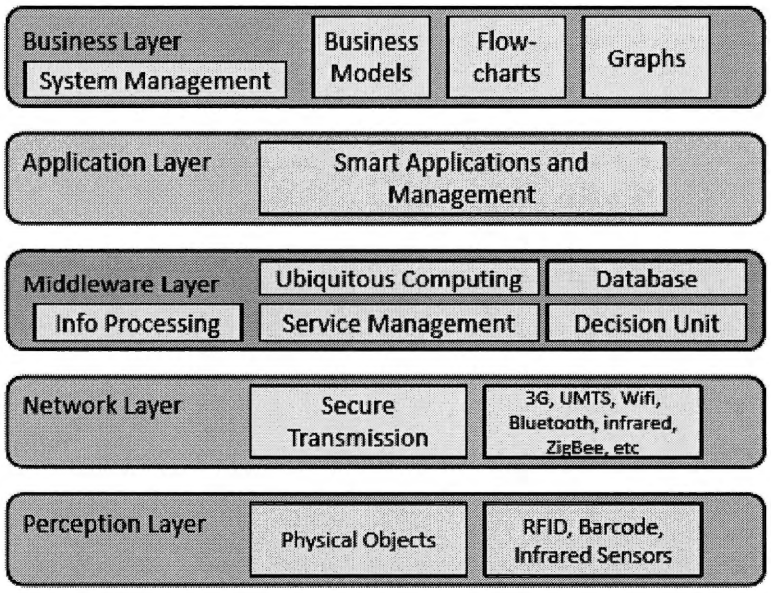

Figure 2.4 Architecture de (kha,)

Comme décrit précédemment, nous retrouvons une composition similaire avec les couches perception, réseau et application qui ont le même rôle. Mais les auteurs ont rajoutés dans ce cas, deux couches, tout d'abord la couche intergiciel qui est responsable de la gestion des différents services implémentés et utilisés par les objets connectés. Les informations sont reçues depuis la couche réseau et sont stockées dans une base de données, permettant ainsi de faire du traitement de celles-ci ainsi que de prendre des décisions automatiques basées sur ces traitements.

De manière similaire à l'architecture de (Wu *et al.*, 2010), il y a également une couche business dans cette architecture qui est responsable de la gestion plus globale du système de l'IdO incluant les services et les applications. Elle a pour objectif de créer des modèles d'affaires, ainsi que des graphes et des chartes de flux basés sur les données reçues. Dans leur article (Chen *et al.*, 2011), les auteurs proposent une architecture (représenter par la figure 2.5) se focalisant sur la sécurité qui est composée de quatre couches (Perception des données, Accès réseau hétérogène, Gestion des données et Services intelligents).

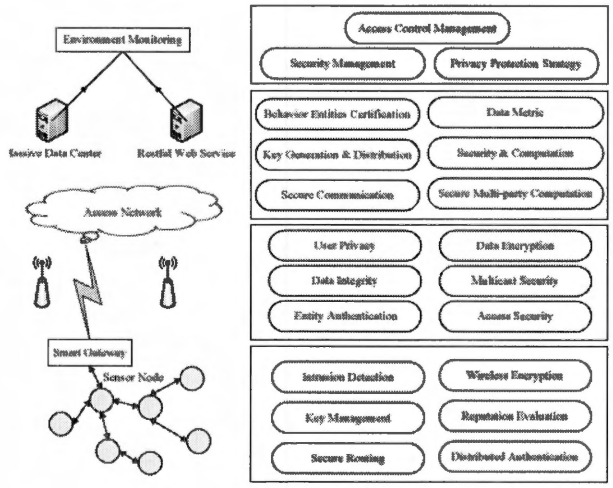

Figure 2.5 Architecture de (Chen *et al.*, 2011)

Chacune des couches aborde différents défis de sécurité liés à l'IdO, la première qui est la couche de perception qui a le même fonctionnement que présenté précédemment sauf que dans ce cas, les auteurs parlent des défis spécifiques à cette couche qui comporte le routage sécurisé, la gestion de clés, la détection d'intrusion, le chiffrement au sein du réseau sans-fil ainsi que l'évaluation de réputation. La seconde couche est la couche accès au réseau hétérogène qui a pour principaux enjeux la sécurité, la protection de la vie privée de l'utilisateur, le chiffrement et l'intégrité des données, la sécurité des accès, l'authentification d'entité ainsi que de la sécurité à multidiffusion (*Multicast* en anglais).

Par la suite, la couche de gestion des données se concentre sur la certification des entités, les données, la génération et distribution de clés, le calcul de sécurité, la communication sécurisée ainsi que le service de calcul multi parti sécurisé. La couche de services intelligents quant à elle aborde les défis de sécurité suivants : la gestion du contrôle des accès, la gestion de la sécurité et les stratégies de protection de la vie privée.

Dans leur article (Yang *et al.*, 2012), les auteurs proposent une architecture (représenter dans la figure 2.6) se focalisant sur la sécurité. Elle est composée des trois couches de base de l'IdO (Perception, réseau et application) et cherche à résoudre les défis liés à la protection de la confidentialité des données en ajoutant quelques intergiciels tels que des serveurs tiers, des mécanismes de chiffrement et déchiffrement, la gestion des accès, permettant ainsi d'établir une architecture plus sécurisée que celle de base de l'IdO.

Cinq enjeux de sécurité sont mis en avant dans l'article. Le premier est le renforcement de la protection du domaine local, visant à gérer et protéger les identités des balises (*Tags* en anglais) à un certain niveau afin de pouvoir transmettre les données du domaine local. Les données sont chiffrées par l'algorithme de chiffrement

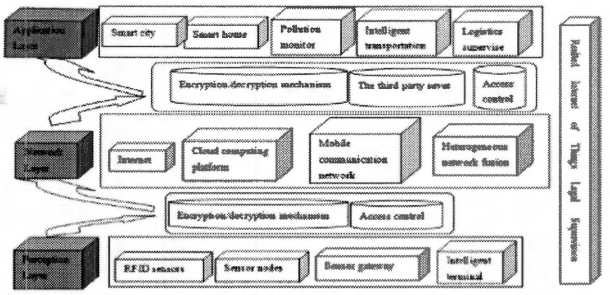

Figure 2.6 Architecture de (Yang *et al.*, 2012)

3DES (Nadeem et Javed, 2005) (qui est le remplacent de DES qui est le standard de chiffrement des données) lorsqu'elles sont envoyées vers une autre zone locale.

Le second point abordé est de faire en sorte que les différentes couches aient des permissions différentes grâce à des contraintes d'authentification afin de permettre le renforcement de la dynamique de la protection de la vie privée. Troisièmement, il faut ajouter un tiers de confiance (comme un serveur tiers) qui peut être considéré comme un module de surveillance permettant de filtrer les données avant qu'elles n'accèdent à la couche application, elles sont donc filtrées par le tiers pour s'assurer que les données sensibles n'ont pas été volées ou modifiées.

Le quatrième enjeu est le traitement anonyme des données relatives à la vie privée des utilisateurs selon un modèle de contrôle de la sécurité et de la protection de la vie privée. Seuls les utilisateurs légitimes ayant passé l'autorisation peuvent voir ces informations traitées. Dernièrement, la chose la plus fondamentale est d'améliorer la qualité de l'ensemble de la population afin de réduire le comportement illégal d'obtention de données sensibles dans la mesure du possible.

Les architectures citées précédemment fournissent un tour d'horizon sur le fonctionnement et la constitution des architectures de l'IdO. Les deux prochains travaux présentés sont les architectures qui s'approchent plus de l'architecture qui

est proposée dans ce mémoire.

Les auteurs (Simpson *et al.*, 2017) proposent dans ce cas, un gestionnaire central de la sécurité qui est construit sur un concentrateur (Hub) ou sur la passerelle du routeur et est positionné pour intercepter tout trafic en provenance et à destination des objets connectés. De plus, la plate-forme de gestion de la sécurité se trouve au sommet du logiciel de communication et déploie des modules qui exécutent des fonctionnalités de sécurité spécifiques.

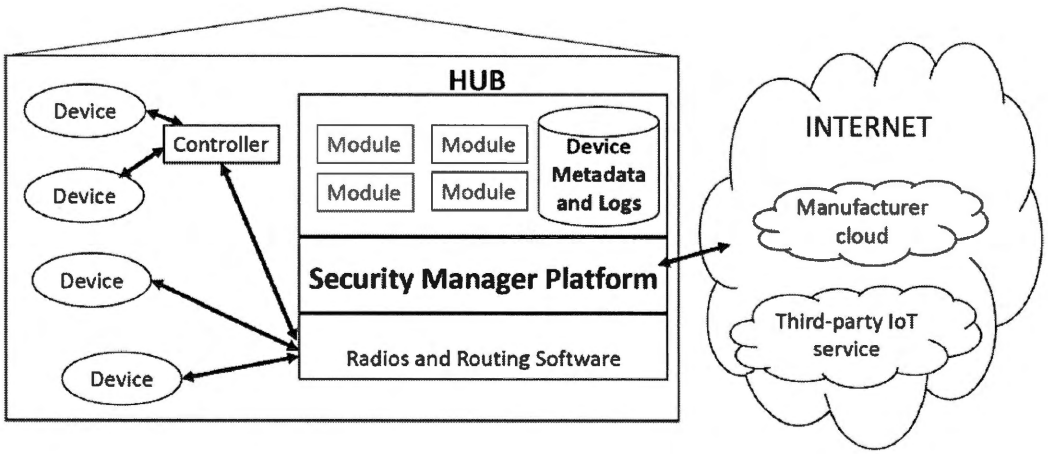

Figure 2.7 Architecture de (Simpson *et al.*, 2017)

L'architecture (figure 2.7) se base tout d'abord sur quatre concepts, le premier étant l'interception des communications ayant le concentrateur en tant que gestionnaire de celles-ci. Le second est d'avoir une vue globale sur tout objet connecté au sein de la maison. Troisièmement, la pré installation ainsi que l'installation des mises à jour des objets connectés. Enfin, le dernier concept mis en avant est de fournir de la flexibilité et de pouvoir ajouter de nouveaux objets à souhait dans le réseau local. En plus de ses concepts, le concentrateur intègre quatre modules de sécurité pour différentes phases au sein d'un cycle de vie de vulnérabilités. Les quatre phases sont les suivantes : vulnérable sans correctif, acquisition d'un

correctif, application d'un correctif et objet compromis.

Dans le dernier cas, les auteurs (Ibrahim *et al.*, 2017) proposent un cadriciel (figure 2.8) pour l'IdO se basant sur les protocoles de sécurité de la couche de transport (TLS/SSL). En élaborant ce cadre, ils ont programmé des capteurs grâce à une carte Arduino et ont géré les données à travers un stockage dans un cloud pour accéder aux objets connectés dans la maison intelligente.

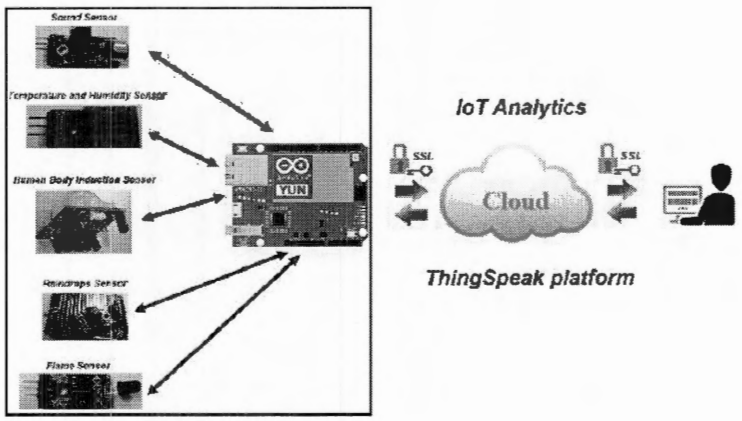

Figure 2.8 Architecture de (Ibrahim *et al.*, 2017)

2.5 Enjeux en termes de sécurité et respect de la vie privée au niveau de l'IdO

Nous avons présenté dans les sections précédentes l'IdO et les différentes architectures existantes qui sont utiles à la compréhension de ce mémoire, la sécurité et la protection de la vie privée qui seront traités dans ce chapitre.

Afin d'avoir une meilleure vue d'ensemble sur les problèmes de sécurité dans le contexte de l'IdO et plus particulièrement en domotique, ce chapitre se concentrera essentiellement sur les différents défis, contraintes, attaques, vulnérabilités et failles recensées dans la littérature à ce jour.

Bien que la sécurité soit l'un des aspects importants dans le contexte des technologies de l'information, l'IdO présente des défis de sécurité nouveaux et uniques. Répondre à ces défis et assurer la sécurité des produits et des services est devenue un enjeu crucial. Plus spécifiquement dans le domaine des maisons intelligentes (domotique), les consommateurs/utilisateurs doivent pouvoir faire confiance aux objets connectés ainsi qu'aux services utilisés, mais non pas forcément la compréhension des risques pour leurs données privées.

Défis de sécurité dans l'IdO

Dans ces articles (Atzori *et al.*, 2010; Khan *et al.*, 2012; Conti *et al.*, 2018; Mathov *et al.*, ; Neshenko *et al.*, 2019), les auteurs décrivent les premiers principaux domaines de recherche importants ainsi que des défis de l'IdO. Pour rester dans la thématique du mémoire, cette section couvre seulement les défis liés à la sécurité.

1. *Identification et authentification.* L'IdO étant composé de milliards d'objets, chaque objet à besoin d'avoir sa propre identité, ce qui implique d'avoir un système d'identification qui peut assigner dynamiquement et gérer des identités uniques pour un grand nombre d'objets ainsi que les accréditations associées. L'authentification est difficile dans un contexte tel que l'IdO car elle nécessite des infrastructures d'authentification appropriées qui ne sont pas adaptées, principalement en raison du manque de ressources des objets connectés.
2. *Sécurité du réseau.* Le système de transmission des données doit être capable de gérer les données provenant d'un grand nombre d'objets sans causer de pertes à cause de la congestion du réseau, de garantir des mesures de sécurité appropriées pour les données transmises et de les protéger des interférences externes ou de tentatives éventuelles de surveillance.

3. *Intégrité, confidentialité et chiffrement des données.* Les informations sont généralement protégées par des mécanismes de chiffrement ainsi que des mots de passe. Dans le cas des mots de passe, la taille de ceux-ci supportés par les technologies de l'IdO ne permettent pas forcément de fournir un niveau de sécurité satisfaisant. Les objets récoltent, prennent des mesures et transfèrent des données, il est donc nécessaire que les objets possèdent un ou plusieurs mécanismes de chiffrements des données suffisamment fort pour garantir la confidentialité des données transmises. Ceci est complexe dans le domaine de l'IdO en raison de différentes capacités des objets connectés et des contraintes énergétiques également.
4. *Sûreté et sécurité physique des objets.* L'IdO est composé d'un très grand nombre d'objets de perception, il est donc nécessaire d'empêcher l'accès d'intrus aux objets qui peuvent causer des dommages physiques ou peuvent modifier leurs fonctionnements.
5. *Oubli numérique / non-effacement des données.* Toutes les informations recueillies au sujet d'un ou plusieurs utilisateurs par les objets connectés peuvent être conservées indéfiniment à partir du moment où le stockage le permet. De plus, les techniques d'investigation numérique peuvent être utilisées pour récupérer facilement toutes ces informations même après plusieurs années. Ce défi n'est pas spécifique à l'IdO de manière générale, mais il reste un défi important dans le domaine de l'IdO.

Défis par couches de l'architecture

Les défis cités dans la section précédente peuvent impacter les différentes couches qui forment l'architecture classique de l'internet des objets.

1. *Couche perception.* Pour la couche perception, les défis de sécurité consistent

principalement en l'authentification, la protection des données sensibles (confidentialité), l'évaluation des risques et la détection d'intrusion.

2. *Couche réseau.* En ce qui concerne la couche réseau, l'authentification, la confidentialité des données ainsi que la détection d'intrusion sont également des facteurs à prendre en compte ainsi que la sécurité de routage.
3. *Couche application.* Quant à la couche application, nous retrouvons également l'authentification, la détection d'intrusion, l'évaluation des risques et la sécurité des données.

Contraintes liées aux objets connectés

Afin d'avoir une meilleure vue d'ensemble sur les défis que représente l'IdO, il faut également prendre en compte les contraintes liées aux objets connectés. Les objets ne possèdent pas les mêmes caractéristiques le matériel informatique « standard ». En effet pour la grande majorité, ils ne possèdent pas une grande puissance de calcul, ont une faible capacité de mémoire et ont un budget énergétique limité. De plus, il faut prendre en compte deux aspects liés aux réseaux, à savoir la taille des noeuds et la bande passante de communication. Des explications plus détaillées sur les types d'attaques, failles, vulnérabilités ainsi que les attaques en elles-mêmes seront développées dans la section suivante.

2.5.1 Attaques existantes au sein des réseaux sans-fils et de l'IdO

Il est possible de classer les attaques de l'IdO selon différentes dimensions (méthodes, attaques, vulnérabilités, couches, orientation, risques, actives, passives, internes ou encore externes), comme nous pouvons le constater dans les articles (Jing *et al.*, 2014; Ziegeldorf *et al.*, 2014; Farooq *et al.*, 2015; Padmavathi et Shanmugapriya, 2009).

1. *Passives.* Les attaques passives sont caractérisées par l'interception de messages sans aucune modification ou encore à l'accès au contenu de la mémoire d'un objet. Il n'y aucun changement effectué directement sur les données du réseau ou du système.
2. *Actives.* Les attaques actives sont les attaques dans lesquelles un ou des changements non autorisés du système (ou réseau) sont menés par un attaquant. Par exemple, cela pourrait concerner la modification des données transmises ou stockées ou encore la création de nouveaux flux de données.

Catégories d'attaques

Les attaques peuvent être séparées en trois catégories. Ci-après, nous avons donné à chacune de ces catégories un exemple d'attaque dans le domaine de l'IdO afin de l'illustrer. Les attaques citées dans le schéma 2.9.

1. *Prise de contrôle.* L'une des trois catégories est la prise de contrôle d'un objet. En effet, il est possible qu'un attaquant puisse prendre le contrôle d'un objet que ce soit une caméra pour espionner, un hub pour contrôler plusieurs objets d'un même réseau, une voiture ou encore un drone.

   La technique de *Spoofing* par exemple (Farooq *et al.*, 2015), représente le cas d'un attaquant diffusant de fausses informations sur plusieurs systèmes dont notamment les systèmes RFID et se sert du contenu pour faire semblant d'être un appareil authentique, ce qui le fait apparaître comme provenant de la source d'origine. De cette façon, l'attaquant obtient un accès complet au système, ce qui le rend vulnérable.
2. *Vol de données.* La seconde catégorie est le vol de données. Dans ce cas, le but de l'attaquant est d'accéder à des informations que ce soit des données de personnes ou d'entreprises. Par exemple, le concept de l'attaque de

l'homme du milieu (Farooq *et al.*, 2015) a pour but d'intercepter les communications entre deux parties, ce qui est dans notre cas soit deux objets, un objet et une application, un objet ou l'infonuage ou encore un objet et l'utilisateur.

3. *Interruption de services.* La dernière catégorie est l'interruption de services. L'objectif est ici de rendre un service indisponible empêchant ainsi les utilisateurs d'accéder au service. Le *Botnet Mirai* est un exemple concret de cette catégorie (Kolias *et al.*, 2017), il a pour but de provoquer un DDOS contre un ensemble de serveurs cibles en se propageant grâce des objets connectés ayant un faible niveau de sécurité.

2.5.2 Classification des attaques

Cette section est consacrée à la présentation des différentes attaques classifiées dans le schéma 2.9.

Combinaison de prise de contrôle, vol de données et interruption de service

— Les *attaques physiques* représentent le cas d'un attaquant ayant un accès physique à un objet. Il peut alors choisir de modifier l'objet lui-même, ce qui lui permettrait d'accéder au réseau ainsi qu'aux flux de données de celui-ci.

— L'*attaque de gouffre (Sinkhole)* est une attaque qui a pour but d'amener les nœuds voisins à partager des informations à un noeud corrompu avec des informations trompeuses sur le chemin de routage. Ce qui a pour conséquence que les flux de données provenant d'un nœud particulier sont déviés vers le nœud compromis. Le trafic est alors réduit au silence pendant que le système est trompé, croyant que les données ont été reçues d'un nœud non compromis. De plus, cette attaque entraîne une consommation d'énergie

plus élevée qui peut provoquer un déni de service (DOS).

