RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Saad Dahlab blida 1

Faculté Des Sciences

Département d'Informatique

Mémoire de fin d'étude

Pour l'obtention du diplôme de Master en Informatique

OPTION : SECURITE DES SYSTEMES D'INFORMATION

THEME

Mise en oeuvre d'une solution de préservation de la vie privée des usagers mobiles dans une ville intelligente

Réalisé par :
- **LARBI BOUARMRANE Mohamed**
- **MERZOUK Fares Mounir**

Devant le jury composé de :
- **AROUSSI Sana Président**
- **BOUSTIA Narhimene Promotrice**
- **BERRAMDANE Djamila Examinatrice**

Encadré par :
- **BENNA Amel**
- **MEZIANE Abdelkarim**

Organisme d'accueil :
- **CERIST**

10 septembre 2020

Remerciements

Nous souhaiterions remercier toutes les personnes qui ont participé à la création de ce projet, nous ont aidé et supervisé pour sa réalisation et sa conclusion.

Tout d'abord nous souhaiterions remercier *notre promotrice*, madame l'attaché de recherche **Amel BENNA**, de nous avoir supervisé tout au long de ce projet et de nous avoir apporté ses remarques et conseils pour l'aboutissement de ce projet. Nous vous remercions pour la patience et l'attention que vous nous avez accordé au cours de ce projet.

Nous souhaiterions aussi remercier *notre co-promoteur*, le maître de recherche **Abdelkrim MEZIANE**, pour nous avoir accueilli au sein du CERIST et de nous avoir proposé ce sujet de master. Nous souhaiterions également le remercier pour l'aide qu'il nous a apporté au cours du projet.

Nous souhaiterions, enfin, remercier *notre co-promotrice* de l'université de Blida, le Docteur **Narhimene BOUSTIA**, d'avoir accepté de nous encadré pour ce projet. Nous la remercions également pour toutes les connaissances qu'elle nous a enseigné au cours de nos deux années de master.

Dédicaces

Je dédie ce travail :

A mes parents, **Mustapha et Nachida**, qui ont su me conseiller et me soutenir tout au long de mon cursus scolaire et à qui je dois tout. les mots ne suffiront jamais pour décrire tous les efforts que vous avez fait pour faire de moi l'homme que je suis aujourd'hui.

A mon très cher frère, **Anis**, avec qui j'ai partagé des moments remplis de joies et de bonne humeur pendant plusieurs années et sur qui j'ai pu compter dans les moments de difficulté. Je te souhaite une vie pleine de réussite et de réaliser tous tes rêves. Tu resteras, pour toujours, mon meilleur ami et mon sang.

A mon grand frère et sa compagne, **Adel et Dalila**, qui, malgré la distance, ont su m'apporter le soutient et les conseils qui m'ont guidé durant mes choix universitaires. Vous resterez toujours présents dans mon coeur et j'espère que l'avenir me guidera dans vos pas afin que je puisse réaliser tout ce que vous avez pu réaliser jusque là.

A mon meilleur ami, **Zaki**, avec qui j'ai partagé tellement de souvenirs joyeux et qui a su me montrer son soutient moral et une épaule sur laquelle me reposer dans mes moments de détresse. Je ne te remercierais jamais assez d'être la personne que tu es.

A ma très chère amie, **Amina**, qui a su m'apporter l'ambiance et la bonne humeur au moment où j'en avais le plus besoin. Je te souhaite une vie pleine de succès et de bonnes surprises et j'espère que la vie fera en sorte que nos deux *destins* se recroisent de nouveau.

A ma meilleure amie d'enfance, **Amira**, avec qui j'ai partagé une jeunesse formidable pleine de fou rire et de bons moments. Tu seras toujours présente dans mon coeur malgré la distance qui nous sépare. Je te souhaite de vivre ta vie au maximum et de garder le sourire.

A mes chers amis, **Nabil, Chafik, Alaa, Tarek** et bien-sûr **Emmanuel**, avec qui j'ai partagé des souvenirs inoubliables dans l'espoir d'en partager encore d'autres dans le futur. Je vous souhaite de réaliser vos rêves et d'accomplir vos objectifs malgré les obstacles.

Et enfin, *à mon cher binôme et bras droit*, **Adam**, sans qui ce travail n'aurait jamais vu le jour. Je n'aurais su trouver mieux comme partenaire de travail. Tu as su me rassurer dans les moments de stresses et de doutes. Je te souhaite un avenir professionnel radieux.

Fares

Je dédie ce travail :

A mes parents, **Hocine et Nacera**, pour avoir fait de moi l'homme que je suis aujourd'hui. Je vous remercie de tout mon coeur pour tous les sacrifices que vous avez fait pour moi depuis ma naissance. On dit qu'on ne choisit pas ses parents, mais si j'avais eu le choix, je vous aurais quand même choisi.

A mon jeune frêre, **Mouadh** et *mes très chères soeurs* **Maroua, Radjaaa, Meriem, Hala**, pour m'apporter chaque jour l'ambiance et les fou-rires dans ma vie. Je vous souhaite d'avoir une vie pleine de succès et de combler vos parents de bonheur en les rendant fiers de vous.

*A toute **ma famille** maternelle et paternelle*. Je vous remercie tous autant que vous êtes. Je suis fier de partager mon nom avec une famille aussi forte et aussi soudée comme les doigts de la main.

*A tous **mes amis**,* sans cité les noms pour n'en oublier aucun, pour toutes les sorties, les soirées, les nuits et les journées passées à vos cotés. Au plaisir de partager encore plus de souvenirs à notre prochaine rencontre.

A ma deuxième famille, **les SCOUTS de Hadjout**, plus particulièrement, *mes collègues ainsi que mes disciples*. Je tiens spécialement à remercier mes mentors de m'avoir formé dés mon plus jeune âge et de m'avoir aidé sur mon développement personnel. Je vous remercie de m'avoir accorder une si grande hospitalité et de m'avoir tant de connaissances au fil des années.

A mon club d'informatique de l'université de Blida, **le CSCC**, pour m'avoir appris beaucoup de choses dans le domaine de l'informatique, particulièrement, dans le domaine de sécurité qui, plus tard, est devenu mon choix de spécialité grâce à eux.

A toute la famille **MERZOUK** pour m'avoir considéré comme votre propre fils, pour nous avoir permis de travailler dans les meilleures conditions de travail, plus particulièrement, *mon cher ami*, avant d'être mon binôme, **Fares** avec qui j'ai travaillé dur au cours de ce projet. Merci pour ton soutient moral, ta patience et ta compréhension tout au long de ce voyage. Je te souhaite beaucoup de bonheur et de succès sur ton chemin qui, j'espère, sera long et joyeux.

Adam

Résumé

Dans une ville intelligente, tout est inter-connecté, des hôpitaux jusqu'aux restaurants en passant par les bâtiments administratifs et les moyens de transport. Cette interconnectivité digitale doit passer par un partage de données des différents secteurs afin de pouvoir faire profiter le citoyen de la ville intelligente de tous les services qu'il demande de la plus simple des façons. Néanmoins, une moindre faille dans ce partage de données pourrait exposer les utilisateurs de ces services à des personnes malveillantes. Par ailleurs, tout type de données des citoyens peuvent être stockés dans divers appareils électroniques pour généraliser le concept d'intelligence. Le smart-phone est largement le plus utilisé pour ce genre de tâches. Toutefois, les smart-phones ne sont pas assez performants pour gérer les données sensibles des utilisateurs et se retrouve généralement confrontés au problème de sur-collecte de données. Grâce à cette généralisation du smart-phone, un second problème fait surface concernant les applications Location-Based Services (LBS). Malgré leur potentiel énorme, les risques de sécurité pour ces applications LBS pour la vie privée des usagers mobiles est considérable et doit absolument être traité au risque de freiner leur exploitation. Afin de remédier à tous les problèmes de sécurité cités précédemment, ce rapport propose quelques solutions de sécurisation, entre autres, le stockage de données dans un cloud-mobile pour la résolution du problème de sur-collecte de données, un système de localisation par zone pour la sécurisation du système de localisation des applications LBS, et enfin, l'algorithme K-anonymity pour l'anonymisation des données sensibles des citoyens de la ville intelligente au cours du partage de données.

Mots clés

Ville intelligente, smart-phone, sur-collecte, localisation, sécurisation, cloud-mobile, partage de données, anonymisation.

Abstract

In a smart city, everything is interconnected, hospitals, restaurants, administrative buildings and means of transport. This digital interconnectivity involve sharing data from different sectors in order to provide smart city citizens all the required services. However, a smaller flaw in this sharing could expose users to malicious individuals. Moreover, all kind of citizen data are stored in various electronic devices to generalize the concept of intelligence. The smartphone is widely used for this kind of task. However, smartphones are not efficient enough to handle sensitive user data and usually ends up in data over-collection issues. Thanks to this generalization of the smartphone, a second problem arises, the Location-Based Services (LBS) applications. Despite their enormous potential, the security risks to the mobile users's privacy are considerable and must absolutely be solved at the risk of hampering their growth. In order to remedy all the security problems mentioned above, this report proposes some security solutions, among others, the data storage in a cloud-mobile for the problem of data over-collection, a localization system by zone to secure the location system of LBS applications, and finally, the K-anonymity algorithm for the anonymization of sensitive data of the smart city's citizens during data sharing.

Keywords : smart city, smartphones, over-collection, privacy, localization, security, cloud-mobile, anonymization.

ملخص

في المدينة الذكيّة ، كل شيء مترابط ، من مستشفيات إلى مطاعم ومباني إداريّة وحتى وسائل النقل. لذلك وجب على هذا الترابط الرقميّ أن يتضمّن مشاركة البيانات من مختلف القطاعات الخدماتية فيما بينها حتى تتمكن من تزويد مواطني المدن الذكية بجميع الخدمات الّتي يحتاجونها بأبسط الطرق. وعليه ، قد يؤدي وجود خلل صغير في مشاركة هاته البيانات إلى تعريض مستخدمي هذه الخدمات لأشخاص ضارين. و يمكن أيضا تخزين أي نوع من بيانات المواطن في مختلف الأجهزة الإلكترونية لتعميم مفهوم الذكاء. يعتبر الهاتف الذكي الجهاز الأكثر استعمالا لهذا النوع من المهام. ومع ذلك ، فإن الهواتف الذكيّة ليست متطوّرة بما يكفي للتعامل مع بيانات المستخدم الحسّاسة وهي أيضا تواجه مشكلة الإفراط في جمع البيانات عموما. وهذا ما نتج عنه مشكلة ثانية تتعلق بتطبيقات الخدمات القائمة على الموقع (LBS). على الرغم من إمكانيّاتها الهائلة ، إلا أن المخاطر الأمنيّة التي تمس خصوصيّة مستخدمي الهاتف الذكيّ بسبب هذه التطبيقات تعتبر كبيرة ويجب معالجتها مع المخاطرة بإعاقة عملها. من أجل معالجة جميع المشاكل الأمنيّة التي سبق ذكرها، يقترح هذا العمل بعض الحلول الأمنيّة ، من بين تلك الحلول تخزين البيانات في سحابة مخصصة للهواتف النقالة لحلّ مشكلة الإفراط في جمع البيانات ، ونظام تحديد المكان حسب المنطقة لتأمين نظام تعريف المواقع في تطبيقات LBS ، وأخيراً ، خوارزمية K-anonymity لإخفاء الهوية وحماية البيانات الحساسة لمواطني المدينة الذكية أثناء مشاركة البيانات.

الكلمات الرئيسية : المدينة الذكية ، الهواتف الذكية ، الإفراط في جمع البيانات ، تحديد المكان ، الحماية ، مشاركة البيانات ، سحابة مخصصة للهواتف النقالة ، إخفاء الهوية.