— L'*attaque par rejeu* permet à l'attaquant d'enregistrer et de stocker des données transmises précédemment pour répéter ses données plus tard ou retarder la session en cours.

— La *capture de nœuds* correspond à la prise de contrôle d'un nœud sur le réseau par l'attaquant. Par la suite, cette situation peut permettre de contaminer d'autres noeuds.

— Les *rançonciciels* sont des logiciels malveillants ayant pour objectif de prendre en otage les données personnelles d'un utilisateur. Ils chiffrent les données personnelles de l'utilisateur et demandent à celui-ci une rançon contre l'envoi de la clé qui permettra à l'utilisateur de déchiffrer et de récupérer ses données.

— L'*accès aux étiquettes RFID* est parfois possible en raison de l'absence d'authentification appropriée. En effet, il existe un grand nombre de systèmes RFID qui peuvent être accessibles par une personne sans autorisation. Au-delà de simplement lire les données, l'attaquant pourrait aussi possiblement les modifier ou les supprimer.

— Le *clonage d'étiquette (Tags)* est possible lorsque les étiquettes sont déployées sur différents objets qui sont visibles et leurs données peuvent être lues et modifiées avec certains types d'attaques. Dans cette situation, ils peuvent facilement être capturés par un attaquant qui peut alors créer une réplique de l'étiquette et donc la compromettre d'une manière dont le lecteur ne peut pas distinguer l'original de la contrefaçon.

— L'*attaque Byzantine* se réfère à une classe d'attaques dans laquelle l'adversaire a le contrôle d'un noeud authentifié au sein d'un réseau lui permettant d'agir de manière arbitraire.

— L'*attaque de répudiation* survient lorsqu'une application ou un système n'adopte pas de contrôle pour suivre et enregistrer correctement les ac-

tions des utilisateurs, ce qui permet une manipulation malveillante ou une identification de nouvelles actions. Elle peut être utilisée pour modifier les informations de création d'actions exécutées par un utilisateur malveillant afin de consigner de mauvaises données dans les fichiers journaux.

— L'*injection de code malveillant* est une attaque dans laquelle un attaquant compromet un nœud afin d'injecter un code malveillant dans le système, ce qui pourrait même entraîner un arrêt complet du réseau ou d'obtenir son contrôle total.

— Un *initié malveillant* est un individu se situant à l'intérieur du réseau qui manipule les données pour un gain personnel ou les avantages d'une tierce partie. Les données peuvent être facilement extraites et ensuite modifiées à partir de l'intérieur.

— Les *maliciels* sont des logiciels malveillants développés dans le but de nuire à un système sans le consentement de l'utilisateur dont le système est infecté.

— L'*attaque de fabrication* est un type d'attaque dans laquelle l'attaquant insère des objets compromis dans le système, sans la connaissance ou l'implication de l'expéditeur. Ainsi, fondamentalement, l'utilisateur ne sait pas son système a été compromis. Cette attaque implique la création, la modification et la suppression non autorisées d'informations, de systèmes informatiques et d'éléments de réseau.

— *Hello Flood* est une attaque qui a pour objectif de neutraliser les nœuds du réseau, permettant ainsi de dégrader les performances du réseau cible et finalement de diviser la grille du réseau, afin de prendre le contrôle d'une partie du réseau de capteurs en insérant un nouveau nœud.

— Un *Zombie* correspond à un objet ou produit qui n'est plus maintenu à jour par le constructeur. Cet objet ou produit peut être utilisé pour une durée indéterminée et risque d'apporter de potentielles failles et vulnérabilités.

Un zombie offre à l'attaquant un point d'entrée sur le réseau lié à l'objet.

— Le principe de *la contamination* est qu'un objet possède un défaut permettant à un attaquant de l'exploiter. Le défaut de cet objet permet ainsi à accéder au réseau lié à l'objet, ce cas reste semblable à celui d'un zombie, sauf que dans ce cas cela peut être un objet qui est toujours maintenu.

Prise de contrôle et vol de données

— L'*attaque Sybil* est une attaque dans laquelle l'attaquant présente des identités multiples et en les utilisant pour avoir une influence disproportionnée pour un seul nœud, ce qui peut entraîner la compromission d'une partie considérable du système. La vulnérabilité face à une attaque Sybil pour un système de réputation dépend de la facilité de génération de nouvelles identités, de la facilité du système de réputation à accepter l'entrée d'entités n'ayant pas de lien de confiance avec des identités de confiance et si le système de réputation traite toutes les entités de manière identique.

— L'*attaque par force brute* est une attaque qui consiste à tester toutes les combinaisons possibles pour découvrir une clé ou un mot de passe. Dans le contexte de l'IdO, certains dispositifs ayant une puissance de calcul limitée, cela les amènent à utiliser des secrets de taille suffisamment petite pour faciliter une attaque par force brute.

— *Attaque par imitation.* L'authentification dans l'environnement distribué de l'IdO est très difficile, car elle nécessite des infrastructures d'authentifications qui sont difficile à déployées principalement en raison du manque des ressources des objets connectés. Ce qui peut permettre à des nœuds malveillants d'utiliser une fausse identité pour mener une attaque malveillante ou de collision qui agit sur une fonction de hachage cryptographique ayant pour but de trouver deux entités de cette fonction qui produisent la même

valeur.

— L'*attaque « trou de ver » (Wormhole)* consiste dans le fait qu'un ou plusieurs nœuds créent une fausse route qui est plus courte que l'originale dans le réseau, ce qui rend confus les mécanismes de routage qui reposent sur la connaissance de la distance entre les nœuds.

— Dans l'*attaque « On-Off »*, un attaquant tente de perturber le système de réputation en place en se comportent bien et mal de manière alternative de telle sorte que confiance est toujours regagnée juste avant qu'une autre attaque se produise. La plupart des programmes de rachat de confiance *(trust redemption schemes)* ne parviennent pas à discriminer entre une attaque on-off et des erreurs temporaires, spécialement lorsque la majorité du comportement de l'attaquant est bon.

— Dans une attaque de *Spoofing*, l'attaquant diffuse de fausses informations sur des systèmes (notamment RFID dans le contexte de l'IdO) et se sert du contenu pour faire semblant d'être authentique, ce qui le fait apparaître à partir de la source d'origine. De cette façon, l'attaquant obtient un accès complet au système le rendant ainsi vulnérable.

— La méthode du *trou noir* se réfère aux places dans un réseau où le trafic entrant ou sortant est discrètement en train de disparaître, sans informer la source du trafic.

— Une attaque par *clonage de nœuds* exploite le fait que la structure matérielle d'un noeud pouvant être simple, il est possible par conséquent qu'elle soit facilement copiée par l'attaquant.

Vol de données

— L'*ingénierie sociale* est l'acte de manipuler les gens afin qu'ils donnent accès à une information sensible tel qu'un mot de passe ou des informations

bancaires.

— Dans une *attaque de première main*, l'attaquant obtient des informations directement de l'utilisateur. Ainsi une divulgation d'information pourrait se produire accidentellement suite par exemple à la présence d'une faille dans la sécurité du système qui entraîne une fuite d'informations. Il également possible de tromper l'utilisateur en utilisant des approches du type d'ingénierie sociale.

— Une *attaque par commérage* consiste en une transmission de données personnelles d'une entité autorisée à une entité non autorisée. La principale différence avec l'attaque de première main est que l'utilisateur n'est plus la source directe de fuite des informations. En effet, la divulgation d'informations peut avoir lieu sans que l'utilisateur ciblé n'en soit conscient.

— Une *attaque d'homme du milieu* a pour but d'intercepter les communications en s'interposant entre deux entités. Dans notre cas, il peut s'agir soit de deux objets, un objet et une application, un objet ou le nuage, ou encore un objet et l'utilisateur.

— Une *attaque par hameçonnage* est une attaque de type *spoofing* se basant sur le courrier électronique. Dans cette attaque, la victime est attirée par l'ouverture d'un courrier électronique par lequel l'attaquant accède aux informations d'identification de cette victime ou d'autres informations plus sensibles.

— Une *attaque de falsification* concerne la modification des données par un attaquant.

— Une *attaque de sessions* est une technique consistant à intercepter une session initiée entre deux machines afin de la détourner.Dans la mesure où le contrôle d'authentification s'effectue uniquement à l'ouverture de la session, un attaquant réussissant cette attaque parvient à prendre possession de la connexion pendant toute la durée de la session.

— Dans une *attaque par observation*, l'attaquant essaye de configurer des objets connectés pour collecter des informations sur leurs environnements. En effet, une des caractéristiques importantes de l'IdO est la capacité des objets à observer et à ressentir leurs environnements.

— L'objectif d'une *attaque par inférence* est d'inférer des informations personnelles telles que les activités, la mobilité ou d'autres informations sensibles d'une entité en utilisant les données recueillies à partir d'autres attaques.

— Une *attaque par invasion automatisée* peut être menée dans l'IdO, après avoir recueilli de grandes quantités d'informations, en utilisant une des attaques précédemment citées. Par exemple, un système automatisé peut combiner les données pour effectuer une exploration de données ou une analyse permettant de conduire à un nouveau type d'attaques.

— Une *attaque par écrémage (Skimming)* consiste à lire rapidement des messages transmis pour collecter des données. Par exemple, cette attaque est principalement utilisée pour les cartes de crédits avec un *Skimmer* déposé sur les distributeurs bancaires.

— Une *attaque d'inventaire* se réfère à la collecte non autorisée d'informations sur un utilisateur. Bien qu'il soit normal que les utilisateurs légitimes du système puissent interroger les objets qu'ils possèdent, des entités non désirées pourraient aussi les interroger afin de compiler des données permettant de former une liste d'inventaire.

— Une *attaque par écoute (Eavesdropping)* est possible en raison des caractéristiques sans fil de certaines technologies (notamment RFID) où il est relativement facile pour l'attaquant de renifler des informations confidentielles, comme les mots de passe ou autres données qui passent par d'une étiquette à un lecteur ou d'un lecteur à une étiquette. Ce concept d'écoute est également applicable aux objets connectés.

— Une *attaque par analyse du trafic* consiste à surveiller le trafic afin que

l'attaquant puisse déterminer des patterns récurrents. Cette attaque lui permettrait par exemple de déterminer les types de données qui circulent dans le réseau.

Interruption de service

- Une *attaque par déni de service (DOS ou Denial of Service)*, a pour but de rendre indisponible un service, ce qui permet d'empêcher l'accès à un ou plusieurs utilisateurs légitimes d'un service. DOS est une attaque provenant d'une source unique.
- Une *attaque par déni de services distribué (DDOS ou Distributed Denial of Service)* suit le même schéma que l'attaque DOS sauf pour le fait que l'attaque provient de multiples sources.
- Dans une *attaque par privation de sommeil (Sleep Deprivation)* utilise le fait que les nœuds d'un réseau de capteurs sans fil étant alimentés avec des batteries, ils ont besoin de suivre une routine de sommeil permettant de prolonger leur durée de vie. La privation de sommeil est une attaque qui maintient les nœuds éveillés, ce qui entraîne une plus grande consommation de batterie et par conséquent, diminue aussi la durée de vie des batteries entraînant ainsi l'arrêt des nœuds.
- Dans *attaque par brouillage de fréquence radio*, les étiquettes RFID sont compromises par une sorte d'attaque par déni de service dans laquelle la communication à travers les signaux radio sont perturbés par un excès de signaux de bruits.
- Une *attaque d'épuisement des ressources* est une exploitation de sécurité informatique qui bloque ou interfère avec le programme ou le système ciblé. C'est donc une forme d'attaque par déni de service.
- Une *attaque par reniflage (Sniffing)* correspond au vol ou à l'interception de

données en capturant le trafic réseau à l'aide d'un renifleur (une application visant à capture des paquets réseau).

— Dans l'*attaque « Gray Hole »*, un nœud malveillant de l'attaquant agissant comme un nœud normal efface les messages ou paquets qui le traverse, permettant ainsi de cacher les informations importantes à transférer au prochain nœud.

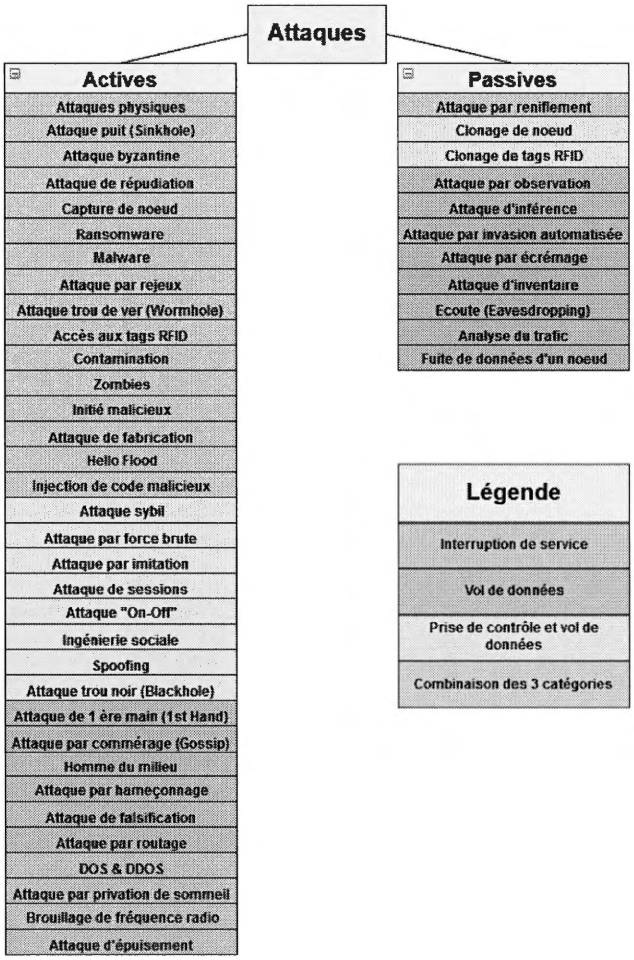

Figure 2.9 Classification des attaques liées à l'IdO

Nous avons réalisé une classification des attaques qui sont illustrées dans la figure 2.9.

2.5.3 Respect de la vie privée et protection des données

La notion de respect de la vie privée est un concept très ancré dans notre civilisation et qui est reconnu comme droit fondamental dans les législations de nombreux pays. L'accomplissement du potentiel de l'IdO va dépendre des stratégies qui respectent les choix individuels en termes de respect de la vie privée.

Les flux de données et les fonctionnalités offertes par les objets permettent d'offrir des services pour les utilisateurs qui n'existaient pas auparavant, mais les préoccupations concernant la protection des données personnelles et les répercussions potentielles de l'utilisation de ces données sont des enjeux majeurs à prendre en compte.

L'IdO redéfinit le débat sur les problèmes de confidentialité et la protection des données personnelles, car la collecte, l'analyse et l'utilisation des données sont au coeur de l'IdO. Cette situation amplifie les inquiétudes quant au potentiel d'une surveillance accrue, de la difficulté à pouvoir exclure certaines collectes de données et de la force de l'agrégation des flux de données pour façonner des portraits numériques d'utilisateurs détaillés.

Bon nombre d'utilisateurs n'hésitent pas à acheter des objets connectés pour des raisons de commodités. Mais pour le grand public, il est difficile de comprendre la portée de ce que peuvent collecter les objets connectés sur l'environnement et les utilisateurs eux-mêmes ce qui peut avoir un impact majeur sur leur vie privée. Il est également difficile pour un utilisateur de savoir quelle est l'option offrant le meilleur compromis fonctionnalité / respect de la vie privée avec la multitude de fabricants agissants sur le marché de l'IdO et de la domotique, ainsi que les produits qui ont les mêmes fonctionnalités (caméras, thermostats, concentrateurs, verrou, etc.). Ainsi, il est possible qu'un bris de vie privée apparaisse suite au choix

d'un objet ne garantissant pas une sécurité assez bonne ou moins développée que celle proposée par des concurrents.

Dans l'article (Ziegeldorf *et al.*, 2014), les auteurs font un survol des dangers liés à la vie privée ainsi que la confidentialité dans le contexte de l'IdO. Ainsi, à partir d'une seule faille sur un réseau ou sur un objet, il est possible d'identifier, de localiser, de suivre, de profiler et d'associer une personne a une ou plusieurs informations récoltées. Voici quelques exemples de risques :

1. *L'identification* consiste à associer un identifiant (persistant), tel qu'un nom, prénom ou adresse d'un utilisateur avec des données le concernant. La menace réside donc dans l'association d'une identité à un contexte spécifique et à des facteurs générateurs d'autres menaces.
2. La *localisation et le suivi* sont des attaques visant à déterminer l'emplacement d'une personne dans le temps et l'espace. Le suivi nécessite une identification afin de lier des localisations différentes à un utilisateur.
3. Le *profilage* consiste à rassembler des informations sur les utilisateurs afin de déduire leurs intérêts en les corrélant avec d'autres données. Les méthodes de profilage sont principalement utilisées pour la personnalisation en commerce électronique, mais aussi pour une optimisation interne basée sur les caractéristiques et les intérêts des clients/utilisateurs.

   Une *attaque d'inventaire* se réfère à la collecte non autorisée d'informations sur l'existence et les caractéristiques d'objets connectés. Par exemple, une personne non autorisée peut interroger et exploiter une faille dans le but de dresser une liste des objets et composants d'un réseau (entreprise, maison, etc.).
4. La protection de la vie privée est menacée lorsque les objets connectés révèlent des informations sensibles pendant les changements de périodes

de contrôle dans leur *cycle de vie.* Le problème a été observé directement en ce qui concerne les photos et vidéos compromettantes qui se trouvent souvent sur les caméras ou téléphones intelligents usagés.

5. La *liaison (Link)* consiste à relier différents systèmes précédemment séparés, de sorte que la combinaison de sources de données révèle des informations permettant de faire des inférences sur l'utilisateur.

2.5.4 RGPD : La régulation générale européenne de la protection des données

Après quatre ans de préparation et de débat, le RGPD[6] (régulation générale européenne de la protection des données) a été approuvé par le parlement européen le 14 avril 2016, puis mise en pratique à partir du 25 mai 2018. Le but du RGPD est de protéger tous les citoyens de l'Union Européenne contre les atteintes à la vie privée et d'assurer la protection des données personnelles. Les principes fondamentaux de la protection des données restent inchangés par rapport à la directive précédente (Directive sur la protection des données 95/46/EC), en revanche de nombreux changements on été proposés dans les aspects réglementaires :

— *Pénalités.* Les organisations en infraction avec le RGPD peuvent se voir infliger une amende allant jusqu'à 4% du chiffre d'affaires annuel mondial ou 20 millions d'euros.

— *Consentement.* Les conditions de demande consentement ont été clarifiées et les entreprises ne sont désormais plus en mesure d'utiliser des termes et conditions illisibles et remplis de jargon juridique. Ainsi, les conditions d'utilisations doivent être claires et fournies sous une forme intelligible et facilement accessible, dans un langage clair et simple.

— *Portée territoriale accrue.* Le RGPD s'applique à toutes les entreprises trai-

6. https ://eugdpr.org/the-regulation/

tant les données à caractère personnel des citoyens et résidents de l'Union Européenne et cela peut importe la localisation de l'entreprise.

Du point de vue des citoyens européens, le RGPD garantit les droits suivants.

- *Droit à l'accès.* Les personnes concernées ont le droit d'obtenir du responsable du traitement la confirmation que des données à caractère personnel les concernant font ou non l'objet d'un traitement, où elles sont stockées et pour quelles finalités.
- *Notifications de brèches de vie privée.* Les notifications d'atteinte à la protection des données sont désormais obligatoires dans tous les états membres de l'Union Européenne où une atteinte à la protection des données est susceptible d'entraîner un risque pour les droits et libertés des personnes.
- *Droit à l'effacement.* La personne concernée peut demander au responsable du traitement d'effacer ses données personnelles, de cesser la diffusion des données et, éventuellement, de faire cesser le traitement des données par des tiers.
- *Portabilité des données.* La personne concernée a le droit de recevoir les données à caractère personnel la concernant dans un format standard. De plus les citoyens ont le droit de transmettre ces données à un autre responsable du traitement.
- *Protection de la vie privée dès la conception.* Ce principe est le plus important dans notre contexte et pour notre proposition. En effet, bien que le concept de la protection de la vie privée dès la conception existe depuis des années, il devient une exigence légale dans le cadre du RGPD.

  Fondamentalement, la protection de la vie privée dès la conception exige l'inclusion de la protection des données dès le début de la conception des systèmes, plutôt que d'y arriver en ajout a posteriori.

  L'*Article 23* (GDP, 2016) prévoit que les responsables du traitement ne

> doivent conserver et traiter que les données absolument nécessaires à l'accomplissent de leur mission, ainsi que de limiter l'accès aux données à caractère personnel à ceux qui doivent procéder au traitement (principe de minimisation des données).

En ce qui concerne l'impact de la RGPD sur l'IdO, bons nombres des activités de traitement des données liées au fonctionnement de l'IdO relèveront du champ d'application matériel de la RGPD, étant donné que les objets connectés figurant au sein de l'IdO ont tendance à traiter des données personnelles. Par conséquent, la protection des données devrait être intégrée dans toutes solutions de l'IdO dès le début et tout au long du cycle de vie des objets (principe de *Privacy by Design).*

De plus, une évaluation de l'impact sur la protection des données pourrait, selon toute vraisemblance, devoir être effectuée. Les concepts de transparence, d'équité et de capacité à respecter les droits des personnes concernées devraient être intégrés dans la conception des objets. Tout cela devrait être documenté et mis en évidence dans le cadre du principe de responsabilité du RGPD.

Dans le contexte de l'IdO, l'une des difficultés principales consiste à déterminer si une partie prenante agit en tant que responsable du traitement des données ou en tant que responsable du traitement des données dans une activité particulière et précise. Par exemple, les fabricants d'objets sont considérés comme des contrôleurs pour les données personnelles générées par les objets, car ils conçoivent le système d'exploitation ou déterminent la fonctionnalité globale du logiciel installé. Les développeurs quant à eux sont des contrôleurs lorsqu'ils utilisent les objets connectés pour collecter et traiter des informations sur des individus.

Il peut y avoir également d'autres parties prenantes, telles que les plates-formes de données de l'IdO et les plates-formes sociales peuvent être considérées comme des contrôleurs pour les activités dues traitement, pour lesquelles elles déterminent

les finalités et les moyens. Au contraire, ils peuvent être considérés comme des sous-traitants lorsqu'ils traitent des données pour le compte d'une autre partie prenante de l'IdO qui agit en tant que responsables du traitement. Il est donc important que les parties prenantes liées à l'IdO procèdent à une évaluation des activités de traitement afin d'identifier les rôles respectifs en matière de protection des données et de répartir correctement les responsabilités.

Maintenant que nous avons passé en revue l'IdO ainsi que les différentes architectures existantes ainsi que la sécurité et la protection de la vie privée. Nous allons aborder dans les prochaines sections, les différents composants plus spécifiques lies à ma proposition. La première étant les SDI, l'AM avec plus particulièrement l'AF ainsi qu'une vue d'ensemble sur les différents protocoles d'ASD.

2.6 Briques de constructions

Cette dernière partie de la revue de la littérature introduit trois composants importants à notre proposition, à savoir, les SDI, l'AF ainsi que l'ASD.

2.6.1 Système de détection d'intrusion (SDI)

En sécurité informatique, aussi bien de manière générale que dans le contexte de l'IdO la gestion de la sécurité peut être divisée en trois étapes : Prévention, Détection et Atténuation. Ainsi après l'étape de prévention, la détection d'intrusion prend la relève, cette phase ayant pour but de détecter tout comportement suspect dans un réseau ou système informatique.

Les SDI ont été conçus pour gérer cette seconde phase de la sécurité informatique. Le nombre de fonctionnalités peut varier selon le système utilisé, mais en règle générale les SDI permettent d'identifier, localiser sur le réseau l'intrus ainsi que

de savoir des paramètres importants tels que le temps, l'activité, le type ou encore la couche où l'intrusion est effectuée.

La figure 2.10 représente la classification des différents types et fonctionnalités des SDI qui sera décrite plus en détail dans les prochaines sous-sections.

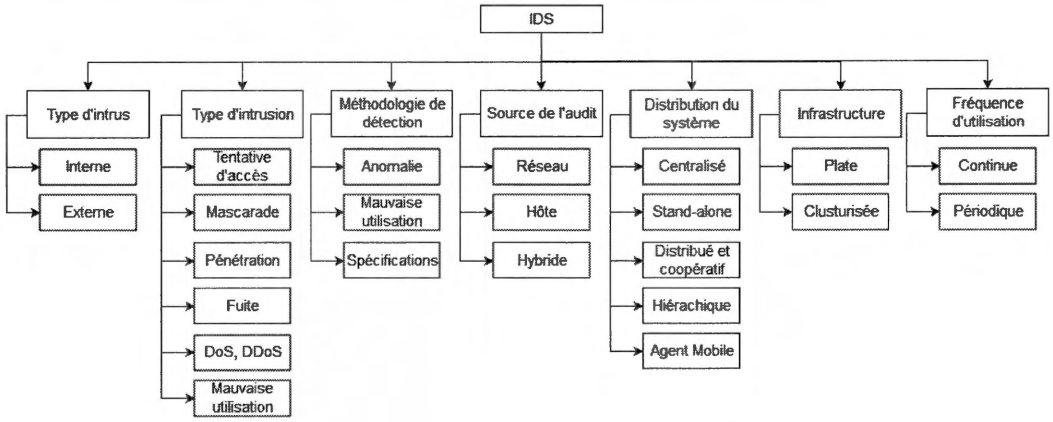

Figure 2.10 Classification des SDI inspirée de (Butun *et al.*, 2014)

Types d'intrus et d'intrusions

Il existe plusieurs types d'intrus : l'intrus externe qui est un attaquant utilisant différentes attaques pour atteindre un réseau ainsi que les systèmes informatiques situés à l'intérieur et l'intrus interne qui prend la forme d'un noeud compromis faisant partie intégrante du réseau ciblé. Il est possible de catégoriser les différents types d'intrusions de la manière suivante :

— Tentative d'effraction : une tentative d'accès non autorisé au réseau.
— Mascarade : un attaquant utilise une fausse identité pour obtenir un accès non autorisé au réseau
— Pénétration : acquisition d'un accès non autorisé au réseau.

— Fuite :un flux d'informations indésirable provenant au réseau
— Déni de Service (DOS, DDOS) : blocage des ressources du réseau (bande passante de communication) aux utilisateurs.
— Mauvaise utilisation : nuire délibérément aux ressources du réseau

Méthodes de détection et sources d'information

Il existe trois types de méthodes de détection d'intrusion (Butun *et al.*, 2014). La première est une modélisation du comportement (statistiques) qui se concentre sur les anomalies détectées qui correspondent à un comportement anormal. La seconde quant à elle se base sur les profils d'attaquants enregistrés précédemment en tant que références afin de détecter des comportements suspicieux. La troisième méthode se base sur un jeu de spécifications et de contraintes définissant le bon fonctionnement d'un programme ou d'un protocole, par la suite, l'exécution du programme ou du protocole s'effectue par rapport aux spécifications et contraintes définies est contrôlée.

Les SDI peuvent récolter de l'information de plusieurs manières. La première implique d'écouter, capturer et examiner les paquets circulant sur le réseau. La seconde méthode est basée sur l'hôte et est concernée uniquement par les événements sur celui-ci (pouvant être un ordinateur par exemple) tandis que la troisième méthode est un hybride des deux premières.

Distribution des systèmes, infrastructure et fréquence d'utilisation

Les systèmes de détection d'intrusion peuvent être localisés de plusieurs façons (Butun *et al.*, 2014) :

— *Centralisé.* Un hôte centralisé surveille toutes les activités du réseau et

détecte les intrusions en analysant les données d'activité de celui-ci.

- *Autonome.* Un SDI est exécuté indépendamment sur chaque noeud et chaque décision est basée uniquement sur les informations collectées sur le noeud.
- *Distribué et coopératif.* Chaque noeud exécute un agent SDI qui participe (en coopérant aux décisions et actions de détection d'intrusion globale) à la détection d'intrusion et à la réponse de l'ensemble du réseau. Si un noeud détecte une intrusion, il peut initier une procédure de détection d'intrusion globale s'il manque de preuves. Dans le cas où il possède assez de preuves, il peut alors prendre la décision d'alerter indépendamment le réseau qu'une attaque est en cours.
- *Hiérarchique.* Ce modèle a été proposé pour les infrastructures réseau multi-couches (*Clustering* en anglais). Les têtes de clusters (*Cluster Heads* en anglais) sont responsables de la surveillance de leurs noeuds membres ainsi que de la participation aux décisions globales de détection d'intrusion.
- *Agent mobile.* Chaque agent est affecté à une tâche spécifique du SDI sur un noeud sélectionné et la détection est réalisée par l'action coopérative des noeuds sélectionnés.

Les SDI peuvent suivre deux formes d'infrastructure, la première plate (*Flat* en anglais), dans laquelle les noeuds sont considérés égaux entre eux, tandis que dans la seconde dite *clusterisée* les noeuds sont inégaux. En ce qui concerne la fréquence d'utilisation des SDI, ils peuvent être utilisés soit de manière continue ou périodiquement.

La prochaine section est dédiée à l'AM et plus particulièrement à l'AF.

2.6.2 Apprentissage machine

L'AM (Mohri *et al.*, 2012) est un sous-domaine de l'intelligence artificielle (IA) s'intéressant au développement d'algorithmes qui apprennent à partir de données permettant ainsi de construire des modèles à partir de ces données qui pourront être utilisées plus tard pour faire des prédictions sur de nouvelles données.

Bien que l'AM soit un domaine de l'informatique, il diffère des approches classiques en informatique. En effet, dans l'informatique « traditionnel », les algorithmes sont des ensembles d'instructions explicitement programmés utilisés par les ordinateurs pour calculer ou résoudre des problèmes. Les algorithmes d'AM quant à eux permettent aux ordinateurs de s'entraîner grâce à des jeux de données et d'utiliser l'analyse statistique afin de prédire les sorties attendues. Ainsi, l'AM facilite la construction de modèles à partir d'ensembles de données afin d'automatiser les processus décisionnels.

Il existe deux formes principales d'AM (illustré dans la figure 2.11).

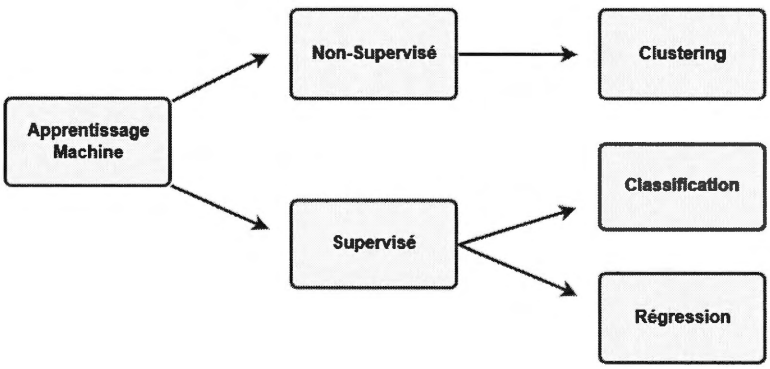

Figure 2.11 Formes d'apprentissage machine

1. Apprentissage supervisé (*Supervised Learning* en anglais) (Kotsiantis *et al.*,

2007) Le but de cette forme d'apprentissage est de construire un modèle qui fait des prédictions en se basant sur des données d'entraînements. Plus précisément, elle s'entraîne sur un ensemble connu de données d'entrée et des réponses associées et forme un modèle pour générer des prédictions pertinentes pour de nouvelles données.

L'apprentissage supervisé utilise la classification et la régression pour développer des modèles prédictifs :

La *classification* cherche à prédire une classe parmi un ensemble fini (par exemple : si un courriel est authentique ou s'il s'agit d'un pourriel). Les modèles de classification classent les données d'entrée en catégories.

La *régression* prédit des réponses parmi un ensemble continu (par exemple : les changements de températures ou les fluctuations de la demande d'électricité).

2. Apprentissage non-supervisé (*Unsupervised Learning* en anglais) (Gentleman et Carey, 2008). Cette forme d'apprentissage cherche à révéler des structures intrinsèques cachées dans les données. Elle est principalement utilisée pour mener des analyses à partir d'ensemble de données composé de données d'entrée sans réponses associées. La technique de *clustering* (Hastie *et al.*, 2009) est la plus communément utilisée pour l'analyse exploratoire des données afin de trouver des motifs ou des regroupements cachés dans les données. Elle consiste à diviser la population ou les points des données en un certain nombre de groupes de sorte que les points de données des mêmes groupes sont plus semblables aux autres points de données du même groupe et différents des points de données des autres groupes. Il s'agit essentiellement d'une collection d'objets sur la base de la similitude et de la dissimilitude entre eux.

2.6.3 Apprentissage fédéré (*Federated Learning* en anglais)

Les approches standards d'AM nécessitent de centraliser les données sur une seule machine ou dans un centre de données (*DataCenter* en anglais). Mais une nouvelle approche, l'AF (Fed, 2017), a fait son apparition en 2016 et est mise en avant par Google comme permettant d'avoir un apprentissage collaboratif sans centraliser les données d'entraînement.

Il y a plusieurs raisons qui ont poussé à la création de cette nouvelle approche. En effet l'AM centralisé est l'architecture la plus commune, mais la séparation entre les algorithmes d'apprentissage et les données utilisateurs a pour effet que de colossaux volumes de données sont toujours en mouvement ce qui représente de sérieuses contraintes, dont en particulier :

- *Coût de transfert.* Dû à la croissance phénoménale du nombre de données qui ont besoin d'être gérées, le coût de transfert de ces données est élevé.
- *Latence.* L'AM n'est pas la solution la plus appropriée dans beaucoup de situations, car cette approche nécessite d'interagir en temps réel.
- *Confidentialité.* Le fait de devoir transférer des données personnelles et de les avoir stockées dans des serveurs distants peut créer des opportunités à des attaquants pour intercepter ces données.
- *Incompatibilité.* Pour des raisons de confidentialité, certains domaines (santé, assurances, banques ou militaires) ne sont pas autorisés à partager leurs données et de les stocker dans l'infonuagique.

L'AF est une approche spécifique à l'AM ayant pour objectif de former un modèle centralisé de haute qualité avec des données d'entraînement distribuées sur un grand nombre de clients ayant chacune des connexions réseau peu fiables et relativement lentes.

Cette approche a été originellement développée pour les téléphones intelligents en leur permettant d'apprendre de façon collaborative à partir d'un modèle de prédiction partagé tout en conservant toutes les données d'entraînements sur l'appareil, séparant la capacité de faire de l'AM et la nécessité de stocker les données dans l'infonuagique.

Depuis, d'autres articles ont continué à étendre l'utilisation de l'AF, par exemple :

- *Semi-Supervised Knowledge Transfer for Deep Learning from Private Training Data* (Papernot *et al.*, 2016). Dans cet article, les auteurs proposent *PATE*, une approche ayant pour but de protéger la confidentialité des données pendant l'entraînement en transférant les connaissances d'un ensemble de modèles d'enseignants (*Teachers* en anglais) sur des partitions des données à un modèle étudiant. Le modèle étudiant apprend à prédire un résultat choisi par un vote de majorité des modèles enseignants et ne peut pas accéder directement à un enseignant en particulier ou aux données ou paramètres sous-jacents.
- *Federated learning : Strategies for Improving Communication Efficiency* (Konečnỳ *et al.*, 2016). Dans cette étude, les auteurs proposent en utilisant l'AF, deux façons de réduire le coût de communication sur les liaisons montantes (*UpLink* en anglais) dans un cadre d'utilisation de téléphones intelligents où l'efficacité des communications est un enjeu crucial. La première manière est d'effectuer des mises à jour structurées consistant à apprendre les nouvelles mises à jour directement d'un espace restreint paramétré en utilisant un petit nombre de variables. La seconde façon est d'effectuer des mises à jour esquissées (*Sketched Updates) en anglais*, où un apprentissage d'une mise à jour complète du modèle est compressé en utilisant une combinaison de quantification, de rotation aléatoire et de sous-échantillonnage.
- *Practical Secure Aggregation for Federated Learning on User-Held Data*

(Bonawitz *et al.*, 2016). Dans cet article, les auteurs entraînent un réseau de neurones profond avec l'approche de l'AF en utilisant la descente de gradients distribuée à travers des données d'entraînement détenues par les utilisateurs sur leurs téléphones intelligents. Ils utilisent également un nouveau protocole d'agrégation sécurisée efficace pour les données de haute dimensionnalité qui tolère jusqu'à un tiers des utilisateurs ne complétant pas le protocole.

— *Communication-Efficient Learning of Deep Networks From Decentralized Data* (McMahan *et al.*,). Dans cette recherche, les auteurs présentent une méthode efficace pour l'AF des réseaux profonds basée sur le calcul itératif de la moyenne des modèles et conduisent une évaluation empirique approfondie en considérant cinq architectures de modèles et quatre ensembles de données différents.

— *Differentially Private Federated Learning : A Client Level Perspective* (Geyer *et al.*, 2017). Dans cet article, les auteurs proposent un algorithme de confidentialité différentielle (Dwork, 2011) pour le client en préservant les bénéfices de l'utilisation de l'AF. Cet algorithme a pour but de résoudre le problème des attaques différentielles (*Differential attacks)*, elles reposent sur la recherche d'une corrélation entre une certaine différence entre deux textes clairs en entrée et une différence en sortie.

— *Scalable Private Learning with PATE* (Papernot *et al.*, 2018). Dans cet article, les auteurs ont démontré que l'approche *PATE* (Papernot *et al.*, 2016) peut s'adapter à des tâches d'apprentissage avec de grands nombres de classes de sorties non traitées et déséquilibrées de données d'entraînements comportant des erreurs. En introduisant de nouveaux mécanismes d'agrégation pour les ensembles d'enseignants qui sont plus sélectifs et moins bruités et prouvent leurs garanties plus rigoureuses en matière de vie privée différentielle.

Fonctionnement et défis liés à l'apprentissage fédéré

Comme spécifié précédemment, les travaux et l'utilisation de l'AF ont été principalement proposés dans un contexte utilisant les téléphones intelligents. Le cas illustré par la figure 2.12 représente l'utilisation de l'AF sur la prédiction de texte. L'AF sera détaillé dans la section 2.6.3 pour son application dans l'IdO et dans le contexte de la domotique.

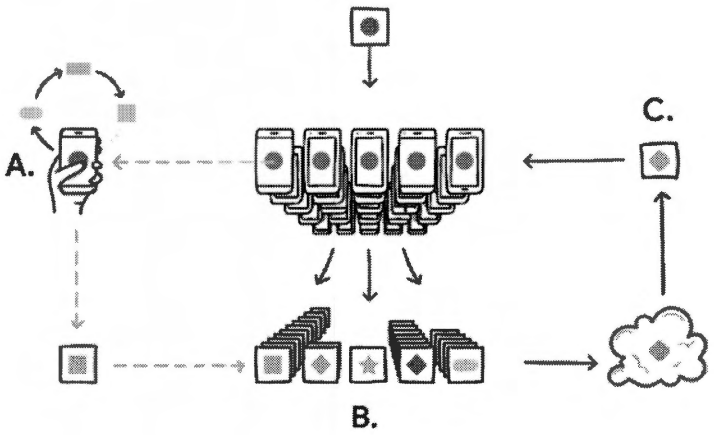

Figure 2.12 Cas d'utilisation de l'apprentissage fédéré (Fed, 2017)

Dans le cas illustré à la figure 2.12, les téléphones mobiles téléchargent le modèle actuellement disponible, qui est amélioré à travers le temps grâce à l'apprentissage effectué directement sur les données de chaque téléphone (représenté par A), puis regroupe les changements apportés grâce à chaque apprentissage sous forme de mise à jour (illustré par B). Cette mise à jour est envoyée à l'infonuagique, en utilisant des communications chiffrées, où toutes les mises à jour sont regroupées afin d'améliorer le modèle partagé commun (représenté par C). Le cycle continu encore et encore afin de continuer à améliorer les modèles de prédictions de textes

sur téléphones.

De plus, dans cet exemple (Fed, 2017), un protocole d'ASD à été développé utilisant des mécanismes cryptographiques permettant qu'un serveur de coordination ne puisse déchiffrer la mise à jour que si des centaines ou des milliers d'utilisateurs y ont participé, ce qui implique qu'aucune mise à jour individuelle de téléphone ne peut être inspectée avant le calcul de la moyenne.

Afin de rendre l'AF possible, il faut répondre à plusieurs défis techniques mais aussi algorithmiques. Dans un système d'apprentissage classique, un algorithme d'optimisation fonctionne sur un grand ensemble de données réparti de manière homogène sur des serveurs dans l'infonuagique. Dans le cas d'une approche d'AF , les données sont réparties sur des millions d'objets et de manière non-homogène. Ces objets, ont des connexions à latence et faible débit et ne sont disponibles que de façon intermittente pour la formation des modèles, la latence est nettement plus élevée que pour des ordinateurs fixes ou portables ayant dans la majorité des cas accès à des réseaux plus haut débit et plus stables.