Table des matières

Table des figures

Liste des tableaux

Liste des acronymes

Acronymes	Description
CERIST	Centre de Recherche sur l'information Scientifique et Technique
CO2	Dioxyde de carbone
CP-ABE	Ciphertext-Policy Attribute-Based Encryption
EPC	Electronic Protocol Code
GPS	Global Positioning System
GSM	Global System for Mobile Communications
IBM	International Business Machines
ID	Identifiant
IDE	Environnement de développement integré
IEEE	Institute of Electrical and Electronics Engineers
IMEI	International Mobile Equipment Identity
iOS	Intelligent Operating System
IoT	Internet of Things
JSON	JavaScript Object Notation
LBS	Location-Based Services
M2M	Machine to Machine
PHC	partenariat Hubert Curien
PP-CP-ABE	Privacy Preserving Ciphertext-Policy Attribute-Based Encryption
QR	Quick Response
RFID	Radio Frequency Identification
SL	Security Level
SNS	Social Network Service
SQL	Structured Query Language
TIC	Technologies de l'information et de la communication
UDID	Unique Device Identifier
WiFi	Wireless Fidelity
WSN	Wireless sensor network (réseau de capteurs sans fil)

TABLE 1 – Liste des acronymes

Introduction générale

Durant les dernières décénies, le développement de la technologie de l'information et de la communication a incité les plus grandes entreprises telles que IBM, Cisco ou encore Microsoft à dépasser leur limites dans le développement de villes capables de stocker des données en masse afin de les exploiter pour faciliter la vie de chacun. Une nouvelle génération de l'urbanisation est né, un lieu appelé "ville intelligente", où les gens sont en totale immersion avec des systèmes intelligents qui leur proposent divers services en toute simplicité. Cependant, pour bien exploiter les fonctionnalités offertes par ces systèmes intelligents, les résidents doivent souvent offrir leurs informations personnelles tels que : leur numéro de compte bancaire et leur mots de passe pour faire des achats en ligne, leur adresse pour recevoir des colis, etc. Par ailleurs, les données sont au cœur d'une ville intelligente, car elles sont constituées de toutes les informations des utilisateurs, ce qui est inestimable à l'ère du Big-Data. Bien évidemment, cela représente des risques de confidentialité pour l'utilisateur. La sécurité représente un atout primordiale dans le développement d'une ville intelligente. Dans la mesure où la totalité du système est informatisé, le risque de fuite peut aussi bien être interne qu'externe.

En cette nouvelle génération d'urbanisation, le smart phone est devenu le dispositif électronique le plus utilisé par les gens dû au fait que tous les systèmes traditionnels connus ont intégré leur services dans les téléphones portables à cause de sa portabilité et son accessibilité. Les smartphones jouent un rôle irremplaçable dans une ville intelligente. Il semble pratique que nous stockions des données dans des smartphones et que nous les utilisions partout et à tout moment. Selon une étude réalisé par Yibin et al [1], les problèmes de sécurité sont moins fréquents dans les systèmes d'exploitation d'Apple appelé iOS (Intelligent Operating System) à cause d'une sécurité accrue que dans les systèmes non iOS, par exemple, Android. Dû au fait que les smart-phones stockent tout type de données en grande quantité, cela représente un risque majeur pour l'utilisateur d'être exposé à l'exploitation de ses données personnelles par des personnes malveillantes. La sur-collecte de données des applications, l'exposition de la vie privée dans les réseaux sociaux ou encore le partage d'informations personnelles sont des problèmes qui pourraient facilement compromettre la sécurité de l'individu sans qu'il en soit conscient.

L'objectif de notre projet de fin d'études est de proposer des solutions permettant la sécurisation des données personnelles des usagers mobiles avec l'utilisation de techniques de stockages et d'algorithmes d'anonymisation des données sensibles et ainsi garantir la confidentialité et l'authenticité des informations des citoyens. Ce projet de fin d'étude rentre dans le cadre d'un projet PHC-Tassili sur la recommandation de services dans les villes intelligentes entre l'université D'Angers et le CERIST.

Structure du rapport

Après cette introduction, ce rapport se compose en deux parties. La première partie est réservée à l'état de l'art sur les villes intelligentes et les solutions proposées à certains problèmes de confidentialité et de sécurité des données des villes intelligentes. Cette partie comprend 2 chapitres : un premier chapitre qui décrit la ville intelligente et ses problèmes de sécurité et de confidentialité. Par la suite, un deuxième chapitre qui introduira un des problèmes de confidentialité, en l'occurrence, la sur-collecte de données et son impact sur les villes intelligentes, et enfin, les solutions proposées pour remédier à cette sur-collecte des données. Pour la deuxième partie du rapport, il y aura un chapitre d'analyse et conception qui mettra en avant les solutions choisies. Ensuite, on passera à la réalisation et l'implémentation de la solution et aux tests des résultats obtenues. Enfin, nous allons conclure le rapport avec une conclusion générale qui résumera tout ce qui a été fait tout au long du projet.

Chapitre 1

Villes intelligentes et problèmes de sécurité

1.1 Introduction

Les villes intelligentes peuvent être considérées comme un grand système d'information qui consiste en plus petits sous-systèmes connectés à d'autres systèmes. Le défi de développer une ville intelligente devient de plus en plus complexe avec la nécessité de développement de structures de sécurité. Le Smart phone est le dispositif électronique le plus souvent utilisé pour la collecte de données de l'utilisateur et l'interaction avec les systèmes inter-connectés de la ville intelligente. Cette inter-connectivité des systèmes pourrait causer du tord aux citoyens de la ville intelligente si, par malheur, leurs données privées venaient à être exposé aux mauvaises personnes.

Au cours de ce chapitre, nous allons voir une définition de la ville intelligente et citer les différents secteurs où le système intelligent a été intégré. Nous allons, par la suite, décrire l'architecture de la ville intelligente en citant ses caractéristiques les plus importantes. Enfin, nous citerons quelques problèmes de sécurité auxquels la ville intelligente fait face.

1.2 L'essor des villes intelligentes

Au cours des deux dernières décennies, le secteur urbain a beaucoup évolué pour conduire au développement de villes intelligentes. Cet aspect est vite devenu le centre d'intérêt des grandes compagnies pour la simplification du secteur social et éducatif. Plus de la moitié de la population sur terre vit en ville. L'utilisation de systèmes intelligents conduit à encore plus de villes intelligentes. La figure 1.1 illustre un exemple de ville intelligente issue du projet "**Go Green in the City**" de Schneider Electric [1].

FIGURE 1.1 – Achitecture "Go Green in the City" [1]

L'objectif du projet "**Go Green in the City**" [1] est d'améliorer et d'intégrer les systèmes traditionnels dans une ville intelligente. Une ville intelligente se dispose de différents systèmes intelligentes dans les différents secteurs tels que l'énergie, l'eau, les bâtiments, l'intégration, le service publique et la mobilité. D'aprés l'étude de Khatoune et al, l'article [2] cite quelques secteurs importants tels que :

Secteur gouvernemental : L'E-gouvernance est la proposition du travail gouvernementale par le biais d'un support électronique. Elle permet aux citoyens de remplir leurs responsabilités civiques et sociales à travers un portail web. L'objectif principal des services proposés par le gouvernement électronique inclut : Garantir l'accès à tous les services d'information à travers un site web gouvernemental ; assurer une coordination efficace entre tous les départements gouvernementaux ; fournir des méthodes de communication flexibles aux citoyens en utilisant divers types d'applications Internet comme par e-mail, SMS, chat ou réseaux sociaux ; fournir les informations sur les sentiments des citoyens et les alerter en cas de problème ; et enfin, éliminer la paperasserie gouvernementale. Le citoyen de la ville intelligente sera apte d'accéder à partir de chez lui à tous les services gouvernementaux proposés sans avoir à se déplacer. Par exemple, il pourra faire la demande d'un centre de conférence, le paiement de factures et le signalement de problèmes.

Secteur de la santé : Les services de soins dans une ville intelligente peuvent également bénéficié des dispositifs intelligents. Leur but est d'aider les gens à vivre sainement en garantissant l'accès à une gamme d'installation. Dans une ville intelligente, les professionnels du secteur médical auront souvent besoin d'accéder au dossier médical de leur patient à tout moment et n'importe où à l'aide de dispositifs connectés, spécialement dans le cas où la présence physique du professionnel n'est pas possible ou de problème de santé inattendu. Les services médicaux à distance peuvent être garantis grâce à des dispositifs connectés directement à des centres médicaux et à des systèmes d'analyse de données. Toutefois, il existe de nombreux défis auxquels ce secteur fait face ; la coordination entre les services médicaux prestataires, assurance en cas d'erreurs, consommation d'énergie élevée par les nouveaux systèmes de services médicaux, interactions entre les hôpitaux nécessitant des langages communs ou de nouveaux protocoles, des systèmes in-

teropérables de haute qualité et enfin la normalisation des processus. La santé représente la chose la plus importante au monde d'où l'importance du système de connectivité qui doit être présent dans les villes intelligentes.

Secteur des infrastructures essentielles : Le smart grid est un système informatisé construit sur des infrastructures avancées basées sur les TIC (Technologies d'information et de communication). Il gère l'électricité d'une manière durable, fiable et économique. Un réseau intelligent est un réseau utilisant des ordinateurs et des capteurs placés dans le réseau. Selon l'article de Khatoune et al [2], le royaume uni, l'île suédoise de Gotland ou encore les états unis travaillent afin d'améliorer les conditions de consommation d'énergie en exploitant le secteur des énergies renouvelables. Le projet américain CenterPoint Energy Houston Electric (CEHE) [2] montre effectivement l'efficacité du smart grid : amélioration de la fiabilité du système de distribution (évitant des dizaines de millions de minutes d'interruption des clients), les coûts de compteurs réduits, des coûts d'exploitation et d'entretien réduits (diminués d'environ 55 millions de dollars en 2013), une réduction de la consommation de carburant des camions et une réduction des coûts liés à la détection des vols tels que la remarque d'une consommation inhabituelle de la part du client (un coût réduit de 2 millions de dollars en 2013).

Secteur des bâtiments intelligents : Dans une ville intelligente, les bâtiments intelligents sont nécessaires pour diverses raisons : améliorer le confort, fonctionnement efficace des systèmes du bâtiment (c.-à-d. ascenseurs, conduits d'eau, conduits de gaz), et réduction de la consommation d'énergie. Dans un bâtiment intelligent, un système d'automatisation du bâtiment (BAS) contrôle automatiquement le chauffage, la climatisation, l'éclairage et d'autres systèmes. Quant à la promotion des technologies pour les bâtiments intelligents, il a été démontré que la technologie des piles à combustible permettra aux bâtiments intelligents de fournir leur propre électricité avec 50% d'émission de CO2 en moins.

Secteur du transport : Aujourd'hui, la moitié de l'humanité vit dans les villes. La mobilité dans les grandes villes entraîne plusieurs problèmes, tels que la congestion du trafic, l'augmentation de la pollution et la consommation d'énergie. Pour atténuer ces problèmes, les systèmes de transport intelligents (STI) offrent de multiples services, tels que la réduction de la mobilité en facilitant la sélection du mode de transport, l'optimisation de la planification des trajets, la détection des conducteurs présentant des comportements malveillants, l'amélioration de la sécurité des conducteurs et des passagers, la réduction d'émission de CO2, la mise à disposition des informations sur les places de stationnement sur les Smartphones et le suivi des véhicules. Dans ce secteur, la communication véhiculaire est une technologie clé dans les villes intelligentes.

1.2.1 Architecture IOT dans les villes intelligentes

Pour suivre le développement des villes intelligentes, plusieurs architectures ont été conçues. Cependant, il n'existe pas d'architecture IoT uniforme. Étant donné que l'accent de ce travail est de résumer les problèmes de sécurité et de confidentialité dans les villes intelligentes, l'architecture décrite ici est basée sur l'architecture à quatre couches bien connue et l'architecture généralement acceptée proposée par L.Tan et N.Wang dans la figure 1.2.

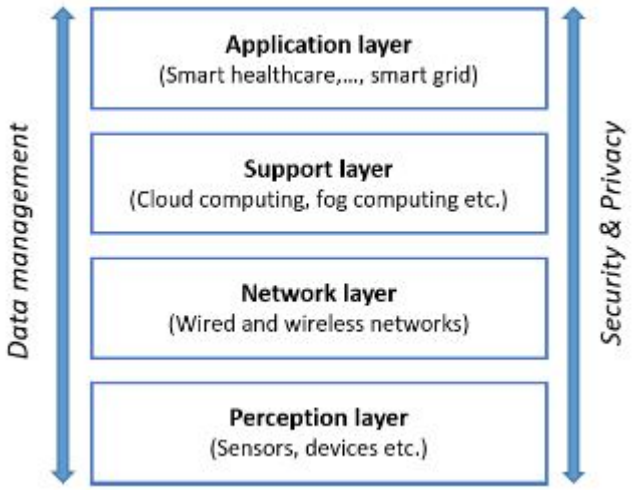

FIGURE 1.2 – Architecture basée IOT d'une ville intelligente. [3]

Comme le montre la figure 1.2 , l'architecture peut être divisée en quatre couches : Application, support, réseau et perception :

— **La couche de perception**, également appelée couche de détection, couche de reconnaissance ou couche de bord, est la couche la plus basse de l'architecture. La couche de perception est principalement utilisée pour la collecte de données à partir d'éléments (par exemple, des appareils hétérogènes, des WSN et des capteurs) dans le monde réel et pour transmettre les informations acquises à la couche réseau pour un traitement ultérieur.
— **La couche réseau** est la couche centrale de l'architecture IoT qui dépend des réseaux de base, tels qu'Internet, WSN et réseaux de communication. La responsabilité de cette couche est de transmettre les données collectées par la couche de perception et de connecter les objets intelligents, les périphériques réseau et les serveurs.
— **La couche support**, qui travaille en étroite collaboration avec la couche d'application, prend en charge les exigences des applications diversifiées via des techniques informatiques intelligentes (par exemple, le cloud computing, le edge computing, le brouillard).
— **La couche application**, en tant que couche supérieure, est chargée de fournir des services ou des applications intelligents et pratiques aux utilisateurs en fonction de leurs besoins personnalisés.

1.2.2 Caractéristiques des villes intelligentes

Cui et al [4] ont proposé cinq caractéristiques de ville intelligente à prendre en considération comme illustre la figure 1.3 :

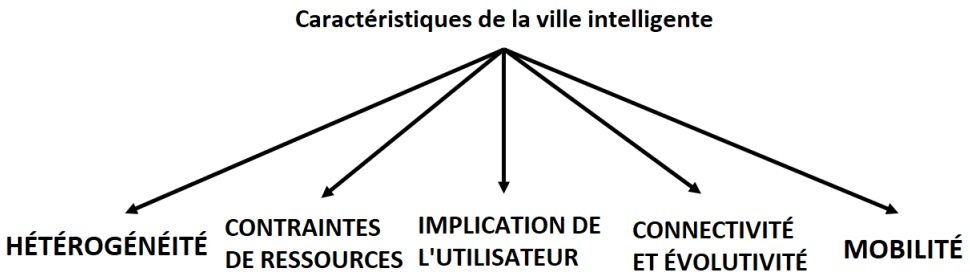

FIGURE 1.3 – Caractéristiques des villes intelligentes. [1]

— **Hétérogénéité :** Dans les systèmes basés IoT, il est important d'avoir une grande hétérogénéité, ce qui veut dire que le système doit être indépendant, distribué, stocké et utilisé par différents utilisateurs. Il fait également référence à la grande variété de nœuds IoT, de protocoles et de technologies de communication, de moyens de mobilité, de performances matérielles diverses, de platesformes, etc. L'architecture IoT varie en fonction de la ville intelligente. Par conséquent, l'absence d'un cadre et d'un service de sécurité communs est un autre problème majeur.

— **Contraintes de ressources :** La plupart des appareils IoT sont limités en ressources, ce qui signifie non seulement une mémoire, une capacité de batterie et des capacités de traitement limitées, mais également des interfaces réseau limitées en raison de normes radio de faible puissance. Pour être plus précis, les appareils embarqués moins chers, plus petits, mais moins énergivores sont largement utilisés dans les villes intelligentes. En règle générale, la mémoire à accès aléatoire et les capacités de stockage de ces périphériques sont limitées, avec des microcontrôleurs 8 bits ou 16 bits. Les réseaux sans fil équipés de la radio IEEE 802.15.4 entraînent des débits de données et des tailles de trame faibles (20-250 kb / s et jusqu'à 127 octets, respectivement) selon l'article [4].

— **Mobilité :** La mobilité urbaine a été considérée comme un moteur important de la croissance et du progrès des villes modernes. Dans les villes intelligentes, la mobilité ne se réfère pas seulement aux mouvements à l'intérieur d'une ville et à la livraison de marchandises d'un endroit à une autre destination, elle signifie également des technologies comme la communication sans fil à l'échelle de la ville et la surveillance en temps réel du trafic, ainsi que la flexibilité réactions aux problèmes. De plus, la mobilité dans les villes intelligentes est personnalisée grâce à une infrastructure de communication bien développée.

— **Connectivité et évolutivité :** La connectivité permet à tout appareil de se connecter au monde intelligent. C'est la caractéristique la plus fondamentale d'une ville intelligente réussie et elle a été considérée comme fondamentale pour faire avancer les plans de ville intelligente [5]. Dans le même temps, l'évolutivité est une caractéristique apparente dans les scénarios de ville intelligente. Les villes intelligentes se développent rapidement de petites à grandes, ce qui entraîne une croissance explosive du trafic de données et de réseaux. Par conséquent, une ville intelligente n'est pas en mesure de bien fonctionner sans systèmes et mécanismes évolutifs.

— **Implication de l'utilisateur :** La définition d'une ville intelligente ne concerne pas seulement les technologies et infrastructures de pointe, les facteurs humains (apprentissage, créativité et éducation) sont également essentiels au développement des villes intelligentes [6], car la construction de villes intelligentes a pour principal objectif de servir les résidents. En outre, la participation des citoyens peut améliorer la qualité de ces applications intelligentes. Par exemple, une compréhension initiale de leurs exigences et préoccupations concernant la sécurité se traduira par un meilleur résultat en termes de stratégies de protection.

1.3 Quelques problèmes de sécurité et de protection de vie privée de l'utilisateur dans les villes intelligentes

Pour introduire l'inter-connectivité, nous allons prendre l'exemple de John, alors qu'il est entrain de ranger son bureau pour quitter le travail, son patron lui demande de rester tard. Pour venir chercher son fils à l'école, John appelle une voiture autonome d'un glissement du pouce et le service envoie à son smart-phone le code Quick Response (QR) pour accéder à la voiture quelques instants plus tard. Son fils James sort de l'école et entre dans la zone de chargement de l'école de véhicules autonomes pour trouver une voiture autonome noire brillante qui l'attend avec la climatisation au réglage préféré de son profil. John continue à mettre la touche finale au projet supplémentaire, il enregistre à moitié consciemment le voyage de James via le système Global positioning System (GPS) de son smart-phone et voit qu'il quitte l'arrêt alimentaire qu'il a approuvé afin d'éviter une dispute avec sa femme. L'application de voiture autonome lui montre que la voiture de James a été ré-acheminée pour éviter les embouteillages et qu'il devrait arriver bientôt chez lui. Pendant ce temps les systèmes intelligents de la maison, en prévision de l'arrivée de James, modifient la température pour s'assurer que la climatisation n'est utilisée qu'en cas de besoin. Avec un bourdonnement sourd, John est informé que son fils est pris en charge, nourri et confortablement assis à la maison (fig 1.4).

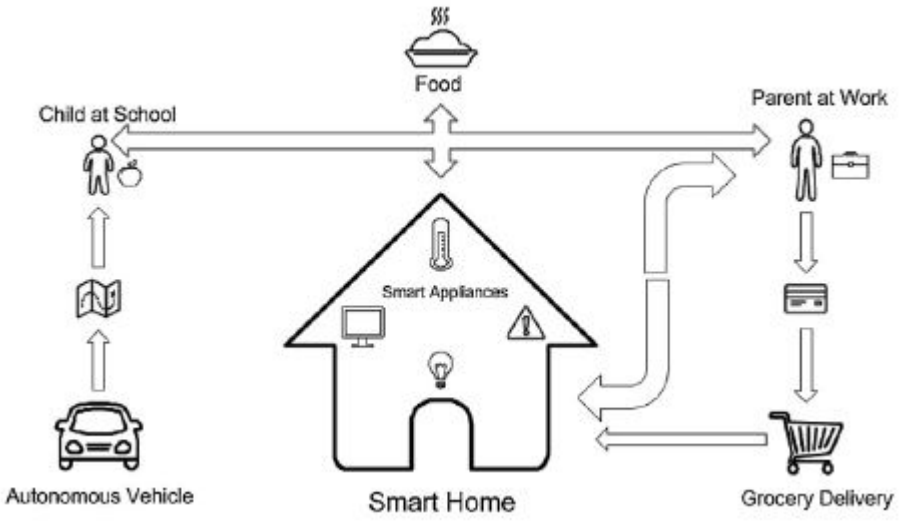

FIGURE 1.4 – Exemples d'interconnectivité dans une ville intelligente [7]

1.3.1 Menaces à la vie privée dans le partage et le datamining de données

Par nature de l'interconnectivité, les données sont transférés et utilisés tout au long du processus de ville intelligente, de multiples parties communiquantes et ayant accès aux informations. Des fabricants de capteurs intelligents jusqu'aux autorités de transport de la ville en passant par les individus accédant à la ville intelligente via leurs smartphones, chaque organisation contribuant à la ville intelligente utilisera et traitera les données de manière unique avec la possibilité de mettre en danger la vie privée.

Les attentes en matière de sécurité et de confidentialité seront centrées sur l'idée de T.Braun et al., qui est la suivante : La sécurité, "en tant que concept n'est pas absolue, mais c'est une tentative dynamique et authentique de prévenir les dommages à la ville intelligente et à ses habitants, à la fois directement et indirectement, par le biais de connexions numériques et physiques" [7].

Les villes intelligentes représentent plusieurs secteurs avec chacun des données qui corespondent à leur besoins, ces secteurs peuvent aranger des partenariats avec l'industrie privée concernant le même secteur pour améliorer la sécurité de la vie privée. Par exemple, les services de santé dans une ville intelligente peuvent s'appuyer sur des partenariats transparents entre les secteurs publics et privés. Alors qu'un hôpital public peut administrer les soins à vue et être un décideur central, la répartition des patients et des médicaments entre les établissements peut être réalisée plus efficacement grâce à un partenariat avec le secteur privé. Dans un tel scénario, les informations sensibles concernant l'état d'un patient, le calendrier de traitement et l'adresse du domicile devront être transférées et analysées par les parties concernées.

D'après Fung et al. [8], Ce processus de transmission de données entre plusieurs organisations dans un but commun aboutit souvent à des techniques appelées mashup de données et intégration de données. Le mashup de données fait référence à la simple jonction

de deux ou plusieurs ensembles de données avec un sujet d'intérêt commun, tandis que l'intégration de données joint plusieurs ensembles de données d'une manière qui modifie les données existantes. Ils expliquent trois problèmes de confidentialité qui surviennent avec les mashups de données privées de grande dimension entre les parties concernées. Premièrement, en combinant plusieurs datasets privés, le dataset résultant révélerait des informations plus sensibles aux autres fournisseurs de données. Deuxièmement, le dataset intégré pourrait faciliter l'identification des individus en fournissant plus de points de données pour la réidentification. Par exemple, le simple fait de savoir qu'un avocat de 30 ans souffre de leucémie n'aide pas vraiment un acteur malveillant à déterminer quel patient a le cancer, mais sachant qu'un avocat de 30 ans né le 2 avril à Burlington, au Vermont, est atteint de leucémie pourrait être retracée à un seul nom. Troisièmement, les données de mashup provenant de plusieurs sources peuvent contenir autant d'attributs de données que les modèles de confidentialité traditionnels, comme K-anonymity [9], rendraient les données protégées inutilisables pour l'analyse.

Les villes intelligentes ne peuvent pas s'appuyer sur des méthodes traditionnelles de protection de la vie privée lorsqu'elles réglementent ou participent au mashup de données. De plus, lors de l'anticipation des attaques contre la confidentialité, les hypothèses sur la connaissance d'un adversaire peuvent nuire à la sécurité de la confidentialité. Ainsi, lors de l'examen des méthodes de préservation de la vie privée individuelle dans les ensembles de données des villes intelligentes, le concept de confidentialité différentielle devrait être pris en compte [10]. La confidentialité différentielle a été considérée comme l'un des modèles de confidentialité les plus solides, car elle garantit empiriquement la confidentialité indépendamment des connaissances de base et de la puissance de calcul de l'attaquant [10].

1.3.2 Menaces sur la confidentialité dans le mashup de données

Étant donné qu'une ville intelligente est en fait un système bien connecté d'objets technologiquement intelligents, la sécurité de la ville intelligente est intrinsèquement plus difficile que la sécurisation d'objets intelligents individuels, tels que les smartphones, les objets IoT et les plates-formes de services. Toutes les vulnérabilités des objets individuels posent un risque alarmant pour la sécurité de la ville intelligente lorsque les connexions entre ces objets sont utilisées pour rendre la ville vraiment «intelligente». Ce phénomène augmente la surface d'attaque de la ville intelligente en offrant aux attaquants potentiels un vaste paysage pour compromettre la ville intelligente. Grâce à l'accès au réseau intelligent, les pirates peuvent collecter méthodiquement des informations sur l'état de sécurité des organisations opérant dans la ville intelligente [11]. Par exemple, si un réseau repose sur une connexion sécurisée à un smartphone, un smartphone secrètement compromis pourrait utiliser la connexion sécurisée à des fins malveillantes. C'est pour cette raison que la somme des vulnérabilités individuelles sera plus élevée que chacun des systèmes dépendants [12]. Cela pose des difficultés en raison de la nécessité précitée de sécurité dans une ville intelligente, sans laquelle les habitants seraient réticents à participer.