Le dernier concept à aborder dans ce chapitre est l'ASD.

2.6.4 Agrégation sécurisée de données

L'agrégation des données (Hu et Evans, 2003) au sein des réseaux de capteurs sans-fil est un processus d'exploration des données et d'informations où les données sont recherchées, recueillies et présentées dans un format résumé et fondé sur des rapports afin d'atteindre des objectifs ou des processus opérationnels précis ou encore effectuer une analyse humaine. Il s'agit de l'étape qui se produit entre le regroupement de données et la phase d'analyse. Les données peuvent provenir d'une ou plusieurs sources différentes et l'agrégation des données peut être implémentée de plusieurs manières. Pour un sujet tel que l'IdO, c'est un processus

important et nous allons aborder sa sécurité dans la suite.

Besoins fondamentaux à prendre en compte pour l'agrégation sécurisée au sein de réseaux sans-fils

1. *Confidentialité des données.* La confidentialité des données garantit que les données n'ont jamais été divulguées à des parties non autorisées et c'est l'un des aspects les plus importants en termes de sécurité dans les réseaux sans-fils.
2. *Intégrité des données.* Bien que la confidentialité des données garantisse que seules les parties visées obtiennent les données non chiffrées, elle ne protège pas les données contre des possibilités de modifications de celles-ci. L'intégrité quant à elle garantit qu'un message transféré n'ait jamais été corrompu.
3. *Authentification des sources.* L'authentification de la source permet à un noeud de s'assurer de l'identité du noeud avec lequel il communique. Sans authentification de la source, un adversaire pourrait tenter d'obtenir un accès non autorisé à des ressources et des informations sensibles et d'interférer avec le fonctionnement d'autres noeuds.
4. *Fraîcheur des données.* La fraîcheur des données est le fait de s'assurer que les données sont récentes et qu'aucun message ancien n'a été rediffusé pour protéger les schémas d'agrégations de données contre les attaques de type rejeu.

État de l'art des protocoles d'agrégation

Le premier protocole d'ASD a fait son apparition en 2003 (Hu et Evans, 2003). Il est focalisé sur le problème d'agrégation de données dans un contexte où un noeud peut être compromis. Depuis la création de ce premier protocole, de nombreux

autres ont fait leur apparition :

— **SIA** (*Secure Information Aggregation*) (Przydatek *et al.*, 2003). Il s'agit d'un cadre utilisant une approche de type agréger-mettre en gage-prouver (*Aggregate-Commit-Prove* en anglais), en recueillant tout d'abord les données (provenant des objets) avant de les agréger et de valider les données collectées à l'aide d'un arbre de hachage de *Merkle* (Merkle, 1980). Un arbre de hachage de *Merkle* est une structure de données ayant pour but de vérifier l'intégrité d'un ensemble de données. Par la suite, les résultats sont communiqués à la station de base une fois que l'exactitude et l'intégrité des données ont été vérifiées.

— **ESPDA** (*Energy-Efficient and Secure Pattern-based Data Aggregation Protocol*) (Çam *et al.*, 2006). Il s'agit du premier protocole à prendre en compte les techniques d'agrégation de données sans compromettre la sécurité en utilisant des modèles de codes pour effectuer l'agrégation. Les modèles de codes sont des éléments représentatifs qui sont extraits des données réelles de telle sorte que chaque modèle de code possède certaines caractéristiques des données réelles correspondantes.

— **SecureDAV** (*A Secure Data Aggregation and Verification Protocol*) (Mahimkar et Rappaport, 2004). Il s'agit d'un protocole d'agrégation basé sur le principe de *Cluster* en utilisant des constructions d'arbre de hachage de Merkle (Merkle, 1980) (*Merkle Hash Trees*).

Les auteurs proposent d'utiliser la cryptographie sur les courbes elliptiques (*Elliptic Curve Cryptography* - *ECC* en anglais) car elle permet une bonne sécurité avec des clés de petite taille, améliorant ainsi le temps de calcul.

— **SRDA** (*Secure Reference-Based Data Aggregation Protocol*) (Sanli *et al.*, 2004). Ce protocole a été proposé pour les réseaux de capteurs sans-fil en cluster. Dans ce cas, les données brutes détectées par les noeuds des

capteurs sont comparées aux données de référence et seules les différences sont transmises. Les données de référence sont considérées comme la valeur moyenne d'un certain nombre de mesures précédentes du capteur.

SRDA a pour but de réduire le nombre de bits d'une transmission, car la communication est l'activité la plus consommatrice d'énergie dans réseaux de capteurs. Ainsi tandis que l'agrégation de données réduit le nombre de paquets, la réduction de la taille des paquets transmis permet d'améliorer d'autant plus les économies d'énergie.

— **CDA** (*Concealed Data Aggregation*) (Girao *et al.*,). Dans ce protocole, les noeuds partagent une clé symétrique commune avec la station de base qui est cachée des agrégateurs intermédiaires. La principale contribution de ce travail est la mise en place d'un chiffrement de bout-en-bout pour le trafic à multi-diffusion inverse entre les capteurs et la station de base. Avec cette approche, les agrégateurs de données exécutent des fonctions d'agrégations qui sont appliquées aux textes chiffrés (données), ce qui a pour avantage que les agrégateurs intermédiaires n'ont pas à effectuer des opérations coûteuses de déchiffrement et de chiffrement.

— **SDAP** (*Secure Hop-byHop Data Aggregation Protocol*) (Yang *et al.*, 2008). Il s'agit d'un protocole basé sur l'approche mettre en gage et attester (*Commit-and-Attest* en anglais) ce qui implique que l'agrégateur mettant en gage un résultat ne peut le dénier par la suite, puis diviser-pour-régner qui est une stratégie qui divise l'arbre du réseau en sous-arbres ce qui permet de rendre l'authentification et la détection de noeuds compromis plus facile. SDAP se compose de trois étapes : la construction de l'arbre d'agrégation, la récolte d'information et la vérification et l'attestation des résultats.

— **SELDA** (*Secure and Reliable Data Aggregation*) (Ozdemir, 2007). L'idée générale derrière ce protocole est que les noeuds observent les actions de

leurs noeuds voisins pour développer des estimations de confiance, à la fois pour l'environnement et pour les noeuds voisins. Ainsi, les noeuds emploient des mécanismes de surveillance pour détecter la disponibilité, la détection et le routage des noeuds, les mauvais comportements de leurs voisins, qui sont quantifiés comme des niveaux de confiance.

Les noeuds échangent leurs niveaux de confiance avec les noeuds voisins pour former un réseau de confiance qui leur permet de déterminer des chemins sûrs et fiables vers les agrégateurs de données qui mesurent eux-mêmes les données provenant des capteurs qu'ils reçoivent en s'appuyant le réseau de confiance.

— **SEDAN** (*Secure and Efficient protocol for Data Aggregation*) (Bagaa *et al.*, 2007). Le protocole proposé par les auteurs est une agrégation sécurisée des données pour les réseaux de capteurs sans-fils se basant sur un mécanisme de vérification à deux sauts de l'intégrité des données. Cette solution se différencie des autres sur le principe qu'elle ne nécessite pas de se référer à la station de base pour vérifier et détecter les données agrégées erronées.

— **RSDA** (*Reputation-based Secure Data Aggregation*) (Alzaid *et al.*, 2008). Ce protocole se base aussi sur la notion de réputation pour les réseaux sans-fils qui intègre les fonctionnalités d'agrégations avec les avantages fournis par un système de réputation pour améliorer la durée de vie du réseau et l'exactitude des données agrégées. Ce protocole associe des clés secrètes symétriques à des emplacements géographiques et elles sont assignées à des noeuds en fonction de leur emplacement.

— **SEEDA** (*Secure End-to-End Data aggregation*) (Poornima et Amberker, 2010). Ce protocole est un schéma hybride qui combine les avantages d'un schéma d'agrégation bout-en-bout et d'un schéma d'agrégation par bonds. Il assure ainsi la confidentialité des données de bout-en-bout et le nombre de bits transmis est pratiquement le même que celui d'un système d'agrégation

saut-à-saut.

— **EEHA** (*Energy-Efficient and High Accuracy Secure Data Aggregation*) (Li *et al.*, 2011). Ce protocole permet d'obtenir une agrégation précise des données sans libérer les détections provenant des capteurs privés et sans introduire une surcharge significative sur les capteurs à faible batterie. L'objectif principal de ce protocole est la défense contre les attaques d'écoutes dans lesquelles un attaquant tente d'entendre la transmission sur des liaisons sans-fil pour obtenir des informations privées.

— **IPHCDA** (*Integrity Protecting Hierarchical Concealed Data Aggregation*) (Ozdemir et Xiao, 2011). Il s'agit d'un protocole d'agrégation de données hiérarchiques qui se base sur des codes d'authentification de messages (CAM) s'appuyant sur une clé secrète partagée ainsi que des schémas de chiffrement homomorphe. L'idée générale du chiffrement homomorphe est qu'il permet à un tiers qui possède une clé publique d'effectuer des calculs arbitraires (addition ou multiplication) sur des messages préalablement chiffrés. Ce qui permet d'obtenir en résultats de ces calculs de nouveaux messages, qui sont les chiffrés des résultats des opérations.

— **RDCA** (*Recoverable Concealed Data Aggregation for Data Integrity*) (Chen *et al.*, 2012). Ce protocole est basé sur le principe de *cluster* et possède la particularité de récupérer toutes les données générées par des noeuds même après le processus d'agrégation. Les auteurs proposent deux schémas différents, *RCDA-HOMO* pour les réseaux homogènes et *RCDA-HETE* pour les réseaux hétérogènes, malheureusement, RCDA-HETE ne garantit pas l'intégrité des données.

— **CRSR** (*Secure Data Aggregation Algorithm*) (Lathamanju et Senthilkumar, 2013). Les auteurs ont proposé et algorithme d'agrégation sécurisé de données pour améliorer la durée de vie du réseau et d'assurer sa sécurité. Cet algorithme utilise une technique d'agrégation de données basée sur

les clusters en utilisant LEACH-KED (qui est un algorithme de hiérarchie adaptative des clusters à basse énergie).

— **PEPPDA** (*Power Efficient Privacy Preserving Data Aggregation*) (Jose *et al.*, 2013). Ce protocole convient bien aux applications critiques et sécuritaires comme les applications militaires par exemple. L'objectif principal de *PEPPDA* est de fournir un système d'agrégation de données sécurisé qui garantit la confidentialité, l'authenticité et la fraîcheur des données individuelles ainsi que l'exactitude et la confidentialité des données agrégées sans introduire une surcharge importante sur les capteurs à faible batterie. L'aspect de protection de la vie privée est assuré par l'utilisation du tranchage et de l'opération d'assemblage au niveau des noeuds. La confidentialité des données est assurée par l'agrégation de données chiffrées de bout en bout. L'authentification des messages s'effectue à l'aide de la clé secrète et de la paire ID de chaque noeud. La fraîcheur des données est obtenue en utilisa la clé de chiffrement variable pour chaque session.

— **EESSDA** (*An Energy-Efficient and Scalable Secure Data Aggregation*) (Wang *et al.*, 2013). Les auteurs ont proposé un protocole économe en énergie , sécurisé nommé EESSDA. Dans EESSDA, l'agrégation sécurisée des données est réalisée par la mise en place d'un canal sécurisé et d'une technologie de découpage en tranches (*Slicing* en anglais) qui consiste à diviser les données détectées en morceaux. De plus EESSDA n'a pas besoin d'opérations de chiffrement et de déchiffrement pendant l'agrégation des données.

— **ECIPAP** (*An Efficient Confidentiality and Integrity Preserving Aggregation Protocol*) (Zhu *et al.*, 2014). Il s'agit d'un protocole efficace d'agrégation préservant l'intégrité et la confidentialité des données. Les auteurs ont mis un effort sur le coût de la communication de l'en-tête dans la phase de vérification des résultats. Ceci qui lui permet, en comparaison à une grande

majorité d'autres protocoles d'être plus efficace et d'avoir un moindre coût sur les ressources à disposition.

Protocoles d'agrégation sécurisée de données				
Protocoles	**Confidentialité**	**Intégrité**	**Authentification**	**Fraîcheur**
SDA (2003)	**Non**	**Oui**	**Oui**	**Oui**
SIA (2003)	**Oui**	**Oui**	**Oui**	**Oui**
ESPDA (2003)	**Oui**	**Non**	**Oui**	**Oui**
SecureDAV (2004)	**Oui**	**Oui**	**Oui**	**Non**
SRDA (2004)	**Oui**	**Non**	**Oui**	**Oui**
CDA (2005)	**Oui**	**Non**	**Non**	**Non**
SDAP (2006)	**Oui**	**Oui**	**Oui**	**Oui**
SELDA (2007)	**Non**	**Oui**	**Oui**	**Oui**
SEDAN (2007)	**Non**	**Oui**	**Oui**	**Oui**
RSDA (2008)	**Non**	**Oui**	**Oui**	**Oui**
SEEDA (2010)	**Oui**	**Non**	**Oui**	**Oui**
EEHA (2011)	**Oui**	**Non**	**Oui**	**Oui**
IPHCDA (2011)	**Oui**	**Oui**	**Oui**	**Oui**
RCDA (2012)	**Oui**	**Oui**	**Oui**	**Oui**
CRSR (2013)	**Oui**	**Non**	**Oui**	**Oui**
PEPPDA (2013)	**Oui**	**Non**	**Oui**	**Oui**
EESSDA (2013)	**Oui**	**Non**	**Non**	**Non**
ECIPAP (2014)	**Oui**	**Oui**	**Oui**	**Oui**

Tableau 2.1 Comparatifs des protocoles d'agrégation sécurisée de données en fonction des besoins fondamentaux de sécurité

En prenant en compte les fondamentaux et caractéristiques des différents protocoles existants en matière d'agrégation sécurisée de données, nous pouvons créer

un tableau comparatif inspiré de (Soni et Randhawa, 2017).

Tout ce qui a été présenté jusqu'à présent consiste de l'introduction et contexte de mon sujet de maîtrise ainsi qu'une étude de la littérature des différents concepts importants. Désormais, la seconde partie sera dédiée à notre proposition.

CHAPITRE III

CONCEPTION D'UNE NOUVELLE ARCHITECTURE - IOTFLA

Ce chapitre et le chapitre suivant ont pour but de décrire les composants, le fonctionnement ainsi que leurs interactions au sein de notre proposition. Le chapitre 3 couvre tout d'abord une mise en contexte de la proposition, puis nous détaillerons la composition de celle-ci avant de poursuivre sur le fonctionnement ainsi que l'impact sur la sécurité et la protection de la vie privée. Dans le chapitre suivant (4), nous nous attarderons sur l'intégration des différents composants présentés dans ce chapitre et nous proposerons différents scénarios d'application de notre architecture. Puis, nous passons en revue le modèle d'adversaire agissant sur notre architecture ainsi qu'une analyse de la sécurité et de la protection de la vie privée. Nous clôturerons la deuxième partie de ce mémoire sur une section décrivant comment nous aurions pu implémenter et simuler notre proposition, avant de conclure.

3.1 Contexte de la proposition

Notre proposition, l'architecture IOTFLA que nous allons présenter dans ce chapitre et le suivant est une architecture théorique et n'a pas été implémentée dans un environnement de test pendant la durée de cette maîtrise. Nous avons conscience qu'elle constitue un niveau de complexité plutôt élevé et que son im-

plémentation soulève de nombreux défis même si les composants requis sont déjà utilisés couramment pour différentes applications.

En prenant en compte toutes les notions, concepts et technologies présentés dans les chapitres précédents, nous pouvons commencer à établir les fondations de la proposition de ce mémoire, qui prend la forme d'une architecture pour les maisons intelligentes. En effet, comme décrit brièvement dans le chapitre 1, nous avons suivi plusieurs pistes au cours des recherches sur la littérature scientifique existante. Nous avons par la suite décidé de regrouper ces différentes idées en une seule et unique proposition d'architecture regroupant les composants, caractéristiques, technologies et spécificités des différentes idées que nous voulions intégrer.

Comme présenté dans la section 2.4, il existe déjà beaucoup de propositions d'architectures pour l'IdO dans la littérature scientifique, certaines étant focalisées sur l'aspect de la sécurité et de la protection des données personnelles. En revanche, il n'existe pas à notre connaissance dans la littérature, une architecture combinant les mêmes concepts et technologies (AF, SDI et de système de prévention d'intrusion (SPI), protocole d'ASD, etc.) que celle proposée dans ce mémoire.

L'idée d'avoir cette combinaison unique des technologies citées précédemment est que l'utilisation de celles-ci représente des propriétés intéressantes pour l'IdO, la sécurité et la protection des données privées. L'AF ainsi que le protocole d'agrégation sécurisée des données permettent principalement de garder les données localement à la maison ainsi que de permettre des mises à jour des modèles d'objets d'une manière sécurisée et protectrice des données privées. De plus, l'analyse et la détection de comportement suspect au sein du réseau de la maison sont effectuées par le SDI.

Comme nous avons pu le présenter dans la première partie l'un des points importants de l'IdO de manière générale, mais d'autant plus au sein de la domotique

est la sécurité des données des utilisateurs. Avec notre proposition, nous essayons d'apporter des mesures pour la sécurité et la protection des données personnelles des utilisateurs tout en ajoutant un aspect innovant avec l'utilisation de l'approche de l'AF . À noter que ce chapitre couvre uniquement les fondations permettant de créer une telle architecture, l'intégralité de l'architecture étant décrite et développée dans le prochain chapitre.

Cette architecture cherche à atteindre différents objectifs basés sur deux axes principaux : la sécurité et la protection des données personnelles (respect de la vie privée). En effet, en prenant en compte ces axes et les concepts précédemment cités, nous avons pu rendre possible la définition et création d'une nouvelle architecture. Comme nous avons pu en discuter précédemment, la sécurité et la protection des données personnelles sont deux concepts importants et étroitement liés dans le contexte de l'IdO.

IOTFLA est tout d'abord, axée sur deux aspects, à savoir la sécurité et la protection de la vie privée dans un contexte de réseau privé de maison intelligente regroupant des objets connectés. Puis dans un second temps, elle se concentre sur un aspect plus innovant qui intervient une fois que certains composants et certaines technologies supplémentaires sont implémentés en prenant en compte les différentes stratégies initiales de sécurité et protection des données.

3.2 Composition de l'architecture IOTFLA

Cette architecture est basée sur l'architecture initiale à trois couches (perception, réseau et application) de l'IdO qui a déjà été évoquée dans le chapitre de l'état de l'art (section 2.3). Le schéma 3.1 représente le scénario tout à fait classique d'une maison intelligente avec des composants supplémentaires intelligents intégrés.

En nous basant sur l'architecture originale de l'IdO, nous pouvons décrire que

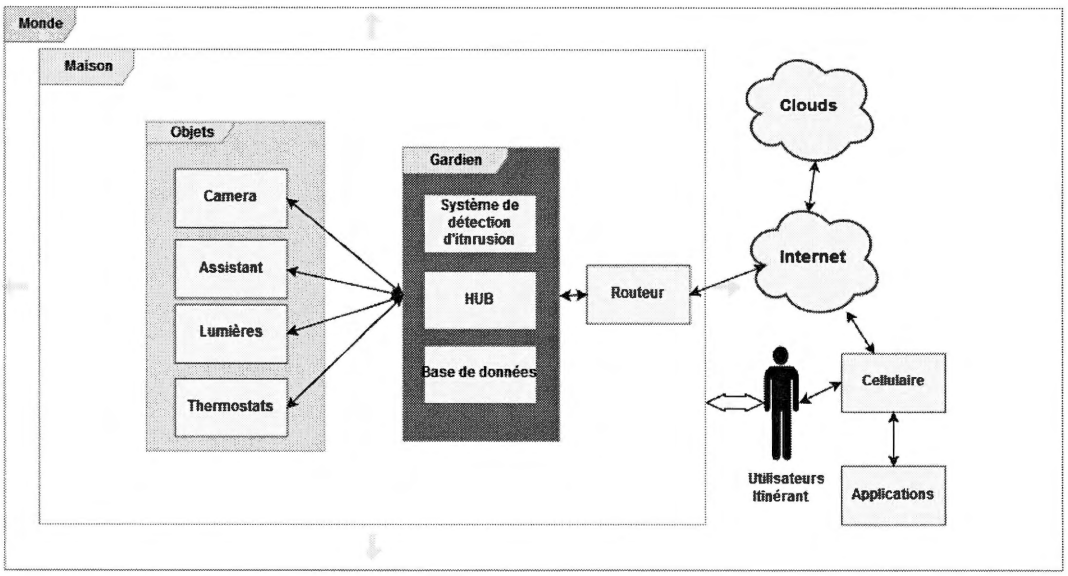

Figure 3.1 Schéma de l'architecture - Fondations

notre architecture impacte directement les trois différentes couches (perception, réseau et application), mais intervient principalement au niveau de la couche réseau (contrôle et transmission des données).