Alors que les smartphones connectent les utilisateurs à la ville intelligente, le réseau s'appuie également fortement sur la communication de machine à machine (M2M) qui automatisera de nombreux processus au sein de la ville intelligente. La communication de machine à machine peut avoir lieu après que les capteurs d'objets intelligents aient

franchi une valeur de seuil ou après avoir reçu des signaux d'un autre appareil. Selon l'article de Braun et al. [7] et l'article de Ijaz et al. [13], les appareils intelligents qui s'engagent dans une communication de machine à machine peuvent présenter un risque pour les villes intelligentes en raison de problèmes de sécurité tels que les attaques physiques, les attaques contre les jetons d'authentification, les attaques de protocole, les menaces de sécurité réseau, les violations de la confidentialité et les attaques de configuration [13].

Les attaques physiques utilisant une communication M2M compromise peuvent être exécutées via des attaques de configuration utilisant des logiciels malveillants pour commettre une fraude en manipulant l'intégrité du logiciel M2M existant et des données associées [13]. Les jetons d'authentification qui accordent à certaines machines l'accès au réseau intelligent peuvent être clonés et utilisés pour infiltrer le réseau. Les menaces liées à la sécurité du réseau, comme l'usurpation d'identité d'appareil et les attaques par déni de service (DoS), entre des appareils intelligents peuvent respectivement infiltrer ou perturber un réseau intelligent. Les attaques de protocole se produisent principalement contre des appareils, tels que les attaques de type man-in-the-middle, OAM et par déni de service (DoS) [13]. Toutes ces attaques perturbent les communications rapides et automatisées entre les appareils intelligents, ce qui permet un accès rapide aux informations dans une ville intelligente.

Un autre défi de sécurité dans la ville intelligente vient des vulnérabilités des étiquettes d'identification par radiofréquence (Radio Frequency Identification RFID), qui sont utilisées dans plusieurs secteurs potentiels des villes intelligentes, notamment l'environnement, l'industrie et la mobilité [13]. La RFID, ou une technologie similaire, a contribué à améliorer la visibilité et la traçabilité des informations en temps réel, mais elle est également sujette aux attaques et aux menaces qui diminuent sa posture de sécurité [13]. Les étiquettes RFID sont susceptibles de recevoir des demandes d'accès non autorisées qui permettent à l'attaquant d'accéder à des informations sensibles, ce qui porte atteinte à la confidentialité des données. Les étiquettes RFID peuvent être interceptées par un lecteur RFID qui produit le bon code de produit électronique (EPC), et les attaquants qui ont accès à des lecteurs RFID ou à un EPC peuvent intercepter et lire des étiquettes confidentielles. La corruption RFID peut être obtenue par la destruction d'étiquettes, le clonage d'étiquettes, l'interférence du signal, le brouillage, les attaques par déni de service et l'écoute clandestine [13]. Dans chaque situation, un attaquant peut perturber la fréquence et éloigner le message du destinataire prévu. La technologie RFID étant déjà utilisée pour plusieurs composants potentiels de systèmes intelligents, les vulnérabilités de sécurité de cette technologie doivent être corrigées.

1.4 Conclusion

Dans ce chapitre, nous avons défini une ville intelligente, ses différents infrastructures dans la gouvernance, la santé, les infrastructures essentielles, les bâtiments et le transport. Nous avons décrit son architecture IOT et ses caractéristiques principales. Les problèmes de sécurité ont également été vu concernant la sur-collecte de données et la vulnérabilité des données sensibles pour l'utilisateur. Dans le prochaine chapitre, des solutions qui ont été proposé par des experts vont être étudié pour le choix de la solution au problème de sécurité et de confidentialité des données.

Chapitre 2

La sur-collecte des données dans les villes intelligentes

2.1 Introduction

Dans le plan de ville intelligente, le smart-phone est l'appareil le plus utilisé dans tous les domaines. Il est donc très intéressant en tant que moyen de collecte de données des citoyens. Ce chapitre va introduire les différents approches de collecte de données, les techniques de collecte de données d'objets en mouvement massif et leur conséquences sur la vie privée des citoyens de la ville intelligente. Nous allons voir aussi quelques solutions qui ont été proposées ultérieurement dans le domaine de la protection de la vie privée.

2.2 Quelques approches de collecte de données de l'usager mobile dans les villes intelligentes

La sur-collecte de données : est la collecte de données plus que suffisante sur la fonction d'origine tout en étant dans la portée de l'autorisation (Yibin et al. [1]).

Les systèmes d'exploitation des téléphones portables ne fournissent que des autorisations à granularité grossière pour réglementer si une application peut accéder à des informations privées, tout en fournissant bien plus d'informations privées que l'application utilise réellement. Entre temps, certains utilisateurs comprennent ce que signifient ces permissions et décident parfois d'annuler l'installation ou de désinstaller les applications qui demandent ce genre de permission dans le but de protéger leurs données personnelles [1]. Cependant, dans le carde de ville intelligente, la confiance doit être présente entre les deux parties afin de ne pas limiter le nombre de citoyens coopérant pour l'évolution de la ville. Ci-dessous sont présentées quelques approches de collecte de données des smartphones :

Suivi de localisation

Les données de localisation sont les données les plus fréquemment utilisées dans les smart-phones. Elles peuvent être utilisées dans des applications dont la fonction principale est la localisation, l'organisation de photos, le service de réseaux sociaux (SNS), les recommandations d'achat et de restauration, ou encore la prévision météorologique. Bien

que les utilisateurs soient avertis chaque fois qu'une application a l'intention de capturer l'emplacement de l'utilisateur, ils choisissent généralement d'autoriser l'accès à leur localisation qui est exigé par l'application pour bénéficier des services que l'application a à offrir en retour. [1].

L'accès aux photos

l'album photo est également un outil trés utilisé dans les smartphones. Les gens l'utilisent non seulement pour garder des souvenirs, mais aussi pour des raisons de commodité, telles que la prise de photos de diapositives pour éviter d'écrire, ou encore imprimer l'écran d'un itinéraire pour le consulter en temps voulu. En conséquence, les smartphones contiennent une quantité assez importante d'images. En raison de la popularité des applications SNS (des applications telles que Facebook ou Instagram [14]), les gens ont souvent l'habitude de publier des images montrant ce qu'ils sont entrain de faire. Presque toutes les applications SNS demandent d'accéder à l'album photo des utilisateurs. Les photos de la vie des utilisateurs révèlent leur vie quotidienne.

L'accès au carnet d'adresses

D'après une étude de Consolvo et al. [15], pour contacter d'autres personnes plus facilement, les utilisateurs sont prêts à créer de nouveaux contacts, à reconstituer les contacts existants avec leur adresse e-mail, leur nouveau numéro de téléphone et leurs remarques. La fonctionnalité du carnet d'adresses offre aux utilisateurs la commodité de communiquer et de travailler. Cependant, l'accès aux carnets d'adresses des utilisateurs à partir d'applications présente de graves risques potentiels pour la sécurité. Le carnet d'adresses comprend les noms d'utilisateur, l'adresse physique, les numéros de téléphone, les adresses e-mail et d'autres notes. Comme pour les photos, les systèmes d'exploitation des smart-phones actuels ne fournissent pas d'autorisation détaillée.

L'accès au calendrier

Les applications de calendrier sont utilisées afin d'organiser le programme des utilisateurs, de suivre les événements et de rappeler aux utilisateurs les événements importants à venir. IOS et Android ont leurs applications d'agenda attachées au système d'exploitation, et il est impossible de les désinstaller à moins d'un jailbreak. Ces applications de calendrier peuvent facilement obtenir l'autorisation du calendrier des utilisateurs, car leur fonction principale consiste à gérer le calendrier des utilisateurs. En conséquence, très peu d'utilisateurs refusent l'autorisation des applications de calendrier à leur calendrier [1]. Par ailleurs, il existe d'autres types d'applications qui accèdent au calendrier telles que les applications de style de vie, de voyages etc. Une fois l'autorisation accordée au calendrier d'un utilisateur, toutes ces applications accéderont à chaque ligne d'informations stockée dans ce calendrier.

Suivi IMEI/UDID

L'IMEI et l'UDID sont tous deux l'ID unique d'un téléphone particulier et ils sont comme des cookies sur le site Web que les utilisateurs ne peuvent supprimer. Bien qu'Apple

ait interdit aux développeurs iOS d'utiliser l'UDID comme méthode pour suivre et identifier les utilisateurs, cette règle n'est appliquée que sur les appareils dotés de la dernière version d'iOS. En effet, Apple a encouragé les développeurs d'applications à utiliser de nouvelles méthodes d'identification des utilisateurs pour suivre leur comportement application par application. D'après le rapport d'Appthority [16], 88% des meilleures applications gratuites Android et 65% des meilleures applications payantes Android accèdent à IMEI / UDID, alors que seulement 57% des meilleures applications gratuites iOS et 28% des meilleures applications payantes iOS accèdent IMEI / UDID. L'identification IMEI / UDID est tout aussi importante que la clé primaire d'une base de données. Il identifie chaque donnée des téléphones et classe toutes les données par appareil.

2.3 Collecte de données d'objets en mouvement massif dans une ville intelligente

Actuellement, de nombreux capteurs et appareils peuvent être utilisés pour détecter et signaler des informations de localisation. Selon les types de capteurs et d'appareils décrits par l'article [17] , les principales sources de données de trace peuvent être regroupées en quatre catégories :

— **Appareils mobiles :** Les appareils mobiles, tels que les téléphones et les tablettes, deviennent omniprésents. Ces appareils portables pourraient fournir diverses données de localisation à l'aide du GPS, du WiFi, du GSM et du Bluetooth. Habituellement, les données de trace des appareils mobiles reflètent approximativement les mouvements de leurs propriétaires.

— **Les véhicules :** De nos jours, de plus en plus de véhicules sont équipés d'appareils GPS pour les services de navigation. Les traces GPS d'un véhicule représentent non seulement la trajectoire du véhicule lui-même, mais aussi celle de son conducteur et de ses passagers. Les traces de véhicules privés et publics ont généralement des exigences de confidentialité différentes.

— **Cartes à puce :** les cartes bancaires et les cartes de transport sont deux types de cartes à puce typiques dans une ville. Chaque activité de consommation avec une carte à puce est associée à un lecteur, dont l'emplacement est généralement fixe.

— **Capteurs flottants :** Un objet équipé d'un module de localisation peut signaler des traces de lui-même, comme le suivi des déchets et des traces de cargaisons avec des étiquettes d'identification par radio-fréquence (RFID). Ces capteurs flottants permettent de collecter des données de trace pour différents types d'objets en mouvement.

La signification sémantique des données de trace dépend de la technologie de localisation. Les méthodes de localisation couramment utilisées pour générer des données de trace comprennent le GPS, le WiFi, le GSM, le Bluetooth et la RFID.

Tech.	Données	Refs.	Exp.	Précision	Couverture
GPS	Coordonnées géographiques	Absolue	physique	1–5 meters (95–99%)	Plein air
WiFi	ID point d'accès + Force du signal ou coordonnées locales	Relative	Symbolique/ physique	1–20 mètres	< 100 mètres depuis un point d'accès
Tour cellulaire	ID tour cellulaire + Force du signal ou coordonnées géographiques	Relative/ absolute	Symbolique/ physique	50–200 mètres en villes	Couverture cellulaire. 5–30km depuis une tour.
BT	ID appareil	Relative	Symbolique	plage de détection du Bluetooth	5–10 mètres pour classe 1 ;20–30 mètres pour classe 2
RFID	ID/position lecteur	Relative/ absolue	symbolique/ physique	plage de détection de RFID	1 mètre pour RFID passif ; 100 mètres pour RFID actif

TABLE 2.1 – Comparaison de technologies de localisation populaires [17].

Dans le tableau 2.1, Gang Pan et al ont effectué une comparaison entre ces technologies de captures de données selon quatre propriétés :

— **Référence,** définit si la donnée est une localisation absolue (par exemple 36˚30'17"N 2˚52'31"E) ou relative (par exemple l'identité d'un Bluetooth mouvant détecté par un téléphone).
— **Expression,** décrit si la trace est physique (par exemple 36˚30'17"N, 2˚52'31"E) ou symbolique (par exemple Université de Saad dahlab, Blida).
— **Précision,** La précision de chaque localisation dans la donnée.
— **Couverture,** représente la plage valide de la méthode.

2.4 Data mining

Il existe plusieurs définitions du data mining. Celle qui a été retenu est présentée par Hand et al. dans leur article sur le "Data mining"[18]. Le data mining est la découverte de structures et de modèles dans des ensembles de données volumineux et complexes. Il comporte deux aspects : la création de modèles et la détection de modèles. La construction de modèles dans le data mining est très similaire à la modélisation statistique, bien que de nouveaux problèmes se posent en raison de la grande taille des ensembles de données et du fait que le data mining est souvent une analyse de données secondaire.

2.5 Conséquences de sur-collecte de données

Les utilisateurs souffrent de la fuite potentielle d'informations de leur vie privée lorsqu'ils profitent de la commodité apportée par les applications mobiles. Dans une ville intelligente, il peut y avoir beaucoup de fuites malgré les efforts pour les éviter.

Avec le développement de la technologie électronique, toutes sortes de smartphones affluent sur le marché et les smartphones usurpent de plus en plus la vie des gens avec des applications goanywhere offrant un large éventail de services d'entreprise, sociaux, financiers et récréatifs. De plus en plus de développeurs d'application partagent leurs applications dans le marché sans trop se soucier des différents obstacles de marketing, d'installation ou de mise à jour. Outre la fonctionnalité qu'apportent les applications,

elles posent d'énormes problèmes de sécurité [1].

La sur-collecte de données présente beaucoup de conséquences pour la vie privée de l'utilisateur, voici quelques unes cités par Yibin et al. [1] :

— **La localisation :** Considéré comme le premier et le plus direct des risques, les problèmes de sécurité physique. Les traces des utilisateurs sont facilement exposées à quelqu'un qui dispose de ses données de localisation en temps réel et précises. En utilisant des méthodes simples d'exploration de données, les habitudes et les coutumes des utilisateurs sont faciles à déduire. Le harcèlement, le cambriolage ou même le meurtre sont des risques à prendre en compte. Le deuxième risque concerne les préoccupations des entreprises, du gouvernement et de l'espionnage martial.

— **L'album photo :** Les photos de vie des utilisateurs révèlent leur vie quotidienne. L'exposition de photos enfreint non seulement les droits des utilisateurs tels que le portrait, mais peut également nuire à la réputation des utilisateurs. De plus, les photos des utilisateurs sur les informations sont beaucoup plus précieuses que les photos quotidiennes. Avec l'aide d'applications de sur-collecte de données, les organisations tierces peuvent obtenir des photos collectées à partir des smart-phones des utilisateurs et les utiliser pour le commerce, comme par exemple, l'affichage de promotion sur des articles qui intéressent un client précis. Ce type de comportement équivaut à voler des actifs aux utilisateurs. Par exemple, un concepteur fait germer une nouvelle idée et dessine un brouillon, puis il prend des photos pour enregistrement. Il sera directement perdu si ses photos sont collectées par une application et vendues à ses concurrents. Plus généralement, il est exaspérant de trouver une photo de soi montrée sur une publicité ou quelque part dépassant ses attentes.

— **Le carnet d'adresse :** Les contacts des utilisateurs ont une grande valeur. Ces données sur les contacts peuvent être utilisées par les développeurs d'applications pour étendre leur clientèle et utilisées par des organisations tierces pour commercialiser des applications ou des services mobiles supplémentaires destinés aux clients de la liste de contacts. En outre, la sur-collecte de données de contacts peut apporter un espionnage potentiel des entreprises. Si un utilisateur branche son smart-phone sur son bureau d'entreprise au travail, il lui donnera la possibilité de se synchroniser avec les contacts des logiciels de messagerie tels qu'Outlook, qui incluent toujours les contacts personnels et d'entreprise. Cependant, l'application demande uniquement à l'utilisateur s'il doit autoriser l'accès au carnet d'adresses, même si les contacts appartiennent à l'entreprise. Cela réduit considérablement la sécurité des données d'entreprise surtout dans le cadre d'une ville intelligente.

— **Le calendrier :** Le principal risque de sur-collecte de données de calendrier est l'espionnage d'entreprise. Les données du calendrier incluent des informations sur la réunion et les événements des utilisateurs. Les informations sur la réunion sont des secrets commerciaux. Ce sera une perte énorme pour une société si les informations de calendrier telles que les informations d'appel, les mots de passe d'appel, les personnes présentes et les sujets discutés ont été divulgués par ses

concurrents. Pour le moins, la divulgation d'informations sur les événements personnels des utilisateurs est également source de harcèlement pour les utilisateurs, si ces informations étaient mises en possession d'organisations tierces, telles que des agences de publicité.

— **L'accès IMEI / UDID :** L'accès principal à IMEI / UDID est lié au fait que le comportement des utilisateurs peut être corrélé entre plusieurs applications et adapté à un utilisateur unique. Même si un utilisateur a différents noms d'utilisateur et mots de passe pour chaque application, ses données peuvent facilement être intégrées par l'ID unique de son appareil. En outre, l'IMEI / UDID peut également être utilisé pour correspondre à des données utilisateur réelles, telles que des noms, des mots de passe, des emplacements et autres. Il permet aux développeurs d'applications et aux réseaux publicitaires de créer un profil complet d'un utilisateur sur plusieurs applications et profils et de les combiner avec d'autres données surcollectées pour une vue approfondie des utilisateurs.

2.6 Quelques solutions proposées pour la résolution des problèmes de confidentialité

Au cours de cette section, quelques solutions de protection des données des citoyens seront mis en avant. Nous verrons comment se fait la sécurisation des applications et des bases de données pour apporter confiance et authenticité aux données sensibles.

2.6.1 W3 Privacy pour les applications Location-Based Service (LBS)

Les services géo localisés (Location-Based Services - LBS) gagnent progressivement en importance. Bien qu'ils aient un énorme potentiel, les risques de confidentialité qu'ils pourraient présenter pour les utilisateurs pourraient freiner leur croissance. Pablo A. Pérez-Martínez et al. [19] ont distingué trois dimensions principales indépendantes pour la confidentialité des données d'utilisateurs d'applications LBS, à savoir la confidentialité de l'identité (**Who ?**), la confidentialité de l'emplacement (**Where ?**) et la confidentialité des requêtes (**What ?**). Ils précisent qu'un LBS est W3-privé lorsque le service est fourni alors que le fournisseur (i) ne sait pas qui est l'utilisateur, (ii) où est l'utilisateur, et (iii) ce que l'utilisateur demande.

— **WHO :** Un bon moyen de protéger l'identité est l'utilisation de routeurs à oignons [20]. Néanmoins, d'après [19], cela ne suffit pas lorsque les fournisseurs LBS exigent l'authentification des abonnés. Dans ces cas, un protocole de paiement anonyme doit être utilisé. Pablo A. Pérez-Martínez [19] suggère d'utiliser la signature partiellement aveugle d'Abe et Okamoto [21]. Le principe de ce dernier est qu'avant d'envoyer des requêtes à un fournisseur LBS, chaque utilisateur s'abonne via ce protocole. Ce faisant, les utilisateurs peuvent s'authentifier sans révéler leur identité et les fournisseurs peuvent les facturer correctement pour le service.

— **WHERE :** Les personnes vivant et se déplaçant dans les villes ne se répartissent pas uniformément. Au contraire, chaque ville a des rues, des avenues et des places avec une plus grande densité de population. Ainsi, il est possible d'obtenir une carte de densité telle que celle illustrée à la figure 2.1. Pablo A. Pérez-Martínez et al supposent que ces cartes sont communes à tous les utilisateurs LBS et peuvent être librement consultées. Ils peuvent être compris comme un ensemble de points en R3(x, y, d), où x et y représentent la longitude et la latitude, et d est la densité de ces coordonnées. En utilisant ces informations, les utilisateurs peuvent déterminer les points de densité maximale, qui sont d'excellents candidats pour être utilisés comme faux emplacements car ils rendent difficiles les attaques d'identification d'objet (OI) et d'identification d'espace restreint (RSI).

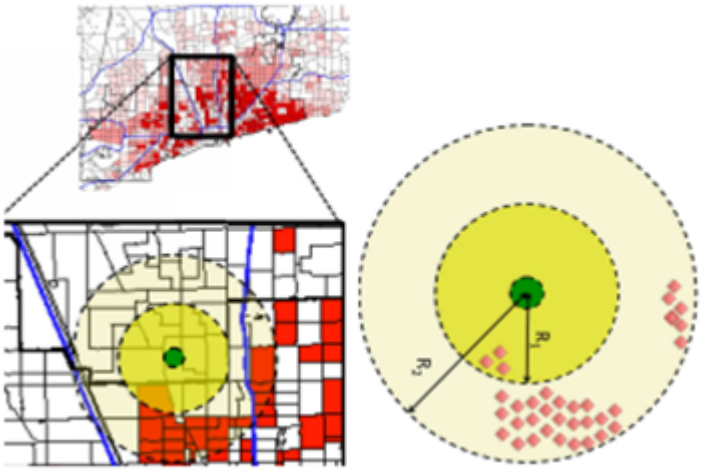

FIGURE 2.1 – (Gauche) Carte de densité de Chicago et les utilisateurs qui y sont localisés. (Droite) Radius d'incertitude atour de l'utilisateur et points candidats [19].

Pour préserver leur «confidentialité de localisation», les utilisateurs définissent une série de rayons d'incertitude $R = \{R1, R2, ..., Rm\}$ dans lequel ils essaient de trouver des points avec au moins une densité minimale $dmin$. Ces points pourraient être utilisés comme de faux points. Tout d'abord, les utilisateurs se localisent (par exemple avec un GPS). Ensuite, ils sélectionnent la zone (définie par le rayon) dans laquelle ils veulent trouver de faux points. Enfin, ils utilisent les informations de la carte de densité pour déterminer les points qui ont une densité supérieure à $dmin$ qui se trouvent dans une plage inférieure à un Ri donné. Si aucun point ne remplissant les conditions requises n'est trouvé, l'utilisateur doit assouplir les contraintes et recommencer la procédure. Une fois les points candidats trouvés, les utilisateurs peuvent sélectionner l'un d'entre eux au hasard (en alternant, les utilisateurs peuvent sélectionner le point le plus proche d'eux, réduisant ainsi l'erreur tout en maintenant l'incertitude).

— **WHAT :** Dans le but de préserver la confidentialité des requêtes, les utilisateurs génèrent un certain nombre de requêtes différentes $\{q1, q2, ..., qk\}$. Ainsi, le fournisseur n'est pas en mesure de déterminer laquelle est la vraie requête. Cette idée ressemble au concept de « k-anonymat » [9] et a été largement utilisée pour ga-

rantir la confidentialité des requêtes avec les moteurs de recherche Internet.

2.6.2 Modèles d'évaluation et de sécurisation de données

Yibin Li et al. [1] affirment que toutes les données sont confrontées au risque potentiel d'être sur-collectées, en particulier dans les smartphones. Le risque de sécurité est la valeur représentant le dégât potentiel à la sécurité lorsqu'un certain risque survient. Dans leur article [1], Yibin Li et al. proposent trois modèles différents pour la protection des utilisateurs dans le cas de la sur-collecte de données :