Nous retrouvons les éléments classiques et communs à toutes maisons intelligentes, à savoir les objets connectés, les utilisateurs itinérants [1] qui sont équipés de téléphones intelligents regroupant plusieurs applications liées aux objets connectés et un routeur permettant de faire la liaison entre le réseau extérieur et le réseau interne à la maison.

En plus de ses composants, l'architecture comprend un point central que nous avons décidé d'appeler *« gardien »*. Il peut être installé sur un ordinateur clas-

1. Utilisateurs et propriétaires de la maison intelligente, nous les avons nommés itinérants, car ils peuvent tant bien être à l'intérieur de la maison qu'à l'extérieur pour interagir avec les objets connectés et leurs applications.

sique ou bien sûr une plate-forme telle que le *Raspberry Pi 3* (en fonction des préférences et besoins de l'utilisateur). Le gardien possède trois sous-composants, le premier est un système de détection d'intrusion et de prévention, le second est un concentrateur (*HUB*) et le dernier est une base de données.

Il existe beaucoup de solutions de SDI (Snort (Roesch *et al.*, 1999), Suricata (Sur, 2018), OSSEC (OSS, 2019), Security Onion (Sec, 2019), etc.) qui sont utilisables pour l'IdO. De manière similaire, des solutions de HUB (concentrateurs) existent aussi bien en tant que versions constructeurs (Samsung Smart Things, WINK HUB, Amazon Echo, Control Amy, etc.) ou dans des solutions ouvertes (Home Assistant, Open HAB, Calaos, Domoticz, Mister House, Open Motics, etc.). Dans notre cas nous nous sommes intéressés plus particulièrement aux solutions de concentrateurs libres car ils permettent plus de flexibilité pour la mise en place et le développement de solution. À l'inverse, un concentrateur vendu directement par un constructeur est généralement restrictif dans ses fonctionnalités et dans sa flexibilité.

Les concentrateurs libres permettent d'automatiser et de gérer l'interaction entre les objets connectés ainsi que d'avoir une plus grande panoplie d'objets connectés, mais également de pouvoir implémenter différentes technologies avec ceux-ci, ce qui n'est pas forcément réalisable avec les concentrateurs de constructeurs. Il est donc plus souhaitable de se diriger vers une solution de concentrateur libre quand le but est d'avoir un maximum de flexibilité avec ce que l'on peut faire.

Pour le concentrateur, nous avons décidé de nous tourner vers la solution libre Home Assistant (Hom, 2019), qui est une solution d'automatisation de domotique proposant un grand choix d'automatisation d'objets connectés. Home Assistant possède également une communauté vaste et active de développeurs. Ainsi, grâce à cette communauté à l'heure actuelle de la rédaction de ce mémoire, Home As-

sistant ne compte pas moins de 1468 différentes applications, objets connectés et protocoles intégrés. De plus, cette solution permet que toute donnée produite par les objets connectés reste en local au sein du concentrateur (ce qui est point important pour la sécurité des données personnelles des utilisateurs).

Dû à tout ce qui est inclus dans la solution Home Assistant, à savoir une mise à jour régulière, une grande communauté active de développeurs, la capacité à faire de gérer ses données et d'analyser les données ainsi que bien évidemment de pouvoir automatiser et surveiller l'état des objets connectés au sein d'une maison, rend la solution libre vraiment attractive pour des personnes souhaitant automatiser et avoir un contrôle sur un réseau de domotique. En ce qui concerne le monde extérieur, Internet, les infonuages et les utilisateurs itinérants sont les seuls composants pris en compte dans notre cas. La grande majorité des objets communiquent vers l'extérieur, par exemple avec les nuages des constructeurs pour récupérer des mises à jour et également pour la mise en place de l'AF, qui sont nécessaires au fonctionnement de certains objets connectés. Les itinérants peuvent être à la fois sur le réseau interne, mais aussi utilisé à distance des applications liées aux objets connectés.

3.3 Fonctionnement

Dans le cas d'un scénario classique d'une maison intelligente, les objets connectés captent des données sur l'environnement et interviennent en fonction de leurs rôles au sein de l'écosystème. Cette situation ne change pas dans notre cas, les premières différences entre un scénario classique et celui que nous proposons apparaissent dans la manière dont les données circulent et sont gérées ainsi que sur le fonctionnement et l'interaction des objets connectés.

En effet, dans le cas le plus simple, un objet connecté (comme un capteur) ayant

récolté des informations sur l'environnement va chercher à les transmettre, soit vers une application locale au réseau ou sur une application se trouvant sur le téléphone intelligent de l'utilisateur itinérant ou encore vers une plate-forme distante. Pour cela, l'objet connecté doit transmettre ces données vers Internet via le routeur de la maison, elles seront par la suite retransmises aux différentes plates-formes concernées.

Bien évidemment, tous les objets connectés ne fonctionnent pas de la même manière de par leurs différences en termes de fonctionnalités, capacités, applications et interagissent en conséquence différemment au sein d'un réseau de domotique. Ainsi, les objets connectés font partie d'un grand spectre, allant de la simplicité avec les capteurs qui auront comme décrit ci-dessus des fonctionnalités et capacités réduites à des concentrateurs, frigidaire ou autres objets connectés ayant de plus grandes capacités et un plus grand éventail de fonctionnalités.

En ce qui concerne l'utilisateur, de son point de vue, le système de connexion à distance change, mais il lui permet toujours d'accéder aux objets connectés, de façon plus sécurisée et avec un contrôle plus fin. Ainsi, si des objets connectés sont reliés à des applications sur le cellulaire de l'utilisateur et qu'il se situe à l'extérieur du réseau de la maison, les objets connectés peuvent tout de même se synchroniser et échanger des informations.

Dans notre cas, les objets connectés cherchent toujours à transmettre les données comme dans le cas classique. En revanche, les données ne passent pas directement par le routeur de la maison intelligente avant d'atteindre les applications et les *nuages*. Les communications et données transmises passent plutôt par un intermédiaire, le *gardien* qui est notre point central de l'architecture. Il s'occupe d'une part de la gestion des transmissions des données provenant des objets connectés, mais également des données provenant de l'extérieur de la maison intelligente.

La gestion des transmissions est effectuée avec la solution Home Assistant intégrée au sein du gardien. Non seulement elle permet d'automatiser les différents objets connectés de façon plus harmonieuse que dans un cas traditionnel, mais elle agit également en tant que passerelle entre les objets, le routeur et le monde extérieur.

Dans la figure 3.2, les interactions entre les différents acteurs sont illustrées, à savoir l'utilisateur, Home Assistant et les objets connectés au sein d'une maison intelligente. Afin de pouvoir intégrer un objet connecté au sein d'une maison intelligente à l'aide de Home Assistant, l'utilisateur doit tout d'abord configurer l'objet spécifique sur le concentrateur afin de pouvoir gérer les flux de données ainsi que gérer et automatiser l'objet. Dans cette figure 3.2, Home Assistant s'occupe de l'automatisation et du contrôle en fonction des utilisateurs, des objets connectés ainsi que de la maison intelligente.

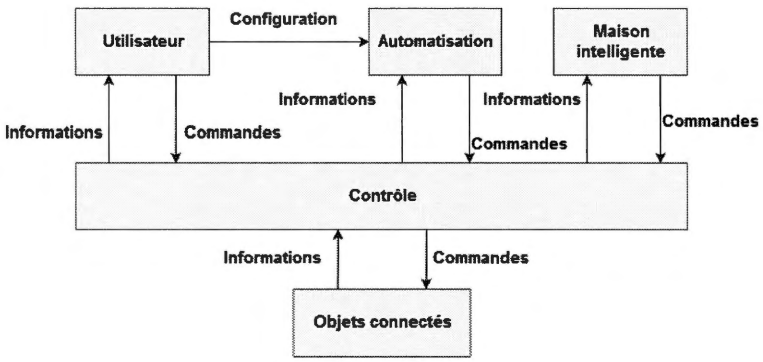

Figure 3.2 Interaction des composants avec Home Assistant

Cette étape permet d'avoir le contrôle sur tous les flux entrants et sortants de la maison (désirables aussi bien qu'indésirables), ainsi que de pouvoir offrir la possibilité à un utilisateur d'accéder à distance et de contrôler les différents objets connectés de manière sécurisée (par exemple : observer l'intérieur de la maison à partir de la vue d'une caméra de surveillance).

Afin de démontrer comment l'automatisation fonctionne, nous montrons dans la figure 3.3, un exemple classique de domotique avec l'interaction et l'automatisation des événements entre une ampoule (lumière) et un détecteur de présence à travers le concentrateur Home Assistant.

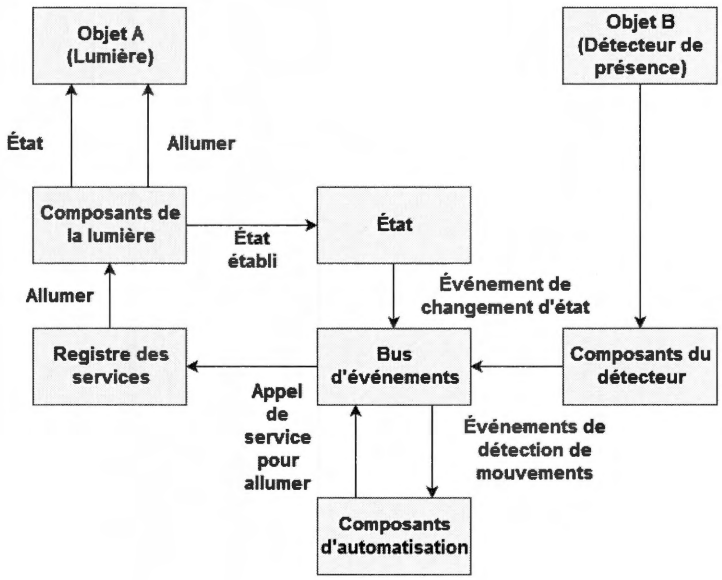

Figure 3.3 Exemple d'automatisation

Dans ce cas, l'objet A qui est une ampoule n'a que deux états, allumé ou éteint. Lorsque le détecteur de présence perçoit une présence dans une pièce spécifique (ou dans la maison), il déclenche un événement lié à la détection de cette présence qui fait appel aux composants d'automatisation d'Home Assistant afin d'activer les services permettant de par la suite d'allumer l'ampoule si elle est éteinte.

De plus, le SDI intégré au gardien s'occupe d'apporter un niveau de contrôle supplémentaire sur le réseau local. Les SDI ont deux sources de données, le trafic du réseau surveillé et des données décrivant des événements sur des machines/objets individuelles (incluant des données dérivées de fichiers de logs, systèmes de tra-

çage, systèmes d'audits, outils de vérification d'intégrité des fichiers et des entrées de registre).

Dans la section 2.6.1, nous avons abordé les différents types, caractéristiques et spécificités des SDI. Dans notre cas, nous nous intéressons aux systèmes de détection d'intrusion qui supposent que le trafic réseau est traité comme la source d'événements qui peuvent déclencher un processus de détection d'intrusion, les NIDS (*Network Based Intrusion Detection System*) (Butun *et al.*, 2014) sont conçus pour surveiller tout le trafic entre les différentes entités d'un réseau et capturer des événements suspects ou anormaux.

Voici les aspects importants de notre SDI (installé au sein du gardien) proposé (illustré dans la figure 3.4) :

1. *Collecte et sélection des données.* La collecte s'effectue sur des données provenant des différentes ressources du réseau (objets connectés), des traces d'audits ou encore les *logs* d'applications. La sélection des données est requise pour effectuer une analyse de sécurité, les données sélectionnées devant être liées à des événements importants du point de vue de la sécurité.
2. *Corrélation et comparaison des données.* Cette étape fait correspondre les événements sélectionnés avec les règles stockées localement pour la détection d'éventuelles violations de sécurité.
3. *Détections d'anomalies, prise de décisions et actions.* Si une anomalie ou violation est détectée, le SDI va prendre une décision qui va déclencher une action sur le réseau local.
4. *Jeu de données de règles.* Les règles sont stockées dans la base de données et sont composées de données des événements traités comme une violation de sécurité observée par le SDI dans le passé. De plus, une mise à jour du jeu de données de règles dans la base de données est effectuée à chaque

violation, anomalie, faille ou vulnérabilité détectée par le SDI, ce qui permet d'améliorer au fil du temps le jeu de données de règles ainsi que de rendre le SDI plus performant et efficace. Les auteurs de (Nguyen *et al.*, 2019) proposent DIOT, un système de détection d'anomalie basé pour l'IoT basé sur l'AF ayant pour but de détecter des objets compromis, dans ce cas il essaye d'apprendre de lui-même grâce à l'AF. Une autre proposition a été présentée par les auteurs (Schneble, 2018) de l'utilisation de l'AF pour les SDI dans les systèmes cyber-physiques médicaux (MCPS). Nous pensons donc que l'AF est une option viable mais encore jeune pour son application sur les SDI ou les systèmes de préventions d'intrusions. Dans notre cas l'utilisation serait principalement pour mettre à jour le jeu de données des règles.

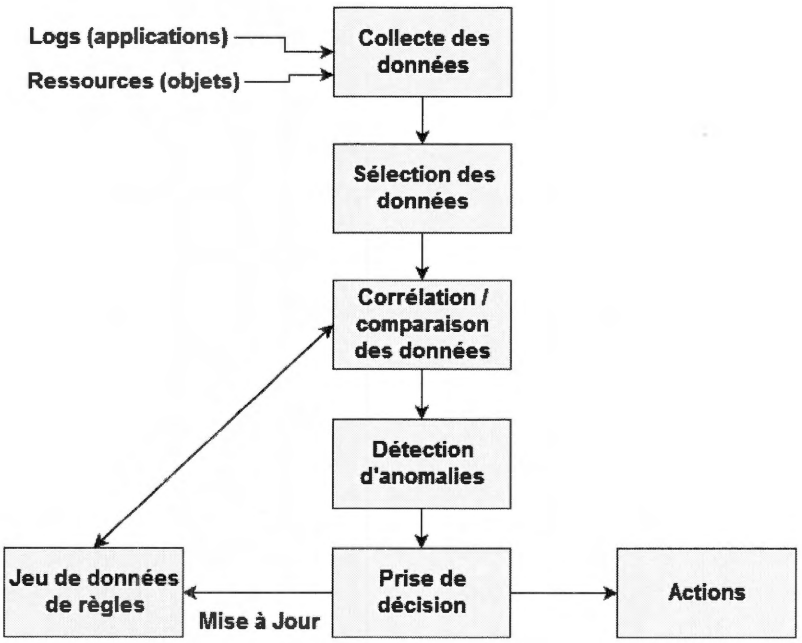

Figure 3.4 Fonctionnement du système de détection d'intrusion

Le troisième composant est une paire de bases de données. La première base de

donnes est pertinente pour le stockage des données provenant des objets connectés à travers le temps, mais également pour les données. La deuxième base de données contient le jeu de règles et les événements du SDI, ce qui permet à l'ensemble de l'architecture d'avoir un contrôle supplémentaire sur les données et de pouvoir les stocker localement.

3.4 Sécurité et protection des données

L'un des plus gros points faibles de l'automatisation de la domotique est la sécurité et la protection des données, qui est due principalement au fait que l'environnement est très hétérogène. Ainsi, chaque objet peut être configuré de différentes manières et possède des capacités de calculs et d'espace hétérogènes. Lorsqu'une maison intelligente est composée d'objets connectés qui contrôlent les serrures, les alarmes, les lumières, la surveillance, la cuisinière, les feux de gaz. Il devient très important d'être vigilant concernant la sécurité et la protection des objets connectés.

En ce qui concerne l'aspect de la sécurité et de la protection des données d'utilisateurs, il existe plusieurs solutions disponibles avec un concentrateur tel que Home Assistant. Le fonctionnement du SDI (figure 3.4) ayant déjà été abordé lors de la section précédente, nous ne reviendrons pas dessus dans les sections suivantes.

L'application de SDI dans un environnement hétérogène tel que l'IdO est plus compliqué que dans les réseaux plus traditionnels. Néanmoins deux approches pour les SDI basés sur des règles comme celui que l'on décrit ont été décrites dans la littérature. Tout d'abord, la première a été proposée par (Jun et Chi, 2014) qui dans ce cas ont conçu l'architecture du SDI sur la base du modèle du traitement des événements, dans ce cas les règles sont stockées dans un répertoire de *pattern* de règles. La deuxième approche proposée par (Eswari et Vanitha, 2013), est un

cadre (*Framework*) de SDI comportant trois phases, à savoir, une phase d'audit local suivie d'une phase d'application des règles puis d'une phase de détection d'intrusion.

D'autres travaux ont été proposés pour des SDI déployés au sein de l'IdO formalisé dans l'étude effectuée par (Zarpelao *et al.*, 2017) confirmant la possibilité de notre proposition, certains basés sur la détection d'anomalies, d'autres sur une approche sur une hiérarchisation d'efficacité d'énergie ou bien sur la détection distribuée, etc.

3.4.1 Points importants de sécurité pour l'utilisation d'Home Assistant

Certains points importants sont à prendre en compte en utilisant une solution telle qu'Home Assistant. Il s'agit d'une base de recommandations dont nous nous sommes inspiré afin de par la suite présenter nos propres recommandations de sécurité :

- Protéger l'interface Web d'accès au concentrateur avec un mot de passe fort (permettant ainsi d'éviter les attaques par force brute ou encore de deviner le mot de passe grâce à de l'ingénierie sociale).
- Sécuriser l'hôte. Par exemple, dans le cas d'un ordinateur sous distribution Linux, il faudrait suivre les recommandations de sécurité spécifiques à cette distribution. Les recommandations spécifiques vont varier de distribution en distribution, mais des guides de sécurité sont fournis par certaines distributions telles que *Debian* ou *Red Hat* par exemple.
- Restreindre l'accès réseau aux objets intelligents en vérifiant les flux de communications, les ports ouverts et les ports utilisés afin de pouvoir visualiser l'état courant du réseau et de pouvoir modifier en fonction des besoins pour améliorer le contrôle général ainsi que la sécurité.

— Utiliser le protocole *SSH (Secure Shell)* pour l'authentification. Il permet de sécuriser les communications en chiffrant les échanges entre deux entités permettant ainsi de protéger les informations échangées.

Le but principal d'utiliser un concentrateur est de toujours pouvoir accéder au réseau de la maison ainsi qu'aux objets connectés. Pour cela, il faut également que les communications entre les utilisateurs itinérants et le réseau local de la maison intelligente soient sécurisées. Une fois encore, plusieurs options sont possibles avec la solution Home Assistant :

— Protéger les communications avec *TLS/SSL (Transport Layer Security/Secure Socket Layer)* qui permet de protéger les échanges d'informations en apportant l'authentification du serveur, la confidentialité et l'intégrité des données.
— Anonymiser les communications avec *Tor* (Syverson *et al.*, 2004), qui est un service de communication anonyme à faible latence basé sur le routage de type oignon (*Onion Routing*). Plus précisément, il s'agit d'un réseau de communication anonyme spécifiquement conçu pour rendre anonyme les applications basées sur *TCP* (telles que navigation Web, messageries instantanées, SSH, etc.).
— Dans certains cas (notamment avec l'utilisation de fureteurs), il faudra en utiliser des certificats qui sont signés par une entité qui possède une identité reconnue par le système (autorité d'authentification/certification).
— Activer le filtrage d'adresses *IP* (en bloquant ou autorisant certaines adresses) et configurer un nombre de tentatives de connexions bas (exemple classique : trois tentatives de connexions, similaires au nombre de tentatives pour des mots de passe).
— Utiliser un proxy qui a pour rôle d'agir d'intermédiaire entre deux entités tentant d'échanger des informations, ce qui permet de faciliter, mais

également de surveiller les communications.

— Mettre en place un réseau privé virtuel (*Virtual Private Network*), ce qui permet de créer une connexion sécurisée sur un réseau qui n'est pas sûr (entre la maison et l'extérieur par exemple). Cette technique permet d'accéder à distance à un réseau en particulier, même si toutes les données protégées ne sont pas toujours chiffrées, il est également possible de valider l'intégrité des données échangées.

— Utiliser un tunnel *SSH (Secure Shell)* permettant de se connecter à l'interface utilisateur d'Home Assistant.