Modèle de sécurité et consommation

Les Smartphones enregistrent différents types de données, certains plus sensibles que d'autres. Pour cela, Yibin et al. identifient différents niveaux de sécurité [1]. Les données les plus sensibles, incluant les informations personnelles telles que la localisation, les adresses, les numéros de téléphones etc., requièrent un niveau de sécurité élevé tandis que les données partagées ou publique requièrent un niveau de sécurité bas. Le reste des données leur seront accordés un niveau de sécurité moyen. Après avoir précisé tous les niveaux de sécurité, les données pourront être stockées dans différents services de stockage dans le Cloud. Les données à sécurité basse sont stockées dans un service de stockage simple. Le niveau de service de stockage augmente en parallèle avec le niveau de sécurité de la donnée. Une relation linéaire existe entre la consommation et la complexité du chiffrement. α représente la relation linéaire entre le niveau de sécurité et le temps d'exécution, et β représente la relation linéaire entre la consommation d'énergie EC et la complexité CX comme suit :

$$SL = \alpha * T, EC = \beta * CX, T = CX \tag{2.1}$$

En conséquence, nous pouvons formuler la relation entre SL et EC comme suit :

$$SL = \frac{\alpha}{\beta} * EC \tag{2.2}$$

Modèle de permission d'application

Jusqu'à présent, pour toutes les applications, des autorisations à granularité grossière sont accordés. C'est le principal coupable de la sur-collecte de données. Ainsi, pour quantifier ces permissions, différentes autorisations sont utilisées. Elles sont accordées, à tous, à certains ou à aucun [1]. Si une application a la permission d'accéder à N données d'utilisateurs, qui a M données au total, la permission de cette application est $N = M$. pour décrire l'influence des permissions pour les données de différents niveaux de sécurité, SL est pris en considération. La permission finale est représentée comme suit :

$$Perm = SL * N/M \tag{2.3}$$

Modèle de risque de sécurité pour la sécurité de la sur-collecte de données

Pour formuler le problème de sécurité causé par la sur-collecte de données, il est nécessaire d'introduire un modèle de risque de sécurité [1]. La sur-collecte de données est une sorte de risque potentiel, qui est le produit de la probabilité de violation de la sécurité et des dommages du problème de sécurité. En conséquence, Yibin et al modélise le risque de sécurité SR de l'application i vers les données d comme suit :

$$SR_d^i = SL^d * Pro_d^i, \tag{2.4}$$

où SR_d^i signifie le risque de sécurité de l'application i sur-collectant les données d. SL^d est le niveau de sécurité des données d et Pro_d^i est la probabilité que l'application i utilise les données d pour endommager la sécurité, ce qui, selon Yibin et al. [1], peut être formulé comme suit :

$$Pro_d^i = 1 - e^{-\lambda * N_i^d / M_i}. \tag{2.5}$$

Dans l'équation 2.5, λ est le coefficient de risque de sécurité du comportement de l'application i sur-collectant les données d, qui peut être ajusté par différentes application et données mais fixé sur un seul scénario. Plus de données sur-collecté amène à une plus grande probabilité de violation de sécurité, qui amène par la suite à un risque de sécurité plus important. Selon les deux équations 2.4 et 2.5, la quantité de données sur-collectée N/M pour formuler les risques de sécurité d'application i envers la données d est présentée dans la formule 2.6 :

$$SR_i^d = SL^d * (1 - e^{-\lambda * N_i^d / M_i}). \tag{2.6}$$

N_i^d/M_i varient entre 0 et 1, ou 0 veut dire que l'application i n'a aucune permission d'accéder à la donnée d de l'utilsateur, et 1 que toutes les données d sont sur-collectés par cette première. Selon différents niveaux de sécurité (SL=1, SL=2, Sl=3), le calcul de la relation entre le risque de sécurité et la quantité de données sur-collectées est montré respectivement dans les figures ci-dessous :

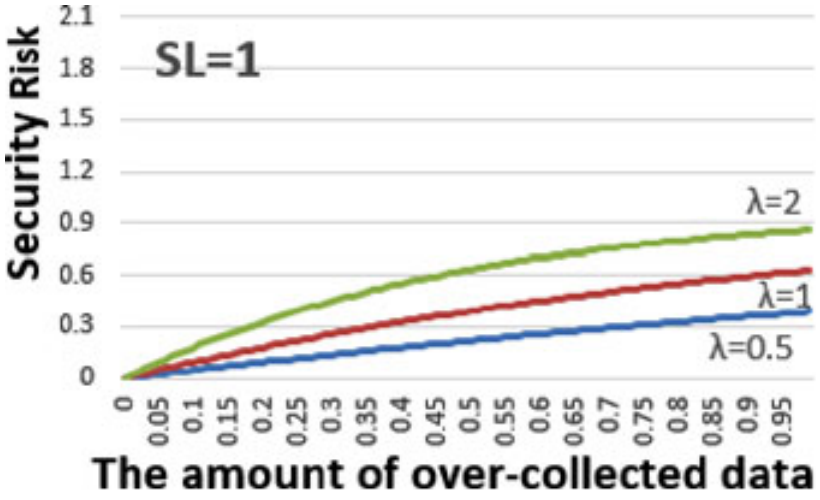

FIGURE 2.2 – la relation entre le risque de sécurité et la quantité de données sur-collectées avec SL=1 [1]

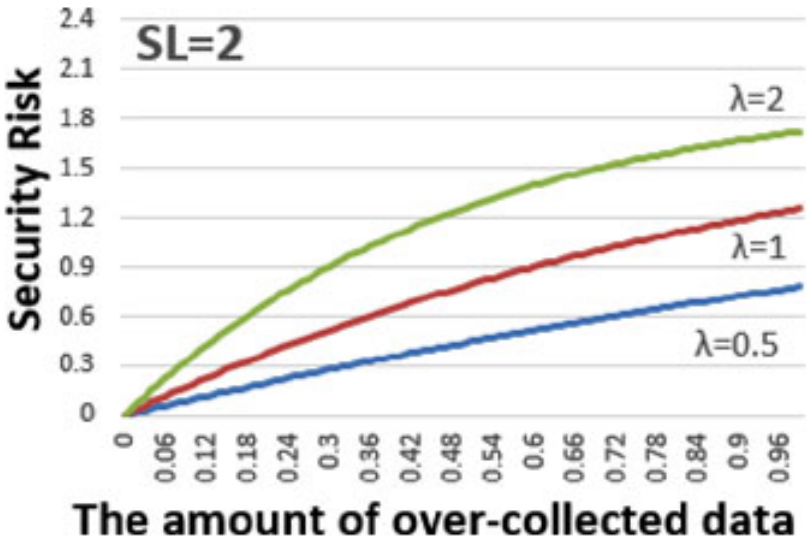

FIGURE 2.3 – la relation entre le risque de sécurité et la quantité de données sur-collectées avec SL=2 [1]

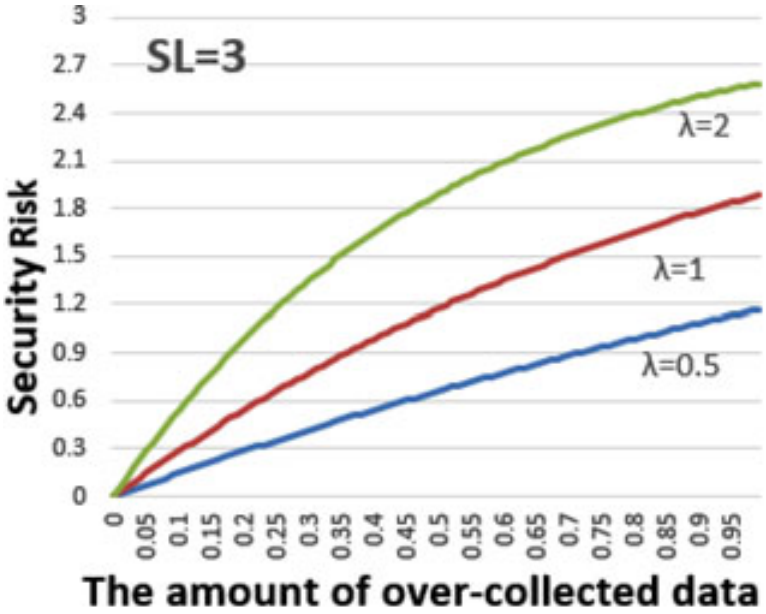

FIGURE 2.4 – la relation entre le risque de sécurité et la quantité de données sur-collectées avec SL=3 [1]

À partir des études effectuées pour calculer les risques des données et les figures 2.2, 2.3 et 2.4, il devient évident que plus les données sont sur-collectées, plus le risque de sécurité est élevé. Plus encore, le niveau de sécurité a un impact direct sur le risque de sécurité. En pratique, de nombreuses applications sur-collectent plusieurs types de données. Yibin Li a donc formulé le risque de sécurité de l'application i envers un utilisateur (U) comme suit :

$$SR_i^u = \sum_{d=0}^{M} SL^d * (1 - e^{-\lambda * N_i^d / M_i}). \tag{2.7}$$

2.6.3 La conception d'une framework cloud-mobile

Pour résourdre le problème de sur-collection de données, Yibin li et al [1] ont proposé une architecture de cloud-mobile où toutes les données des utilisateurs seront stockées et où les smart-phones ne feront que gérer les applications et montrer leurs résultats (figure 2.5).

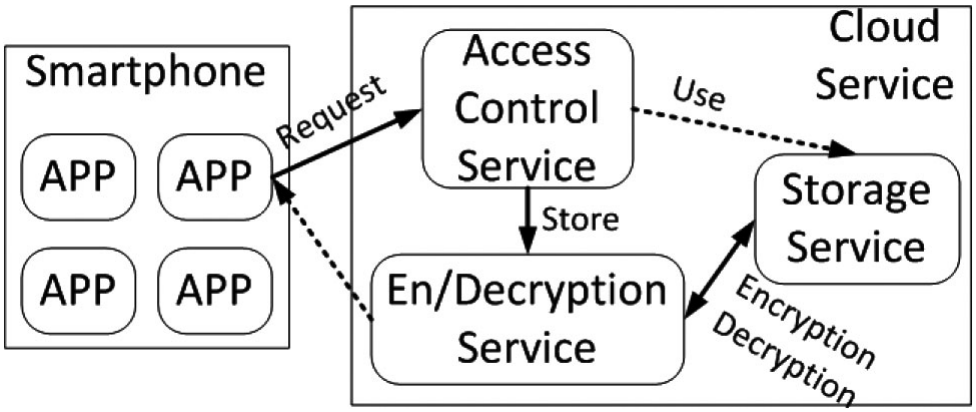

FIGURE 2.5 – l'architecture cloud-mobile [1].

Bien que la sécurité d'un service cloud ne soit pas parfaite de nos jours, les fournisseurs de services cloud sont beaucoup plus professionnels que les développeurs et les utilisateurs d'applications. En utilisant le service cloud, l'opération de cryptage et de décryptage des données peut être effectuée dans le cloud, et les applications fonctionnent comme Data Requester demandant des données au cloud. Selon Yibin li et al [1], le processus peut être implémenté basé sur l'architecture CP-ABE [22] ou alors son amélioration comme PP-CP-ABE [23]. Yibin et al. [1] proposent un algorithme pour décrire la première partie, comme le montre l'algorithme 1. Dans cette partie, l'utilisateur (UR) prend une photo en utilisant la fonction intégrée «photoing and sharing» de l'application. Tout d'abord, nous devons décider si cette application peut accéder à la caméra d'UR. Nous divisons ce type d'autorisations en deux types : l'accès aux données et l'accès au matériel. Si la demande de l'application accède au matériel, nous définissons l'autorisation d'être autorisée, car le matériel ne peut pas faire de mal seul. Si l'application demande d'accèder aux données, nous renvoyons la demande au cloud pour évaluation. Comme mentionné précédemment, le cloud peut utiliser ABE ou d'autres méthodes pour implémenter le service de contrôle d'accès, ce qui n'est pas la clé de notre recherche. Après avoir donné l'autorisation, le service de chiffrement dans le cloud utilisera certaines méthodes de chiffrement pour protéger les données.

Algorithme 1 : Une application stocke des données dans le cloud

Input : appID, userID, data;
Juger le type de cette demande d'accès par $appID$, mettre le résultat dans une variable appelé T;
if T == *hardware* **then**
    UR donne la permission à $appID$;
else
    Envoyer requête au service de controle d'accès comprenant $appID$, $userID$ et $data$;
    Le service de contrôle d'accès décide si cette application a la permission par $appID$, $userID$ et $data$ et accessControlList, mettre le résultat dans une variable appelé P;
    if P == *true* **then**
        Le service de chiffrement crypte $data$;
        Stocker $data$ dans le stockage cloud avec un label $appID$;
    else
        return;

Algorithme 2 : Une application utilise des données du cloud

Input : $appID$, $userID$, informations sur la donnée exigée PD;
Output : contenu concret de la donnée exigée D;
Envoyer requête au service de controle d'accès comprenant $appID$, $userID$ and PD
Le service de contrôle d'accès estime si l'application a la permission par $appID$, $userID$, PD and $accessControlList$, mettre le résultat dans la variable P
if P == *true* **then**
    Service de contrôle d'accès envoie une requête avec $appID$, $userID$ et PD au service de stockage
    Service de stockage cherche la donnée chiffrée $data$ par $userID$ et PD
    Service de stockage envoie $data$ avec $appID$ et $userID$ au service de déchiffrement
    Service de déchiffrement vérifie la permission encore une fois en corespondant $appID$ avec $data$ et mettre le résultat dans la variable P'
    if *P' == true* **then**
        Service de déchiffrement décrypte $data$ en D
        return D
else
    return $none$

Pour libérer totalement la charge de fonctionnement des utilisateurs, Yibin li propose un algorithme pour permettre les autorisations sur les applications, comme indiqué dans l'algorithme 3 [1]. Les utilisateurs n'ont qu'à définir un niveau de sécurité par défaut, similaire au niveau de sécurité d'un navigateur. Ce niveau de sécurité peut être exprimé comme «extrêmement élevé», «élevé», «normal», «faible» et «extrêmement faible». Dans l'algorithme 3, Yibin li calcule le risque de sécurité des données et le risque de sécurité

de l'application en utilisant l'équation 2.6 et 2.7. Puis juge que l'autorisation de cette application dépassera le risque de sécurité par défaut. S'il dépasse le risque de sécurité, l'autorisation échoue. S'il ne dépasse pas le risque de sécurité, l'autorisation passe.