3.4.2 Notre proposition

Chaque réseau de domotique est différent, mais dans l'ensemble, voici les recommandations importantes que nous proposons pour toutes maisons intelligentes.

Premièrement, il faut séparer le réseau local Wi-fi en deux. Le premier sera un réseau pour les invités, qui est très réglementé, avec des clients incapables de « voir »d'autres objets de différents clients. Le second sera le réseau principal de la maison, où les clients sont capables d'interagir avec les objets d'autres clients. Pour une maison typique, un utilisateur peut être confortable à l'idée de donner aux visiteurs l'accès au réseau Wi-fi afin d'accéder à Internet, mais il est peu probable en revanche que l'utilisateur souhaite que ses invités puissent déverrouiller les portes (ou autres) ou autoriser des objets potentiellement compromis d'affecter le réseau local.

Selon les objets connectés faisant partie du réseau de la maison, il est fortement recommandé d'utiliser un VPN faisant la liaison entre le routeur de la maison liant le monde extérieur et le réseau principal de la maison qui permet aux utilisateurs d'interagir avec les objets connectés. Ceci a pour conséquence que non seulement

un mot de passe pour le Wi-fi est requis, mais il faut également les identifiants liés au VPN afin de pouvoir se connecter au réseau contenant Home Assistant et les objets connectés. Pour certains objets, il est possible de se connecter directement en utilisant un VPN mais pour bons nombres d'entre eux cela n'est pas possible, le VPN sera donc principalement utilisé pour accéder au réseau et à Home Assistant depuis l'extérieur par les utilisateurs itinérants. De plus les flux de données seront ainsi chiffrés et protégés contre les attaques de reniflage.

Pour le gardien qui intègre Home Assistant, nous autorisons uniquement des connexions par hôte local par défaut, combinées avec SSL. Le protocole SSL se situe entre la couche application et la couche transport TCP dans le modèle TCP/IP, dans notre cas il se situe entre la couche application et la couche réseau étant donné que la couche réseau de notre architecture comprend les fonctionnalités et paramètres de la couche transport TCP du modèle TCP/IP. Pour la connexion à distance au réseau local, un VPN est de rigueur et si c'est absolument nécessaire d'avoir une connexion à distance (pour certains objets connectés spécifiques) il faut être certain d'autoriser des adresses IP spécifiques. Les connexions IP proviennent tant bien des objets ainsi que des infonuages des constructeurs liés aux objets ou encore aux connexions à distance par les utilisateurs itinérants.

La sécurité parfaite n'existe pas, dans notre cas l'attaquant devra alors spécifiquement cibler le réseau local de la maison, comme expliqué dans la section sur le modèle d'adversaires (section 4.5). Même avec toutes ces précautions, nous recommandons de placer le gardien (et donc Home Assistant) dans un réseau démilitarisé (DMZ) et de très fortement réguler le trafic provenant du *LAN (Local Area Network)* vers le gardien. De plus, l'hôte d'Home Assistant doit uniquement autoriser les connexions au port 443 pour permettre aux applications de fonctionner et utiliser SSH afin d'améliorer le contrôle des accès en limitant le nombre de ports ouverts.

Dans ce cas, le réseau démilitarisé est privé et donc seulement accessible sur une plage d'adresses IP spécifiques privée. De plus, nous préconisons d'utiliser l'accès au tunnel SSH par le LAN et de bloquer toutes connexions externes. De cette manière, si l'utilisateur souhaite réévaluer ou changer des paramètres de configuration de Home Assistant, il devra être sur le LAN (soit localement ou par un accès VPN). Enfin, il faut toujours implémenter un pare-feu à l'aide d'IPtables afin de rendre l'hôte plus robuste.

CHAPITRE IV

IOTFLA - PRINCIPE DE FONCTIONNEMENT ET PROPOSITION D'IMPLÉMENTATION

4.1 Résumé et architecture proposée

Dans le chapitre précédent, les composants clés permettant de poser les fondations de notre architecture ont été détaillés et expliqués. Nous pouvons désormais ajouter les deux derniers composants importants permettant de définir l'architecture dans sa globalité. Nous verrons en particulier, comment tous les composants interagissent et fonctionnent entre eux, mais également l'impact sur la sécurité et la protection des données personnelles. Enfin, nous détaillerons plusieurs scénarios selon les types d'objets connectés utilisés au sein d'une maison intelligente.

4.2 Utilisation de l'approche d'apprentissage fédéré dans l'IdO

Dans la section 2.6.3, nous avons décrit ce qu'était la méthode d'AF (Fed, 2017) ainsi que son fonctionnement dans un contexte d'amélioration de prédiction de textes. Bien qu'étant encore une méthode récente n'ayant que très peu été utilisée dans le domaine de l'IdO et des maisons intelligentes, nous proposons d'intégrer cette méthode au sein de l'architecture. En effet, l'utilisation de l'AF au sein de la domotique pourrait cela pourrait offrir une meilleure personnalisation des services apportés par les objets connectés en prenant en compte des informations

sur l'environnement provenant d'autres objets connectés similaires dans d'autres réseaux de domotique.

Un des premiers cas d'utilisation de l'AF se faisait dans un contexte d'amélioration de modèles de prédiction de textes sur téléphones intelligents. En effet, l'utilisation de l'AF présente plusieurs avantages dans notre contexte. L'AM centralisé classique est l'architecture la plus utilisée de nos jours. Cependant, la séparation entre les algorithmes d'apprentissages et les données des utilisateurs représente un problème majeur pour le domaine de l'IdO (et plus particulièrement des maisons intelligentes) dû aux ressources limitées (aussi bien provenant des objets connectés que des réseaux).

De plus, l'AF permet d'enlever plusieurs contraintes majeures inhérentes à l'AM centralisé. En effet, le coût de communication est réduit, car seuls les modèles sont transférés de temps en temps et non de gros volumes de données en temps réel, ce qui impacte directement la latence possible sur les réseaux. Cela permet également de garder les données d'utilisateurs au sein du réseau local plutôt que de les stocker sur des serveurs distants, permettant donc de renforcer l'aspect de protection de la vie privée.

4.2.1 Modèles d'objets connectés pour l'apprentissage fédéré

Dans le cas de la domotique et de l'IdO, le modèle d'un objet connecté est une structure de données contenant des logiques et des connaissances d'un réseau d'AM entraîné à résoudre un problème spécifique. Depuis quelques années, l'AM se dirige vers une application de cette méthode sur le bord du réseau, à savoir vers les objets connectés et les téléphones intelligents par exemple.

Il est possible d'utiliser une solution comme *TensorFlow Lite* (TFL) [1] pour implémenter l'AF au sein d'objets connectés destinés aux maisons intelligentes. Dans ce contexte, il est possible d'apprendre son propre modèle d'entraînement pour les objets ou d'utiliser des modèles pré-entraînés (de manière similaire au concept d'avoir un modèle générique provenant du constructeur qui est également pré-entraîné). D'ailleurs, sur cette plate-forme, des modèles pré-entraînés existent par exemple pour la classification d'images, la détection d'objets, les réponses intelligentes, l'estimation de position, etc.

TFL a pour but de supporter une variété de scénarios d'apprentissage distribué dans lesquels le code du modèle d'AM qui est développé peut être exécuté sur un grand nombre de clients hétérogènes avec des capacités diverses. Bien qu'à une extrémité du spectre, dans certaines applications, le client pourrait être de puissants serveurs de base de données, de nombreuses utilisations possibles par TFL sont dédiées pour les objets connectés qui ont besoin d'être intégrés malgré des ressources limitées.

Il y a deux niveaux d'agrégation pour l'AF, le premier est une agrégation au niveau de l'objet et le second se situe au croisement de différents objets. Pour l'agrégation locale, cela correspond à une agrégation de plusieurs lots d'exemples appartenant à un même client (objet). Cette étape s'applique à la fois aux paramètres du modèle (variables), qui continuent d'évoluer séquentiellement au fur et à mesure que le modèle est formé localement. Il est également possible d'extrapoler des statistiques (moyenne, précision ou encore d'autres paramètres) que le modèle mettra à jour localement à mesure qu'il se répète dans le flux local de données de chaque client. L'agrégation fédérée fait référence à l'agrégation entre plusieurs objets au sein du système. Il s'applique aussi aux paramètres du modèle, dont la

1. https ://www.tensorflow.org/lite/

moyenne est calculée pour l'ensemble des clients, ainsi qu'au fil du temps ou le modèle est exporté à la suite d'une agrégation locale.

4.2.2 Exemples de modèles de TFL

Couramment, TFL propose d'ores et déjà plusieurs modèles préfaits et prêts à l'utilisation pour l'IdO, de plus il est possible de s'inspirer de ses modèles pré-entraînés pour créer nos propres modèles pour différents cas d'utilisations. Voici quelques exemples :

- *Classification d'images.* Permet de tester une solution de classification d'images d'un modèle pré-entraîné qui peut reconnaître 1000 différents types d'items provenant d'une caméra de téléphone intelligent. Ce cas pourrait par exemple être utilisé pour les caméras externes et internes à la maison intelligente.
- *Détection d'objets.* Il s'agit d'une application caméra qui détecte en continu les objets dans les images vues par la caméra arrière d'un téléphone intelligent.
- *Estimation de la pose.* Il s'agit d'une application qui détecte en permanence les parties du corps dans images vues par la caméra d'un téléphone intelligent.
- *Reconnaissance de voix.* Cette application permet d'effectuer une reconnaissance de voix ou dialogue sur téléphone intelligent.
- *Reconnaissance de mouvements.* Dans ce cas, un réseau de neurones est entraîné afin de reconnaître des mouvements capturés par une caméra.
- *Réponses intelligentes.* Génère des suggestions de réponses dans le contexte d'un échange de messages.
- *Questions et réponses.* Les réponses utilisent des requêtes basées sur les informations extraites d'une archive de messages.

4.3 Intégration d'un protocole d'agrégation sécurisée des données

Le second composant est l'intégration d'un protocole d'agrégation sécurisée des données, comme présenté dans la section 2.6.4. Depuis la création du premier protocole d'agrégation sécurisée des données (Hu et Evans, 2003), un nombre conséquent de protocoles ont été proposés.

Les protocoles peuvent être classés selon plusieurs objectifs de sécurité, à savoir, la confidentialité, l'intégrité, l'authenticité et la fraîcheur. Dans notre cas, la confidentialité et l'intégrité sont les deux critères que nous souhaitons conserver, car ils sont important afin de pouvoir protéger les communications ainsi que les informations échangées. De plus, le protocole choisi pour l'architecture doit être utilisable dans un environnement de maison intelligente. Ceci implique que nous devons prendre en compte certains paramètres supplémentaires, comme le type d'agrégation (saut à saut, cluster, etc...), la consommation énergétique et réseau, la capacité du réseau ainsi que les ressources environnementales disponibles qui sont liés aux capacités des objets connectés).

En se basant sur les objectifs de sécurité pour la confidentialité et l'intégrité, seuls quelques protocoles correspondent aux besoins de sécurité au sein des réseaux sans-fils pour la domotique, les protocoles suivants ont été détaillés dans la section 2.6.4 (en nous basant sur le tableau de classification des protocoles 2.6.4 provenant de l'article (Soni et Randhawa, 2017)).

Le protocole SIA (Przydatek *et al.*, 2003) est conçu pour des grands réseaux de capteurs sans fils, ce qui n'est pas le cas de notre architecture, qui est plutôt un réseau de petite envergure et ne contient qu'un nombre limité de capteurs et d'objets. Nous avons donc renoncé à l'intégration de ce protocole. Tout comme le protocole SIA, le protocole IPHCDA (Yang *et al.*, 2008) est conçu pour de grands

réseaux de capteurs sans-fils, il ne convient donc pas pour notre cas d'utilisation.

Malgré le fait que le protocole SDAP (Mahimkar et Rappaport, 2004) comprend de nombreux points forts (intégrité des données, confidentialité, authentification des sources, etc.), le coût de transmission ainsi que l'énergie utilisée de ce protocole sont élevés, ce qui ne le rend pas adapté à notre contexte. Dans le cas des deux schémas proposés par les auteurs du protocole RDCA (Chen *et al.*, 2012), l'intégrité des données n'est pas garantie ce qui ne convient pas dans notre cas d'utilisation également.

En se basant sur les critères et les contraintes que nous avons précédemment définies, nous avons tout d'abord réduit le nombre grand de protocoles à six. Par la suite en effectuant une recherche plus approfondie sur ses six protocoles, nous avons pu identifier les protocoles les plus adaptés pour notre architecture, ce qui nous laisse SecureDAV (Mahimkar et Rappaport, 2004) et ECIPAP (Zhu *et al.*, 2014). Ces deux protocoles semblent être adaptés pour notre situation, architecture et contraintes. ECIPAP offre plus de bénéfices (Confidentialité, Intégrité et Fraîcheur) que SecureDAV (Confidentialité et Intégrité), ce qui nous pousse à choisir le protocole ECIPAP plutôt que SecureDAV, qui reste néanmoins une bonne seconde option.

4.3.1 Fonctionnement du protocole ECIPAP

Tout d'abord, on fait l'hypothèse qu'un arbre d'agrégation (qui représente un arbre composé d'une station de base et d'un certain nombre de noeuds du réseau) est déjà en place dans la phase de déploiement. Si ce n'est pas le cas alors il est possible d'utiliser Tiny AGregation Service for Ad-Hoc Sensor Networks (TAG) (Madden *et al.*, 2002) afin de construire un réseau basé sur un arbre.

TAG permet d'exprimer des requêtes d'agrégations simples et de les distribuer

et exécuter efficacement dans des réseaux de capteurs sans-fil de faible puissance. Avant le déploiement des noeuds de capteurs, chaque noeud de capteur partage une clé privée (k_i), un entier de grande taille (M) et un (ID_i) unique avec la station de base.

L'algorithme de chiffrement symétrique additif homomorphe et la fonction de hachage SHA-1 (qui est une fonction de hachage cryptographique) également prédéterminés. La primitive de chiffrement homomorphe (figure 4.1) se compose des quatre algorithmes suivants. Tout d'abord, l'algorithme de génération de clé prend en entrée r, un nombre aléatoire, et produit en sortie k_r qui est égale à un hachage de clé k et du nombre aléatoire r auquel est appliqué un modulo M. Par la suite, l'algorithme de chiffrement génère le message chiffré c qui est calculé de la manière suivante, $\text{Enc}(m, k_r, M)$ qui est égale à $(m + k_r)$ modulo M, donc le chiffrement contient le message caché grâce à la clé générée à l'étape précédente. Ensuite, l'algorithme de déchiffrement, $\text{Dec}(c, k_r, M)$ revient à calculer $(c - k_r)$ modulo M , ce qui génère un texte clair qui est égal au message chiffré c moins la clé générée k_r modulo M. Enfin l'algorithme d'addition de message permet d'additionner deux messages chiffrés et de calculer $c = c_i + c_j$ modulo M sans avoir besoin de déchiffrer les messages et tel que le déchiffrement du message c soit égal à $m_i + m_j$.

Le protocole est composé de trois étapes par la suite :

1. *Phase de diffusion de la requête.* Dans chaque tour d'agrégation, la station de base choisie au hasard un nombre aléatoire r et la requête d'agrégation telle que COUNT, SOMME ou MOYENNE. Grâce à sa puissante capacité de communication, elle peut diffuser ce message requête à l'ensemble du réseau avec le nombre aléatoire r en utilisant la méthode d'authentification uTESLA (Perrig *et al.*, 2002) qui a pour but de diffuser les requêtes au-

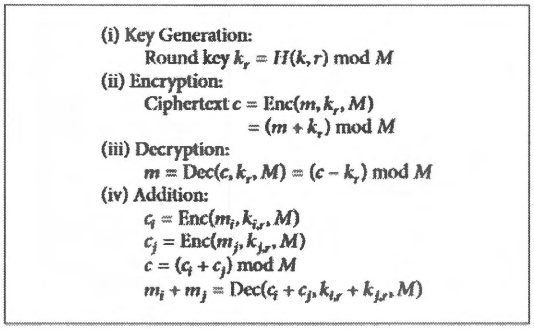

Figure 4.1 Primitive de chiffrement homomorphe (Zhu *et al.*, 2014)

thentifiées ainsi que les nombres aléatoires. Lorsque les noeuds reçoivent les messages requêtes, ils les stockent dans leurs mémoires vives et commencent la phase d'agrégation.

2. *Phase d'agrégation des données.* Les noeuds de capteurs collectent des données sur l'environnement appartenant à $[V_{min}, V_{max}]$ telles que dans cet exemple la température et la valeur et définit la valeur égale à $V_{max} - v$, le compte (count) est égale à 1. Puis la génération de clés temporaires $k_{i,r}$, $k'_{i,r}$, $k''_{i,r}$ est effectuée) en suivant le principe expliqué précédemment :

$$k_{i,r} = H(k_{i,r}) \, mod \, M,$$

$$k_{i,r} = H(k_{i,r}) \, mod \, M,$$

$$k_{i,r} = H(k_{i,r}) \, mod \, M.$$

Puis le déchiffrement est effectué comme ceci :

$$c_{count,i}, k_{i,r}, M,$$

$$c_{value,j}, k'_{i,r}, M,$$

$$c_{\overline{value},i}, k''_{i,r}, M.$$

Les noeuds de capteurs peuvent alors calculer les codes d'authentification de messages (MAC) de la manière suivante :

$$MAC_i = H(c_{count,i}||c_{value,i}||c_{\overline{value},j}||ID_i).$$

Le tuple de données peut alors être créer de la façon suivante :

$$u_i\langle c_{count,i}, c_{value,i}, c_{\overline{value},j}, MAC_i\rangle$$

Lorsque les noeuds préparent les tuples de données, ils les envoient à leurs noeuds parents (illustré dans la figure 4.2).

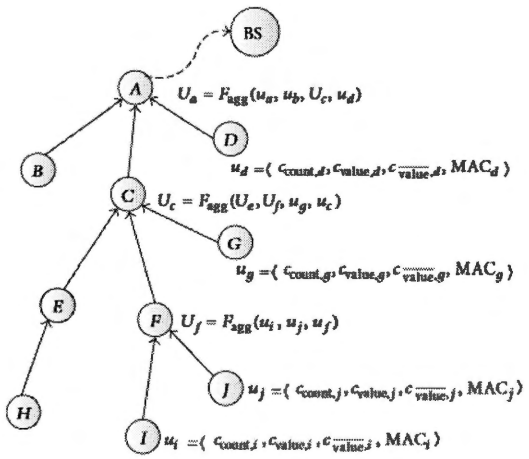

Figure 4.2 Phase d'agrégation du protocole ECIPAP (Zhu *et al.*, 2014)

Il y a $N = 10$ noeuds dans le réseau et le noeud F à deux noeuds enfants I et J. Quand le noeud F reçoit de tuple de données u_i et u_j qui sont envoyés depuis les noeuds I et J, il les agrège avec le tuple de données u_f créer par lui-même tel que : $U_f = \text{Fagg}(u_i, u_j, u_f)$. Le noeud F envoie ce résultat d'agrégation intermédiaire à son noeud parent C. Le résultat est alors transmis au plus haut niveau de l'arbre jusqu'à que le résultat

atteigne la station de base qui peut déchiffre le résultat de l'agrégation (représenter dans la figure 4.3) :

$$\text{count}_{\text{agg}} = \text{Dec}\left(U_a \cdot \text{count}, \sum_{i=1}^{n} k_{i,r}, M\right),$$

$$\text{value}_{\text{agg}} = \text{Dec}\left(U_a \cdot \text{value}, \sum_{i=1}^{n} k'_{i,r}, M\right),$$

$$\overline{\text{value}}_{\text{agg}} = \text{Dec}\left(U_a \cdot \overline{\text{value}}, \sum_{i=1}^{n} k''_{i,r}, M\right).$$

Figure 4.3 Déchiffrement du résultat par la station de base (Zhu *et al.*, 2014)

La station de base BS peut aussi calculer l'agrégation du MAC par la fonction suivante :

$$\text{MAC}_{\text{agg}} = \bigoplus_{i=1}^{n} m_i \cdot \text{MAC}.$$

Figure 4.4 Fonctions d'agrégations (Zhu *et al.*, 2014)

3. *Phase de vérification des résultats.* Lorsque le résultat final de l'agrégation est reçu, la station de base diffuse les données agrégées sur l'ensemble du réseau en utilisant une méthode authentifiée (uTESLA). Pour permettre la vérification des résultats, chaque noeud de capteur enverra un message de vérification à son noeud enfant, elle peut être considérée comme un processus inverse d'agrégation de données (résumé dans la figure 4.5).