Algorithme 3 : Donner la permission à une application

Input : $appID$, $userID$, taille de donnée exigée n et le type de donnée t, risque de sécurité par défaut DSR

Output : La permission pour accéder à p

Envoyer une demande au service de contrôle d'accès incluant $appID$, $userID$ et n ;
Prendre le niveau de sécurité par défaut de l'utilisateur, $SR \leftarrow DSR$;
Prendre le niveau de sécurité sl de cette application ;
Prendre la taille totale de t type de donnée m ;
Calculer la proportion de donnée n envers donnée total m, $d = n/m$;
Calculer le risque de sécurité de donnée n, $sl_n = sl * (1 - e^{-\lambda * d})$;
if $sl > SL$ **then**
    return pas de permission ;
else
    Calculer le risque de sécurité de l'application $appID$, $sl_{appID} = \sum_{d=0}^{M} sl^d * (1 - e^{-\lambda * n_i^d/m})$;
    if $sl_appID > SL$ **then**
        return pas de permission ;
    else
        return permission ;

2.6.4 Protection de la vie privée par K-anonymity

K-anonymity est un concept introduit par Latanya Sweeney et Pierangela Samarati dans un papier publié en 1998 comme solution au problème suivant : "Étant donné les données structurées sur le terrain spécifiques à chaque personne, produire une publication des données avec des garanties scientifiques que les individus qui sont les sujets des données ne peuvent pas être ré-identifiés tandis que les données restent pratiquement utiles" [9].

Une divulgation de données aurait la propriété k-anonymat si les informations de chaque personne contenues dans la divulgation ne peuvent être distinguées d'au moins k - 1 personnes dont les informations figurent également dans la divulgation.

Dans le contexte des problèmes de k-anonymisation, une base de données est une table avec n lignes et m colonnes. Chaque ligne représente un enregistrement relatif à un membre spécifique d'une population. Contrairement à une base de données relationnelle, les lignes n'ont pas besoin d'être uniques. Les valeurs dans les différentes colonnes sont les valeurs des attributs associés aux membres de la population. La table 2.2 illustre une base de données non-anonymisée constituant des dossiers des patients d'un hôpital fictif.

Ehnicity	Date of birth	Sex	ZIP	Marital Status	Problem
Asian	09/27/64	female	02139	divorced	hypertension
Asian	09/30/64	female	02139	divorced	obesity
Asian	04/18/64	male	02139	married	chestpain
Asian	04/15/64	male	02139	married	obesity
Black	03/13/63	male	02138	married	hypertension
Black	03/18/63	male	02138	married	shortness of breath
Black	09/13/64	female	02138	married	shortness of breath
Black	09/07/64	female	02138	married	obesity
White	05/14/61	male	02138	single	chest pain
White	05/08/61	male	02138	single	obesity
White	09/15/61	female	02142	widow	shortness of breath
Black	05/05/61	male	02137	widow	chest pain

TABLE 2.2 – Base de données non-anonymisée d'un hopital fictif [9]

Il y a 6 attributs et 10 enregistrements dans ces données. Il existe deux méthodes courantes pour obtenir le k-anonymat pour une certaine valeur de k.

1. **Suppression :** Dans cette méthode, certaines valeurs des attributs sont remplacées par une '*'. Toutes ou certaines valeurs d'une colonne peuvent être remplacées par '*'. Dans le tableau anonymisé ci-dessous, nous avons remplacé toutes les valeurs de l'attribut «Name» par une «*» (voir table 2.3).
2. **Généralisation :** dans cette méthode, les valeurs individuelles des attributs sont remplacées par une catégorie plus large. Par exemple, la valeur 'Asian' de l'attribut 'Ehnicity' peut être remplacée par 'Person', la valeur 'black' par 'person', etc. La généralisation génère une sorte d'hierarchie où plusieurs valeurs sont remplacés par une autre qui les regroupe comme le montre la figure 2.6 :

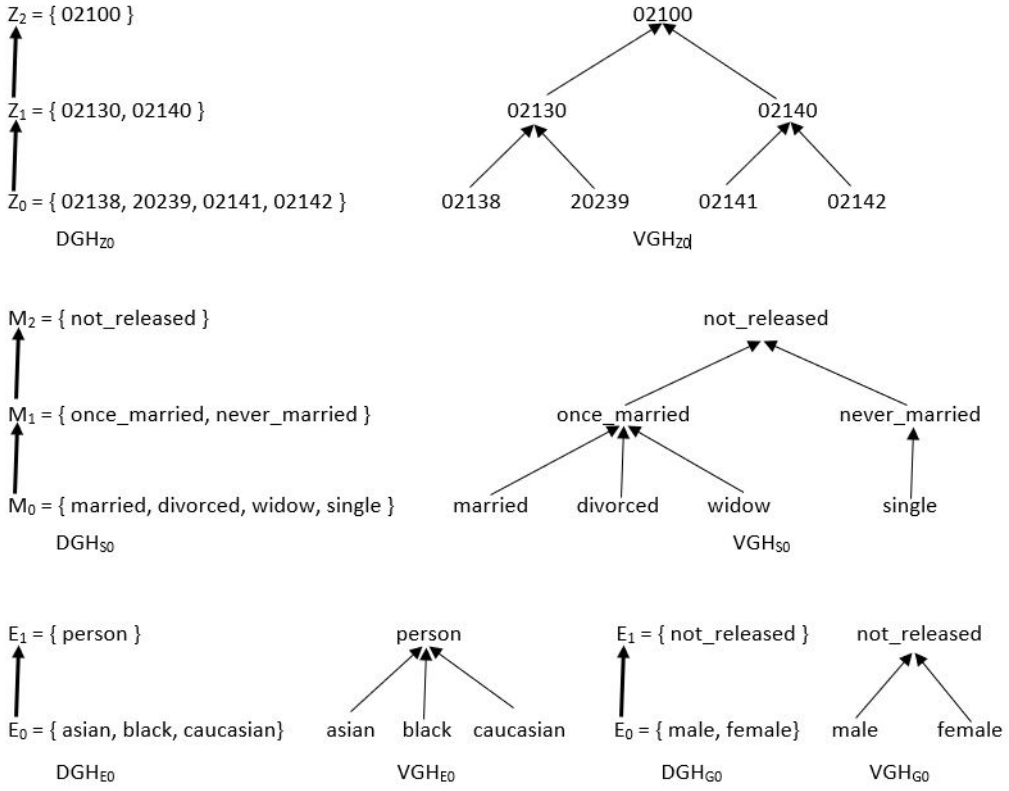

FIGURE 2.6 – Exemple de généralisation dans K-anonymity [9]

Ehnicity	Date of birth	Sex	ZIP	Marital Status	Problem
Person	64	*	021*	been married	hypertension
Person	64	*	021*	been married	obesity
Person	64	*	021*	been married	chestpain
Person	64	*	021*	been married	obesity
Person	63	*	021*	been married	hypertension
Person	63	*	021*	been married	shortness of breath
Person	64	*	021*	been married	shortness of breath
Person	64	*	021*	been married	obesity
Person	61	*	021*	never been married	chest pain
Person	61	*	021*	never been married	obesity
Person	61	*	021*	been married	shortness of breath
Person	61	*	021*	been married	chest pain

TABLE 2.3 – Base de données anonymisée d'un hopital fictif [9]

2.6.5 Avantages et limites des solutions proposées

Chaque solution parmi les solutions représentées ci-dessus répondent à un problème de sécurité bien précis. Avec leur combinaison, la sécurité de la ville intelligente pourrait s'améliorer d'une façon considérable. Néanmoins, avec le développement des villes intelligentes, elles ne suffisent plus et exigent une technologie plus avancée. Avec une sécurité accrue, la confiance des citoyens est gagné, permettant par la suite l'épanouissement de la ville dans le domaine technologique. Certaines des techniques de sécurisation citées ont été pris en considération pour le développement de la solution finale de ce travail de PFE (K-anonymity, cloud-mobile, W3 privacy).

2.7 Conclusion

Dans ce chapitre, nous avons, tout d'abord, vu quelques approches de collectes de données des objets en mouvement massifs tels que les smart-phones, véhicules, cartes à puce et capteurs flottants. Ensuite, Plusieurs solutions ont été traité pour la sécurisation des données des utilisateurs, c'est-à-dire, l'anonymisation de leur informations personnelles, le stockage sécurisé des données à travers le cloud, la sécurisation de leur système de localisation de leur smart-phones (W3-Privacy). L'algorithme K-anonymity a également été étudié afin d'expliquer son fonctionnement et démontrer son rôle dans la sécurisation des données personnelles des gens. Dans le prochain chapitre, nous entamerons l'analyse du système d'informations et les solutions qui ont été choisis pour la résolution des problèmes de sécurité.

Chapitre 3

Analyse et conception

3.1 Introduction

Après avoir introduit la ville intelligente et les problèmes de sécurité auxquels elle fait face, il est maintenant nécessaire de mettre des bornes à l'environnement étudié. L'objectif de ce chapitre est, en premier lieu, la capture des besoins et l'introduction des acteurs. Par la suite, Il va décrire les solutions qui ont été retenues à partir du chapitre précédent avec plus de précision et d'autres qui ont été légèrement modifiées pour la préservation de la vie privée des usagers mobiles dans une ville intelligente.

3.2 Méthode d'analyse et identification des besoins

Afin de réaliser une telle opération, ce rapport passe par plusieurs étapes au cours de ce chapitre. Nous proposons un modèle en cascade pour décrire la déroulement de ce chapitre :

FIGURE 3.1 – Méthode d'analyse et conception proposée en un modèle en cascade

Il est important de décrire chaque méthode en détails en citant les outils utilisés pour la réalisation des objectifs.

3.2.1 Capture des besoins

Dans cette partie, nous allons identifié et analysé les besoins des développeurs d'applications mobiles basés sur la localisation dans le but d'intégrer un système qui permet

la sécurisation des données de leurs utilisateurs. Nous avons identifié des besoins fonctionnels et non fonctionnels :

Besoins fonctionnels

Les besoins fonctionnels représentent les actions que le système doit assurer selon la demande du client. Ils déterminent quels types de sécurité il souhaite appliquer sur son application, à ses données et indiquer le niveau de sécurité à appliquer aux informations en question.

La figure 3.2 représenté ci-dessous décrit le système du projet. Trois parties sont à prendre en considération : Les données et toutes les informations nécessaires de développement d'application mobile, les besoins des utilisateurs mobile et les fonctionnalités que notre solution propose. Pour garantir une sécurité et une confidentialité des données, on doit tenir compte des besoins des clients et avoir l'accès à leurs données :

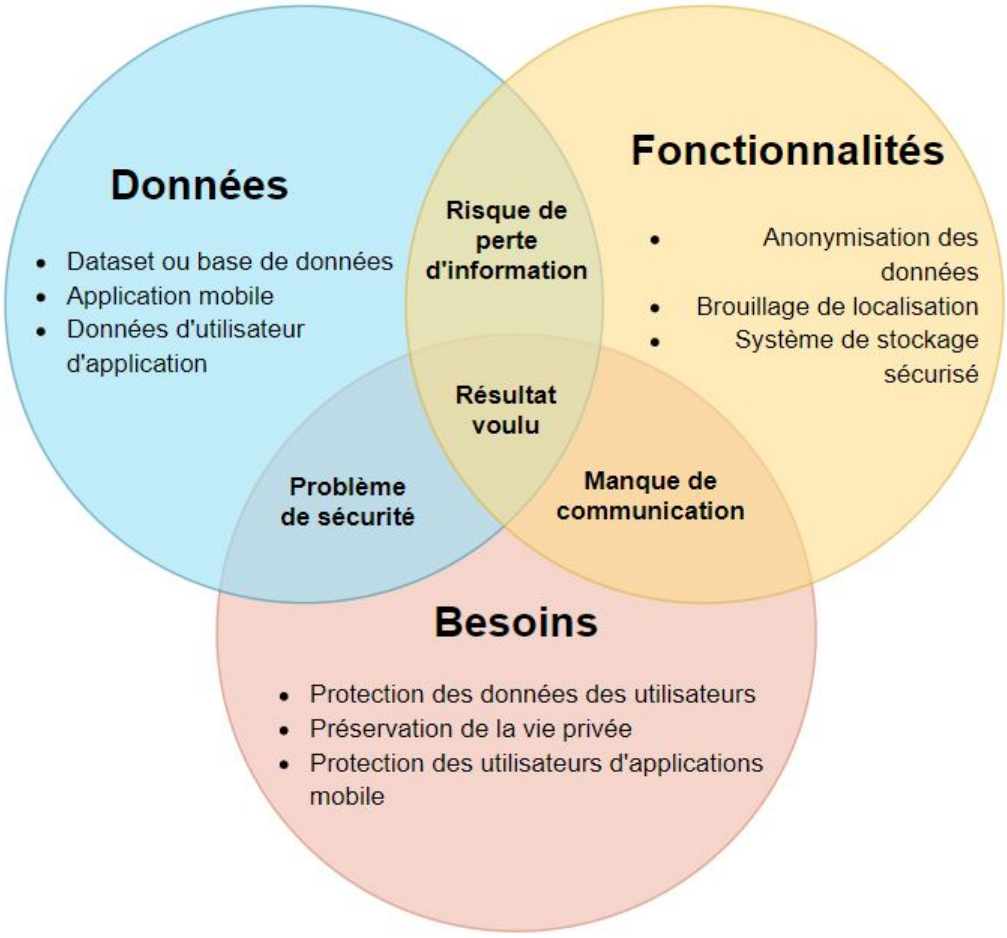

FIGURE 3.2 – Diagramme de relations entre les fonctionnalités et le client

Besoins non fonctionnels

Il existe également des besoins auxquels la solution doit obligatoirement répondre pour le bon déroulement de l'analyse fonctionnelle, par ailleurs, la solution doit être :

— compatible avec le langage de programmation des applications mobiles.
— indépendante du type de dataset auquel elle a à faire (ne doit pas être dépendante d'un seul dataset en particulier).
— ne doit pas dépendre d'un seul type d'application.