   La station de base envoie les tuples de données finaux U'_a au noeud A. Le noeud A peut supprimer les données envoyées par les noeuds B, C et D provenant du tuple U'_a. L'opération de calcul inverse en tant que $\ominus$. Par exemple, nous pouvons avoir le tuple de données créé par le noeud A comme

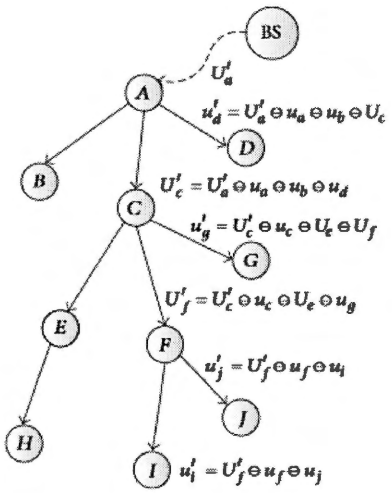

Figure 4.5 Phase de vérification des résultats du protocole ECIPAP (Zhu *et al.*, 2014)

ceci $u'_a = U'_a \ominus u_b \ominus U_c \ominus u_d$. En comparant U'_a avec le tuple de données u_a créé par le noeud A, il peut vérifier si ses propres données ont été au résultat final d'agrégation.

Chaque noeud de capteur peut vérifier si ses propres données ont été ajoutées aux données d'agrégations en comparant celles-ci aux données envoyées par les noeuds parents. Si le résultat passe la vérification, alors chaque noeud de capteur prépare un message d'authentification CAM tel que $CAM(k_i||r||OK)$ et l'envoie à la station de base par saut. Si la vérification échoue, le message $CAM(k_i||r||NO)$ sera envoyé. Lorsque les noeuds de capteurs intermédiaires reçoivent ces messages d'authentification, ils les agrègent à l'aide de la fonction d'agrégation MACagg (figure 4.4).

La station de base peut également calculer ce message d'authentification avec ses propres données stockées. La comparaison de ces deux messages reçus (celui vérifié et celui non vérifié) d'authentification permet donc de

vérifier si toutes les données de détection sont ajoutées au résultat final de l'agrégation. Si le message d'authentification réussit la vérification, la station de base accepte ce résultat d'agrégation, sinon elle l'ignore.

4.4 Fonctionnement et scénarios d'utilisation

Maintenant que tous les composants ont été présentés, nous pouvons introduire l'architecture complète de notre proposition. Une autre particularité de l'architecture est qu'elle n'est pas fixée sur un seul et unique cas d'usage, car elle permet de prendre en compte trois scénarios différents qui dépendent exclusivement des objets connectés composant une maison intelligente.

L'objectif d'utiliser l'approche d'apprentissage fédéré dans le domaine des maisons intelligentes est d'améliorer le fonctionnement et le comportement des modèles des objets connectés à travers le temps. Les modèles sont par la suite régulièrement envoyés aux nuages des constructeurs respectifs afin de mettre à jour les modèles génériques communs avec les autres modèles reçus d'autres utilisateurs possédant les mêmes objets connectés provenant d'autres maisons intelligentes, ce qui est semblable au cas initial présenté avec l'amélioration des modèles de prédictions de textes. Une fois que les modèles génériques communs ont effectué leurs mises à jour, les nuages des constructeurs redistribuent les nouveaux modèles génériques communs améliorés à tous les objets connectés concernés et ceci toujours basé de façon similaire au cas des modèles d'amélioration de prédiction de textes.

Les objets connectés bénéficient d'une amélioration sur le temps basé sur le comportement de son propriétaire, mais également des comportements des autres utilisateurs possédant les mêmes objets connectés. En revanche, l'utilisation de l'approche d'apprentissage fédéré entraîne des coûts additionnels sur les réseaux et les communications quand les mises à jour sont effectuées. Dans le cas d'un scénario

que l'on peut qualifier de « classique », quand une mise à jour doit être effectuée, cela demande uniquement d'avoir le réseau sans-fil de la maison intelligente connecté au monde extérieur.

Par la suite, le constructeur (unique à chaque objet) envoie la mise à jour du nouveau modèle générique commun aux différents réseaux possédant l'objet connecté en question, cette étape pouvant demander dans certains cas l'approbation de l'utilisateur et par la suite l'objet connecté applique la nouvelle mise à jour du modèle.

Un autre point important à prendre en compte de l'impact de l'utilisation de l'apprentissage fédéré est que les mises à jour des modèles ne se font pas à sens unique, mais dans les deux directions. En effet, dans un cas une mise à jour du modèle va de l'objet connecté au nuage du constructeur et dans l'autre du nuage du constructeur à l'objet connecté. Ceci qui va forcément entraîner une charge supplémentaire en comparaison à un scénario classique.

Enfin, étant donné que dans notre scénario nous avons un composant qui est clé et central à notre architecture *IoTFLA*, notre gardien est placé entre les objets et le routeur qui font donc la jointure entre le réseau de la maison intelligente et le monde extérieur (donc les infonuages), le coût d'envoi et de réception des mises à jour peut également être augmenté.

Comme présenté précédemment, TFL est conçu pour faciliter l'AM sur les périphériques, au lieu d'envoyer des données d'un serveur à l'autre. Il marche avec une grande variété d'objets connectés, allant de petits micro-contrôleurs jusqu'à de puissants téléphones intelligents. L'application de l'AF dans l'IdO est donc encore jeune, mais commence à être de plus en plus développée et de nouveaux modèles préfaits voient régulièrement le jour.

Quelques types de modèles ont été également montrés avec l'utilisation de TFL. La majorité des cas présentés sont des utilisations par des caméras ou des téléphones intelligents, ce qui est restreint, mais ce sont des modèles prédéfinis pour apprendre à utiliser l'AF dans le contexte de l'IdO. Il est possible d'utiliser les mêmes approches de TFL pour créer des modèles spécifiques à une utilisation unique d'un type d'objet connecté.

Nous avons songé à quelques exemples pour les objets connectés au sein de la domotique. L'application de l'AF sur les caméras de surveillance interne et externe pourrait améliorer la détection de mouvement et diminuer le temps de réponse du déclenchement d'une alerte en cas d'intrusion d'un individu. Il serait également possible d'avoir l'AF appliqué à un compteur électrique afin de diminuer sur le temps les consommations énergétiques de la maison au global, ou encore sur la reconnaissance de voix et de dialogue des solutions de concentrateurs. Il est possible d'imaginer une large application de l'AF sur des objets connectés simples permettant d'améliorer leurs façons d'agir sur le temps et donc d'améliorer leurs efficacités à effectuer leurs tâches.

Dans notre proposition, nous partons donc de ce constat et que même si l'AF peut être appliqué à de petits micro-contrôleurs et peut couvrir une large gamme d'objets connectés, nous avons prévu le cas où certains objets connectés ne pourraient quand même pas appliquer ou utilisé directement l'AF. Pour cela, nous avons divisé notre proposition en trois scénarios expliqués ci-dessous :

Scénario 1 : Objets connectés auto-améliorés

Dans le premier scénario (illustré dans la figure 4.6), nous supposons les objets connectés au sein du réseau de la maison intelligente ont les capacités nécessaire afin d'utiliser l'AF pour pouvoir améliorer leurs propres modèles.

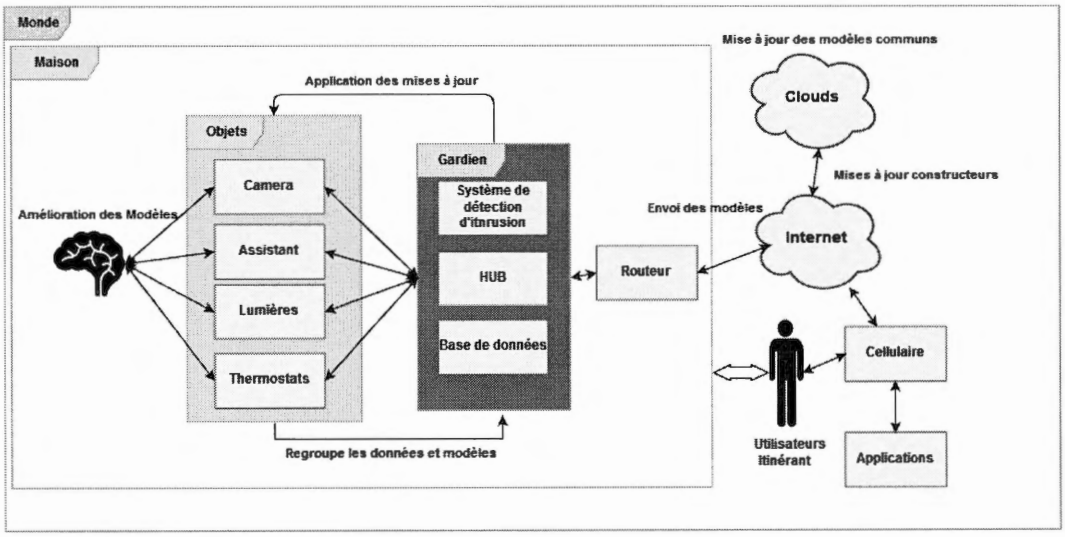

Figure 4.6 Schéma de l'architecture - Scénario 1

Scénario 2 : Le gardien supervise tout

Dans le second scénario (illustré dans la figure 4.7), contrairement au premier scénario, nous supposons que chaque objet connecté dans la maison intelligente n'a pas les capacités et ressources et donc n'est pas capable d'utiliser l'approche d'AF. Dans ce cas, les modèles d'objets connectés sont stockés et améliorés avec le temps au sein du gardien. En revanche cela demande davantage de communication entre les objets connectés et le gardien, ce qui peut augmenter les consommations réseau et énergétiques.

Scénario 3 : La maison hybride

Le troisième scénario est un hybride (illustré dans la figure 4.8), nous supposons que certains objets connectés sont capables d'utiliser la méthode d'AF et que d'autres objets connectés n'en sont pas capables. La manière globale de fonctionner

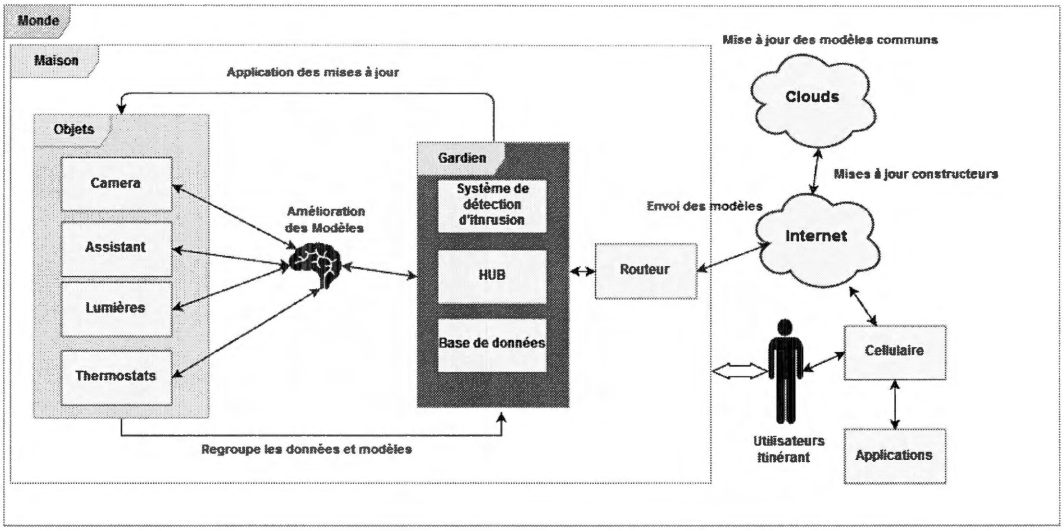

Figure 4.7 Schéma de l'architecture - Scénario 2

ne change pas, il s'agit d'une fusion entre le fonctionnement du premier et second scénario. Les modèles sont améliorés soit au sein du gardien ou sur les objets connectés eux-mêmes s'ils en sont capables.

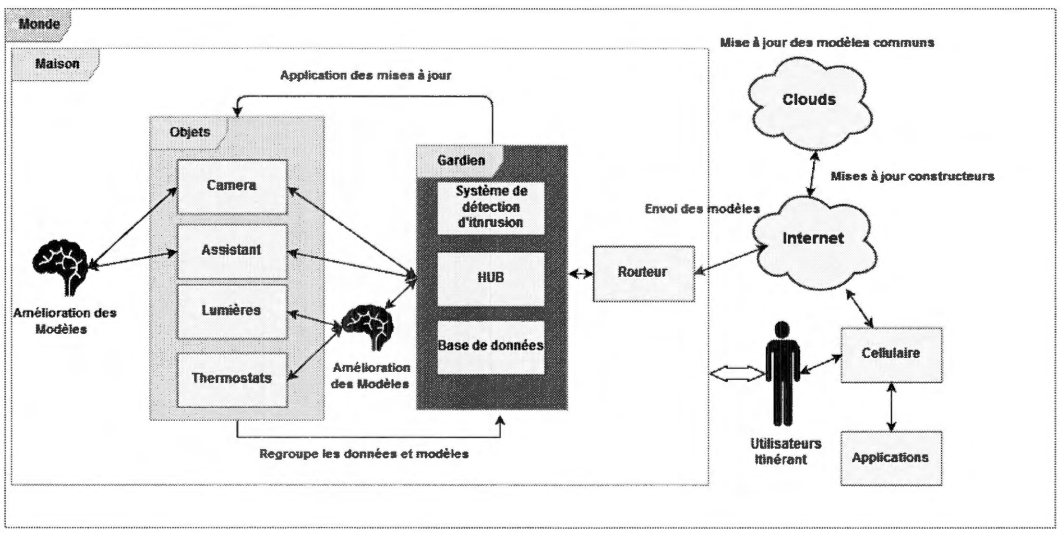

Figure 4.8 Schéma de l'architecture - Scénario 3

4.5 Modèle d'adversaire et analyse de sécurité de notre proposition

Notre proposition, comme dans (Miettinen *et al.*, 2017), ressemble à une installation typique de réseau de domotique. Nous supposons que lorsque des objets connectés sont initialement connectés au réseau de la maison intelligente, ils présentent possiblement des vulnérabilités de sécurité. Cependant, ces objets sont aussi bénéfiques de par leurs fonctions au sein d'un réseau de domotique et qu'ils ne sont pas initialement compromis par l'adversaire. Néanmoins le fait d'avoir un objet connecté infecté avant même d'être au sein du réseau reste une possibilité.

Nous considérons deux types d'attaques, à savoir les attaques locales et les attaques à distance. Dans les attaques locales, l'attaquant tente d'exfiltrer des données ou des informations d'identification de sécurité, de se fier à objet compromis pour infecter d'autres objets ou pour manipuler les informations collectées sur le gardien. Dans le cas d'attaques à distance, l'adversaire peut accéder au réseau des objets connectés (donc de la maison) en utilisant diverses attaques ou techniques que nous avons pu illustrer dans la section 2.5.2.

Dans le chapitre précédent, les principes proposés de sécurité et de protection des données privées ont été détaillés. En plus de ce que nous avons déjà préconisé et prévu d'utiliser au sein de l'architecture, l'ajout de ces deux derniers composants permet des bénéfices supplémentaires par rapport à un scénario plus standard de domotique.

IOTFLA aborde plusieurs aspects en termes de sécurité et protection de la vie privée, le premier est d'essayer de garder un maximum des données sur les objets connectés, sur le gardien donc sur le réseau de la maison intelligente. Le second aspect est de réaliser l'entraînement des différents modèles propres à chaque objet connecté localement et d'utiliser le protocole d'agrégation sécurisée des données

afin de permettre les mises à jour de modèles en utilisant l'approche d'AF.

De plus, l'utilisation de l'AF limite les transferts de données entre la maison intelligente et le monde extérieur (serveurs distants ainsi que les applications), ce qui permet principalement de limiter le nombre d'opportunités pour de potentiels attaquants, ainsi que d'améliorer l'aspect de la protection des données privées en gardant les données en local.

Les bénéfices principaux de cette proposition résident dans le fait que les communications sont à faibles coûts et bas en latence dû à la limitation des échanges de données, mais également la protection privée des données dues au traitement local des données et à l'agrégation sécurisée.

L'architecture prend aussi en compte deux autres aspects importants. Tout d'abord elle permet aux utilisateurs de mieux comprendre ainsi que de contrôler le réseau et ses objets connectés grâce à la mise en place d'Home Assistant au sein du gardien qui permet d'automatiser le fonctionnement des objets. De plus, l'utilisation de l'AF devrait permettre d'améliorer l'efficacité et le fonctionnement des modèles des objets connectés faisant partie intégrante de la maison tout en préservant la vie privée (précédemment illustrer dans les différents scénarios) en faisant le traitement des données localement et en utilisant un protocole d'ASD pour transférer les modèles des objets connectés vers les infonuages des constructeurs.

En revanche, l'un des défis pour la sécurité et la protection de la vie privée de l'utilisation de l'AM directement sur les objets est qu'ils doivent être capables d'apprendre entre eux, mais sans échanger directement les données perçues à cause du caractère privé des données (ce qui est bien plus le cas dans la domotique). Ceci souligne le besoin d'avoir le contrôle sur les échanges et les interactions entre les objets connectés au sein d'un réseau de domotique.

En ce qui concerne l'ASD, d'une manière générale, on peut penser que les protocoles de préservations de confidentialité et de l'intégrité doivent satisfaire aux conditions suivantes :

— Les données (provenant des objets) chiffrées ne doivent être déchiffrées qu'à la station de base et non sur les noeuds intermédiaires afin de garantir la confidentialité de bout en bout.
— Si les adversaires manipulent les données, la station de base peut vérifier leur intégrité par le contrôle du résultat final du regroupement.
— La consommation énergétique doit être raisonnable.

Les deux premières conditions sont plus orientées sur la sécurité et la protection des données et la troisième répond plus aux contraintes des réseaux sans-fils.

Maintenant que nous avons présenté l'architecture *IoTFLA* dans son entièreté, nous pouvons ouvrir le dernière section de ce mémoire avant de conclure, abordant comment nous pourrions implémenter une telle architecture et les contraintes liées.

4.6 Proposition d'implémentation de l'architecture

Ce chapitre a pour but d'approfondir certains points concernant l'architecture IoTFLA, en particulier comment serait-il possible d'implémenter l'architecture en déterminant les différentes options qui nous sont disponibles.

Afin de proposer une implémentation pour tester notre architecture, il nous faut prendre en compte chaque composant faisant partie de IoTFLA et comment nous pouvons les faire fonctionner et interagir entre-eux. Nous suggérons également qu'il est possible d'effectuer une simulation plutôt que d'avoir une version physique des composants.

De nos jours, il existe de nombreux simulateurs de réseaux de capteurs sans-fils

ainsi que de simulateurs pour l'internet des objets qui permettent de mettre en place des environnements de test pour des cas similaires au nôtre (des solutions telles que *DPWSIM* (Han *et al.*, 2014), *iFogSim* (Gupta *et al.*, 2017), *IoTSim* (Zeng *et al.*, 2017), *SinIoT* (Sotiriadis *et al.*, 2014), *CupCurban* (Mehdi *et al.*, 2014), *Cooja* (Osterlind *et al.*, 2006), *OMNET++* (Varga et Hornig, 2008), *NS-3* (Henderson *et al.*, 2008), *QualNet* (Varga et Hornig, 2008), *FitIoTLAB* (Des Roziers *et al.*, 2011), *Smart Santender* (Sanchez *et al.*, 2014), *JOSE* (JOS, 2016) , etc.).

Dans l'article (Chernyshev *et al.*, 2018), les auteurs offrent une comparaison d'un grand éventail de solutions basées sur des critères tels que l'étendue, la dernière mise à jour en date, le nombre de citations, le type, le langage utilisé, la mobilité, l'aspect fonctionnel, le nombre de standards intégrés, le domaine d'application ou encore la simulation d'attaques de cybersécurité. Bien entendu ces solutions nous sont disponibles si nous souhaitons avoir une version simulée plutôt que physique. Même dans le cas d'un environnement simulé intégralement, nous avons besoin d'objets intelligents, d'un nuage, d'un ordinateur, d'un routeur, d'un réseau, d'un téléphone intelligent, d'une base de données et d'un système de détection d'intrusion.

Nous allons partir du principe que nous envisageons de mettre en place un environnement de test entièrement simulé pour IoTFLA.

La première étape serait le déploiement du réseau de la maison intelligente ce qui inclut le déploiement de l'arbre pour le protocole d'agrégation sécurisée de données ECIPAP (Zhu *et al.*, 2014), le routeur, les objets intelligents et le gardien. Le *gardien* sera mis en place sur un ordinateur, étant donné que le composant principal du *gardien* est la solution d'automatisation Home Assistant et qu'il est possible de le déployer sur un système d'opérations (*Linux*, *MacOS* ou *Windows*).

Nous suggérons en revanche d'orienter le choix vers celui qui offre le plus de flexibilité et de contrôle afin de prendre en compte l'aspect d'évolutivité de la maison intelligente ainsi que des objets connectés. Dans notre cas n'importe quel système d'opération basé sur une architecture Linux semble être le choix le plus approprié (avec des systèmes d'opération tels que *Debian*, *Fedora*, *CentOS*, *ArchLinux*, etc.) pour déployer et gérer la solution Home Assistant, le système de détection d'intrusion et le gestionnaire de base de données qui est utilisé pour Home Assistant qui est par défaut *SQLite*.

La seconde étape consisterait à la mise en place et la configuration entre chaque composant intégré dans l'étape précédente, comment les connecter et faire en sorte qu'ils interagissent entre eux correctement. Ensuite, nous appliquerions les recommandations de sécurité et de protection des données personnelles que nous avons détaillées dans la section 3.4.

Durant cette étape, le réseau pour les hôtes est créé, de manière à être très régulé et restrictif. Le réseau principal est déployé avec l'ajout d'un réseau virtuel privé. Il existe plusieurs solutions permettant de mettre en place un réseau virtuel privé, comme *OpenVPN* par exemple. Le VPN aura pour rôle de permettre d'accéder depuis le monde extérieur au réseau principal et donc aux objets connectés.

Par la suite, la mise en place du pare-feu avec la solution IPTables au sein du système d'exploitation sera nécessaire afin d'améliorer l'imperméabilité du gardien. Il serait également possible de placer le gardien dans une zone démilitarisée ou dans un sous-réseau et de ne permettre l'accès uniquement par SSH à travers le LAN. Ceci permettrait de refuser toutes connexions extérieures à l'exception du port 443 et de SSH pour autoriser ainsi l'utilisation d'applications.

Une fois que les politiques de sécurité sont mises en places, la troisième étape consisterait à configurer l'automatisation des objets intelligents avec la solution

Home Assistant (ce qui inclut également la gestion des communications et de la sécurité). Par la suite, nous appliquerions le protocole ECIPAP pour l'agrégation sécurisée des données ce qui permettrait d'initialiser la mise en place de l'approche d'AF (en prenant en compte les trois différents scénarios expliqués dans le chapitre précédent).

La dernière étape serait la mise en place du SDI au sein du gardien, permettant ainsi d'avoir le réseau de la maison et de ses composants activés fonctionnant normalement. Une fois tout cela activé, nous pourrions par la suite connecter le réseau de la maison intelligente aux différents acteurs du monde extérieur autorisés (nuages, applications et utilisateurs itinérants). Afin de pouvoir tester les trois différents scénarios détaillés dans le chapitre précédent, il faut prendre en compte que les objets connectés ont différentes fonctions et fonctionnalités (caméra, thermostats, capteurs, verrou, alarme, lumière , etc.) mais également qu'ils ont différentes capacités de calculs, ressources et consommations énergétiques qui peuvent varier selon le type d'objet ainsi que de ses fonctionnalités.

En effet, un objet ayant peu de fonctionnalités telles qu'un détecteur de mouvement n'a pas les mêmes besoins en capacité de calculs et de ressources qu'un objet effectuant plus qu'une simple perception et récolte d'un seul et unique type de données. Le capteur dans cet exemple n'est pas forcement connecté à une alimentation électrique et donc possède des contraintes énergétiques importantes qui sont liées à sa propre batterie. Si nous prenons dans le cas contraire, un objet disposant de plus de fonctionnalités telles qu'un concentrateur dispose d'une capacité de calculs et de ressources plus grandes, mais n'a probablement pas de contraintes énergétiques aussi fortes que le détecteur de mouvement, car il est possible qu'il soit connecté par alimentation électrique.

CONCLUSION

L'IdO peut avoir un très grand impact sur la sécurité et la protection de la vie privée, étant donné que le marché de l'IdO n'est pas axé sur la sécurité et sur la protection de la vie privée dans la majeure partie des cas, mais plutôt une course à l'innovation ayant pour but de rendre des objets du quotidien intelligents. Dans le contexte de la domotique plus particulièrement la sécurité et la protection de la vie privée des utilisateurs sont deux aspects importants qui ne sont par forcément couverts de manière appropriée par les différents constructeurs d'objets connectés.

Pour répondre à ces enjeux nous avons fait une proposition dédiée spécifiquement pour le domaine de la domotique de l'IdO axé sur la sécurité et la protection de la vie privée. En combinant différentes technologies et composants nous avons pu faire émerger l'architecture théorique IoTFLA qui est construite de manière à prendre en compte un grand spectre de défis, vulnérabilités et dangers connus qui sont liés à l'IdO. Plus précisément, en combinant tous les différents composants présentés dans les chapitres 3 et 4, IoTFLA propose tout d'abord une mise en avant de la sécurité ainsi que de la protection des données personnelles.

IoTFLA est une architecture complexe qui n'a pas était implémentée et qui reste pour l'instant théorique. Nous espérons qu'en proposant cette architecture à la communauté scientifique, elle inspirera à d'autres idées et concepts dans le but de rendre l'environnement des maisons intelligentes plus sécurisés mais également de le rendre plus intelligent et effectif à travers le temps. Nous pensons que l'approfondissement et l'application d'un composant important tel que l'AF appliqué à l'IdO et des réseaux sans-fils de domotique dans le futur pourrait avoir un impact

bénéfique sur la protection de la vie privée ainsi que la sécurité.

Nous avons également lors des recherches réfléchi à différentes variantes ou encore d'autres possibilités pour l'architecture IoTFLA en termes d'améliorations et de perspectives d'évolutions. Notre architecture et ses composants ont été sélectionnés et choisis en se basant sur nos préférences, critères (protection de la vie privée, sécurité, automatisation, etc.) , paramètres et contraintes. Il est possible de créer une architecture qui est similaire à celle proposée tout en choisissant différentes solutions pour chaque composant (automatisation, SDI, serveur d'infonuagique, base de données, objets, téléphone intelligent, routeur, etc.) ou encore apporter des composants et solutions supplémentaires.

Nous voulons également souligner le fait que l'architecture IoTFLA peut être soumise à des changements ou évolutions en ajoutant des composants ou en sélectionnant des solutions plus appropriées ou de nouvelles qui apparaîtront dans le futur si elles sont plus efficaces et disponibles dans le contexte de l'IdO et des maisons intelligentes.

Afin d'illustrer les perspectives d'évolutions de notre architecture, voici quelques exemples. Une possibilité par exemple, pourrait être de remplacer notre système de détection d'intrusions par une solution telle qu'IoT Sentinel (Miettinen *et al.*, 2017) ou encore intégré IoT Sentinel en tant que part entière de notre système de détection d'intrusions. De plus, des techniques telles que la modification, manipulation et injection de trafic ou encore l'inférence de comportements (Miettinen *et al.*, 2017) sur des réseaux tels que ceux des maisons intelligentes peuvent être des ajouts potentiels pour contrecarrer certain type de dangers.

BIBLIOGRAPHIE

(2016). Internet of things.
https://fr.idate.org/internet-of-things-news2016/.

(2016). Large-scale open testbed jose.
https://www.nict.go.jp/en/nrh/nwgn/jose.html.

(2016). Regulation (eu) 2016/679 of the european parliament and of the council.
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

(2017). Federated learning : Collaborative machine learning without centralized training data. https://ai.googleblog.com/2017/04/federated-learning-collaborative.html.

(2017). Gartner says 8.4 billion connected "things" will be in use in 2017, up 31 percent from 2016.
https://www.gartner.com/en/newsroom/press-releases/.

(2017). Internet of things.
https://www.forbes.com/sites/louiscolumbus/2017/12/10/2017-roundup-of-internet-of-things-forecasts.

(2018). Suricata. https://suricata-ids.org/.

(2019). Home assistant. https://www.home-assistant.io/.

(2019). Ossec. https://www.ossec.net/.

(2019). Security onion. https://securityonion.net/.

Alzaid, H., Foo, E. et Nieto, J. G. (2008). Rsda : Reputation-based secure data aggregation in wireless sensor networks. *9th International Conference on Parallel and Distributed Computing, Applications and Technologies*, 419–424.

Atzori, L., Iera, A. et Morabito, G. (2010). The internet of things : A survey. *Computer networks*, 2787–2805.

Bagaa, M., Lasla, N., Ouadjaout, A. et Challal, Y. (2007). Sedan : Secure and efficient protocol for data aggregation in wireless sensor networks. *32nd IEEE Conference on Local Computer Networks*, 1053–1060.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A. et Seth, K. (2016). Practical secure aggregation for federated learning on user-held data. *CoRR*.

Butun, I., Morgera, S. D. et Sankar, R. (2014). A survey of intrusion detection systems in wireless sensor networks. *IEEE communications surveys & tutorials*, 266–282.

Çam, H., Özdemir, S., Nair, P., Muthuavinashiappan, D. et Sanli, H. O. (2006). Energy-efficient secure pattern based data aggregation for wireless sensor networks. *Computer Communications*, 446–455.

Chen, C.-M., Lin, Y.-H., Lin, Y.-C. et Sun, H.-M. (2012). Rcda : Recoverable concealed data aggregation for data integrity in wireless sensor networks. *IEEE Transactions on parallel and distributed systems*, 727–734.

Chen, D., Chang, G., Jin, L., Ren, X., Li, J. et Li, F. (2011). A novel secure architecture for the internet of things. *5th International Conference on Genetic and Evolutionary Computing*, 311–314.

Chernyshev, M., Baig, Z., Bello, O. et Zeadally, S. (2018). Internet of things (iot) : research, simulators, and testbeds. *IEEE Internet of Things Journal*, 1637–1647.

Conti, M., Dehghantanha, A., Franke, K. et Watson, S. (2018). Internet of things security and forensics : Challenges and opportunities.

Des Roziers, C. B., Chelius, G., Ducrocq, T., Fleury, E., Fraboulet, A., Gallais, A., Mitton, N., Noël, T. et Vandaele, J. (2011). Using senslab as a first class scientific tool for large scale wireless sensor network experiments. *International Conference on Research in Networking*, 147–159.

Dwork, C. (2011). Differential privacy. *Encyclopedia of Cryptography and Security*, 338–340.

Eswari, T. et Vanitha, V. (2013). A novel rule based intrusion detection framework for wireless sensor networks. *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 1019–1022.

Farooq, M., Waseem, M., Khairi, A. et Mazhar, S. (2015). A critical analysis on the security concerns of internet of things. *International Journal of*

Computer Applications.

Gentleman, R. et Carey, V. (2008). Unsupervised machine learning. *Bioconductor Case Studies*, 137–157.

Geyer, R. C., Klein, T. et Nabi, M. (2017). Differentially private federated learning : A client level perspective. *CoRR*.

Girao, J., Westhoff, D. et Schneider, M. Cda : Concealed data aggregation in wireless sensor networks. *ACM Workshop on Wireless Security*, 3004–3049.

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

Gupta, H., Vahid Dastjerdi, A., Ghosh, S. K. et Buyya, R. (2017). ifogsim : A toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments. *Software : Practice and Experience*, 1275–1296.

Han, S. N., Lee, G. M., Crespi, N., Heo, K., Van Luong, N., Brut, M. et Gatellier, P. (2014). Dpwsim : A simulation toolkit for iot applications using devices profile for web services. *2014 IEEE World Forum on Internet of Things (WF-IoT)*, 544–547.

Hastie, T., Tibshirani, R. et Friedman, J. (2009). Unsupervised learning. *The elements of statistical learning*, 485–585.

Henderson, T. R., Lacage, M., Riley, G. F., Dowell, C. et Kopena, J. (2008). Network simulations with the ns-3 simulator. *SIGCOMM demonstration*, p. 527.

Hu, L. et Evans, D. (2003). Secure aggregation for wireless networks. *Symposium on Applications and the Internet Workshops*, 384–391.

Ibrahim, J. M., Karami, A. et Jafari, F. (2017). A secure smart home using internet-of-things. *9th International Conference on Information Management and Engineering*, 69–74.

Jing, Q., Vasilakos, A. V., Wan, J., Lu, J. et Qiu, D. (2014). Security of the internet of things : perspectives and challenges. *Wireless Networks*, 2481–2501.

Jose, J., Princy, M. et Jose, J. (2013). Peppda : Power efficient privacy preserving data aggregation for wireless sensor networks. *International Conference on Emerging Trends in Computing, Communication and*

Nanotechnology, 330–336.

Jun, C. et Chi, C. (2014). Design of complex event-processing ids in internet of things. *2014 Sixth International Conference on Measuring Technology and Mechatronics Automation*, 226–229.

Kevin, A. (2009). That 'internet of things' thing, in the real world things matter more than ideas. *RFiD Journal*, 97–114.

Khan, R., Khan, S. U., Zaheer, R. et Khan, S. (2012). Future internet : the internet of things architecture, possible applications and key challenges. *10th International Conference on Frontiers of Information Technology*, 257–260.

Kolias, C., Kambourakis, G., Stavrou, A. et Voas, J. (2017). Ddos in the iot : Mirai and other botnets. *Computer*, 80–84.

Konečnỳ, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T. et Bacon, D. (2016). Federated learning : Strategies for improving communication efficiency. *CoRR*.

Kotsiantis, S. B., Zaharakis, I. et Pintelas, P. (2007). Supervised machine learning : A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 3–24.

Lathamanju, R. et Senthilkumar, P. (2013). Crsr algorithm : A secure data aggregation algorithm in wsn. *International Journal of Advanced Research in Electronics and Communication Engineering*.

Li, H., Lin, K. et Li, K. (2011). Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks. *Computer Communications*, 591–597.

Madden, S., Franklin, M. J., Hellerstein, J. M. et Hong, W. (2002). Tag : A tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review*, 131–146.

Mahimkar, A. et Rappaport, T. S. (2004). Securedav : A secure data aggregation and verification protocol for sensor networks. *IEEE Global Telecommunications Conference*, 2175–2179.

Mathov, Y., Agmon, N., Shabtai, A., Puzis, R., Tippenhauer, N. O. et Elovici, Y. Challenges for security assessment of enterprises in the iot era. *CoRR*, 1906–1922.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S. *et al.* Communication-efficient learning of deep networks from decentralized data. *20th International Conference on Artificial Intelligence and Statistics*,

1273–1282.

Mehdi, K., Lounis, M., Bounceur, A. et Kechadi, T. (2014). Cupcarbon : A multi-agent and discrete event wireless sensor network design and simulation tool. *7th International ICST Conference on Simulation Tools and Techniques, Lisbon, Portugal, 17-19 March 2014*, 126–131.

Merkle, R. C. (1980). Protocols for public key cryptosystems. *IEEE Symposium on Security and Privacy*, 122–122.

Miettinen, M., Marchal, S., Hafeez, I., Asokan, N., Sadeghi, A.-R. et Tarkoma, S. (2017). Iot sentinel : Automated device-type identification for security enforcement in iot. *37th International Conference on Distributed Computing Systems (ICDCS)*, 2177–2184.

Mohri, M., Rostamizadeh, A. et Talwalkar, A. (2012). *Foundations of machine learning.*

Nadeem, A. et Javed, M. Y. (2005). A performance comparison of data encryption algorithms. *International Conference on Information and Communication Technologies*, 84–89.

Neshenko, N., Bou-Harb, E., Crichigno, J., Kaddoum, G. et Ghani, N. (2019). Demystifying iot security : an exhaustive survey on iot vulnerabilities and a first empirical look on internet-scale iot exploitations. *IEEE Communications Surveys & Tutorials*, 2702–2733.

Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N. et Sadeghi, A.-R. (2019). Dïot : A federated self-learning anomaly detection system for iot. *IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 756–767.

Osterlind, F., Dunkels, A., Eriksson, J., Finne, N. et Voigt, T. (2006). Cross-level sensor network simulation with cooja. *31st IEEE conference on Local computer networks*, 641–648.

Ozdemir, S. (2007). Secure and reliable data aggregation for wireless sensor networks. *International Symposium on Ubiquitious Computing Systems*, 102–109.

Ozdemir, S. et Xiao, Y. (2011). Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. *Computer Networks*, 1735–1746.

Padmavathi, D. G. et Shanmugapriya, M. (2009). A survey of attacks, security mechanisms and challenges in wireless sensor networks. *CoRR.*

Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I. et Talwar, K. (2016). Semi-supervised knowledge transfer for deep learning from private training data. *5th International Conference on Learning Representations.*

Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K. et Erlingsson, Ú. (2018). Scalable private learning with pate. *6th International Conference on Learning Representations.*

Perrig, A., Szewczyk, R., Tygar, J. D., Wen, V. et Culler, D. E. (2002). Spins : Security protocols for sensor networks. *Wireless networks*, *8*(5), 521–534.

Poornima, A. et Amberker, B. (2010). Seeda : Secure end-to-end data aggregation in wireless sensor networks. *7th International Conference on Wireless And Optical Communications Networks*, 1–5.

Przydatek, B., Song, D. et Perrig, A. (2003). Sia : Secure information aggregation in sensor networks. *1st International Conference on Embedded Networked Sensor Systems*, 255–265.

Roesch, M. *et al.* (1999). Snort : Lightweight intrusion detection for networks. *Lisa*, 229–238.

Sanchez, L., Muñoz, L., Galache, J. A., Sotres, P., Santana, J. R., Gutierrez, V., Ramdhany, R., Gluhak, A., Krco, S., Theodoridis, E. *et al.* (2014). Smartsantander : Iot experimentation over a smart city testbed. *Computer Networks*, 217–238.

Sanli, H. O., Ozdemir, S. et Cam, H. (2004). Srda : secure reference-based data aggregation protocol for wireless sensor networks. *IEEE 60th Vehicular Technology Conference*, 4650–4654.

Schneble, W. H. A. (2018). *Federated Learning for Intrusion Detection Systems in Medical Cyber-Physical Systems.* (Thèse de doctorat). University of Washington Libraries.

Simpson, A. K., Roesner, F. et Kohno, T. (2017). Securing vulnerable home iot devices with an in-hub security manager. *IEEE International Conference on Pervasive Computing and Communications Workshops*, 551–556.

Soni, A. et Randhawa, R. (2017). Secure data aggregation protocols in wireless sensor networks. *International Journal of Engineering and Technology*, *9*, 3790–3797.

Sotiriadis, S., Bessis, N., Asimakopoulou, E. et Mustafee, N. (2014). Towards simulating the internet of things. *28th International Conference on Advanced*

Information Networking and Applications Workshops, 444–448.

Syverson, P., Dingledine, R. et Mathewson, N. (2004). Tor : The secondgeneration onion router. *Usenix Security*.

Varga, A. et Hornig, R. (2008). An overview of the omnet++ simulation environment. *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*, p. 60.

Wang, T., Qin, X. et Liu, L. (2013). An energy-efficient and scalable secure data aggregation for wireless sensor networks. *International Journal of Distributed Sensor Networks*, p. 843485.

Wu, M., Lu, T.-J., Ling, F.-Y., Sun, J. et Du, H.-Y. (2010). Research on the architecture of internet of things. *3rd International Conference on Advanced Computer Theory and Engineering*, p. 484.

Yang, X., Li, Z., Geng, Z. et Zhang, H. (2012). : 388–393.

Yang, Y., Wang, X., Zhu, S. et Cao, G. (2008). Sdap : A secure hop-by-hop data aggregation protocol for sensor networks. *ACM Transactions on Information and System Security*, p. 18.

Zarpelao, B. B., Miani, R. S., Kawakani, C. T. et de Alvarenga, S. C. (2017). A survey of intrusion detection in internet of things. *Journal of Network and Computer Applications*, 25–37.

Zeng, X., Garg, S. K., Strazdins, P., Jayaraman, P. P., Georgakopoulos, D. et Ranjan, R. (2017). Iotsim : A simulator for analysing iot applications. *Journal of Systems Architecture*, 93–107.

Zhu, L., Yang, Z., Xue, J. et Guo, C. (2014). An efficient confidentiality and integrity preserving aggregation protocol in wireless sensor networks. *International Journal of Distributed Sensor Networks*, p. 565480.

Ziegeldorf, J. H., Morchon, O. G. et Wehrle, K. (2014). Privacy in the internet of things : threats and challenges. *Security and Communication Networks*, 2728–2742.