3.2.2 Identification des acteurs et de leurs objectifs

Dans ce cas d'étude, il existe deux acteurs principaux :

— **Le Client :** développeur d'une application basée LBS ayant des données sensibles d'utilisateurs.
— **L'administrateur :** acteur qui propose la solution de sécurité et de son implémentation.

Après avoir identifié les acteurs, il est important de préciser les objectifs visés dans la réalisation de ce projet :

Objectifs -Niveau 1-	**Objectifs -Niveau 2-**	**Cas d'utilisation**
Sécurisation de données	Sécurisation de données d'application	- cas d'étude - Capture des données sensibles - Application de la sécurité
	Anonymisation de dataset	- Exploration et localisation des données sensibles - Appliquer l'algorithme d'anonymisation
	Protection des usagers mobiles	- Analyse du type de système de localisation d'application LBS - Sécurisation du système de localisation

TABLE 3.1 – Objectifs et cas d'utilisation de la sécurisation des données mobiles

3.3 Conception du système

Dans la section précédente, nous avons analysé les besoins des clients vis-à-vis des solutions de sécurité proposées. Il faut donc résoudre le problème de stockage de données

des applications mobiles LBS concernant les données sensibles des utilisateurs, de sécurité des données de base de données au moment du partage d'information et finalement d'exposition de la vie privée des utilisateurs.

3.3.1 Diagramme de cas d'utilisation

Le premier diagramme de cas d'utilisation 3.3 met en avant le type de client à qui notre solution de sécurité a à faire, c'est-à-dire, au développeur d'application LBS qui veulent, soit sécuriser leur application et les données stockés, ou alors sécuriser leur base de données et protéger les informations de leur clients.

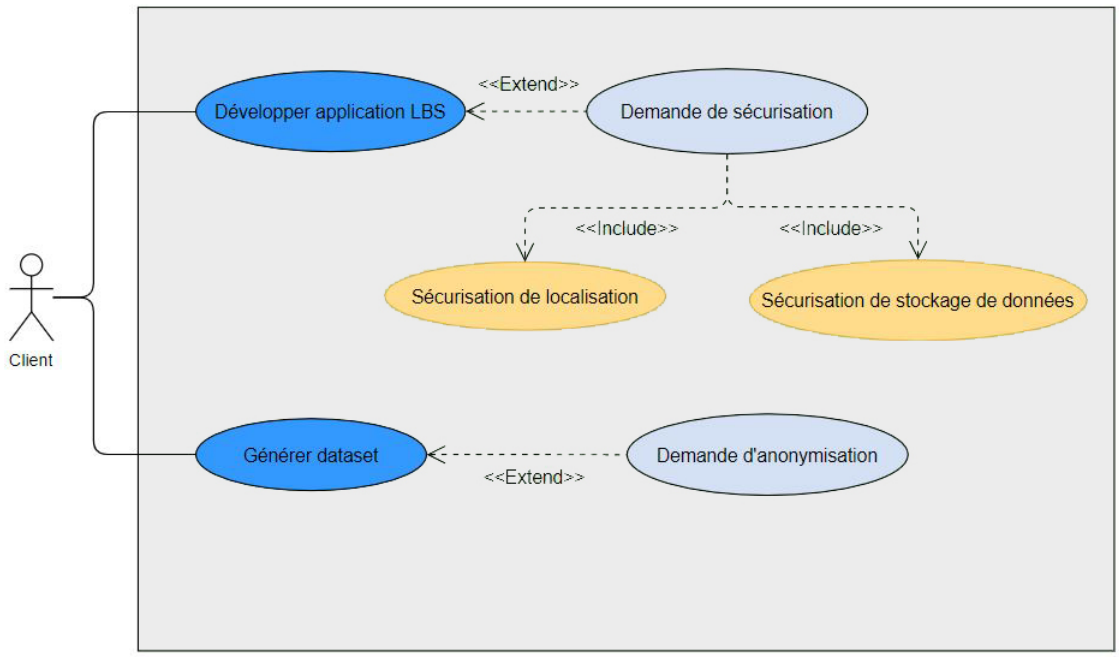

FIGURE 3.3 – Diagramme de cas d'utilisation du client

Le diagramme de cas d'utilisation qui suit décrit comment se déroulera l'application de chaque technique de sécurité à travers différentes étapes ; l'analyse des besoins du client, le data mining et l'application de la solution 3.4 :

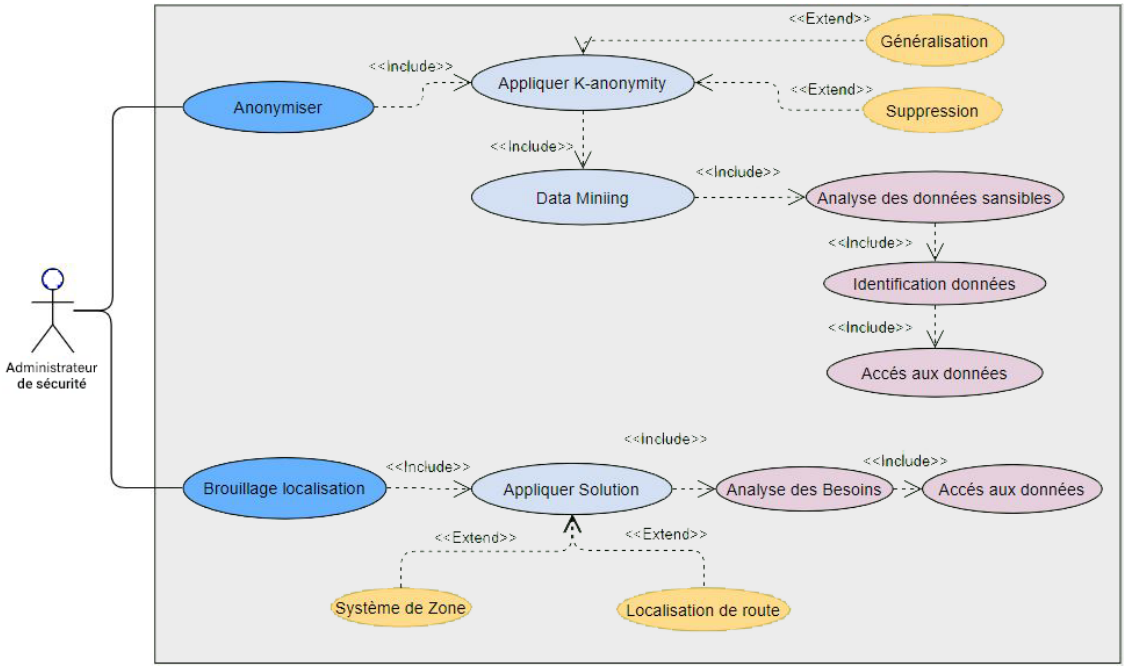

FIGURE 3.4 – Diagramme de cas d'utilisation de l'administrateur de sécurité

3.3.2 Diagramme d'activité

Le diagramme d'activité 3.5 qui suit décrit les opérations possibles par l'administrateur dans le cadre d'une application mobile basée sur LBS et dont la vie privée de l'utilisateur doit être protégée (il est à noté qu'il est important de réétudier les besoins du client et la ré-application du data mining en cas de désaccord avec client par rapport au résultat obtenu)

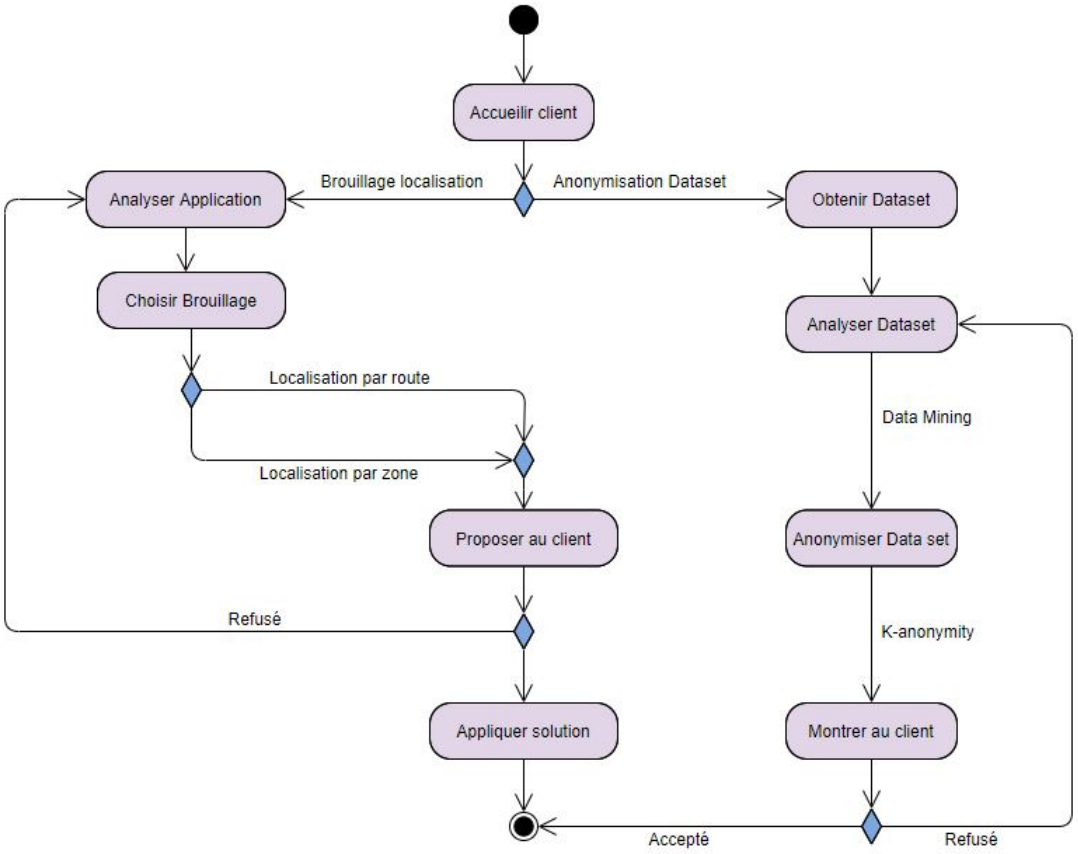

FIGURE 3.5 – Diagramme d'activité de l'administrateur de sécurité

3.3.3 Aperçu général de notre solution de sécurité

Pour remédier aux problèmes dans le cadre d'une application mobile LBS, nous nous sommes basés sur une combinaison d'un ensemble d'approches et solutions décrites dans les chapitres précédent, à savoir, l'anonymisation de données à travers l'algorithme k-anonymity décrit à la section 2.6.4, la protection des données des utilisateurs par le biais d'un cloud-mobile de stockage décrit à la section 2.6.3 et enfin le brouillage de localisation avec diverses techniques de sécurité (décrites à la section 2.6.1). La figure 3.6 illustre notre approche de protection des données des utilisateurs d'une application mobile basée sur la localisation.

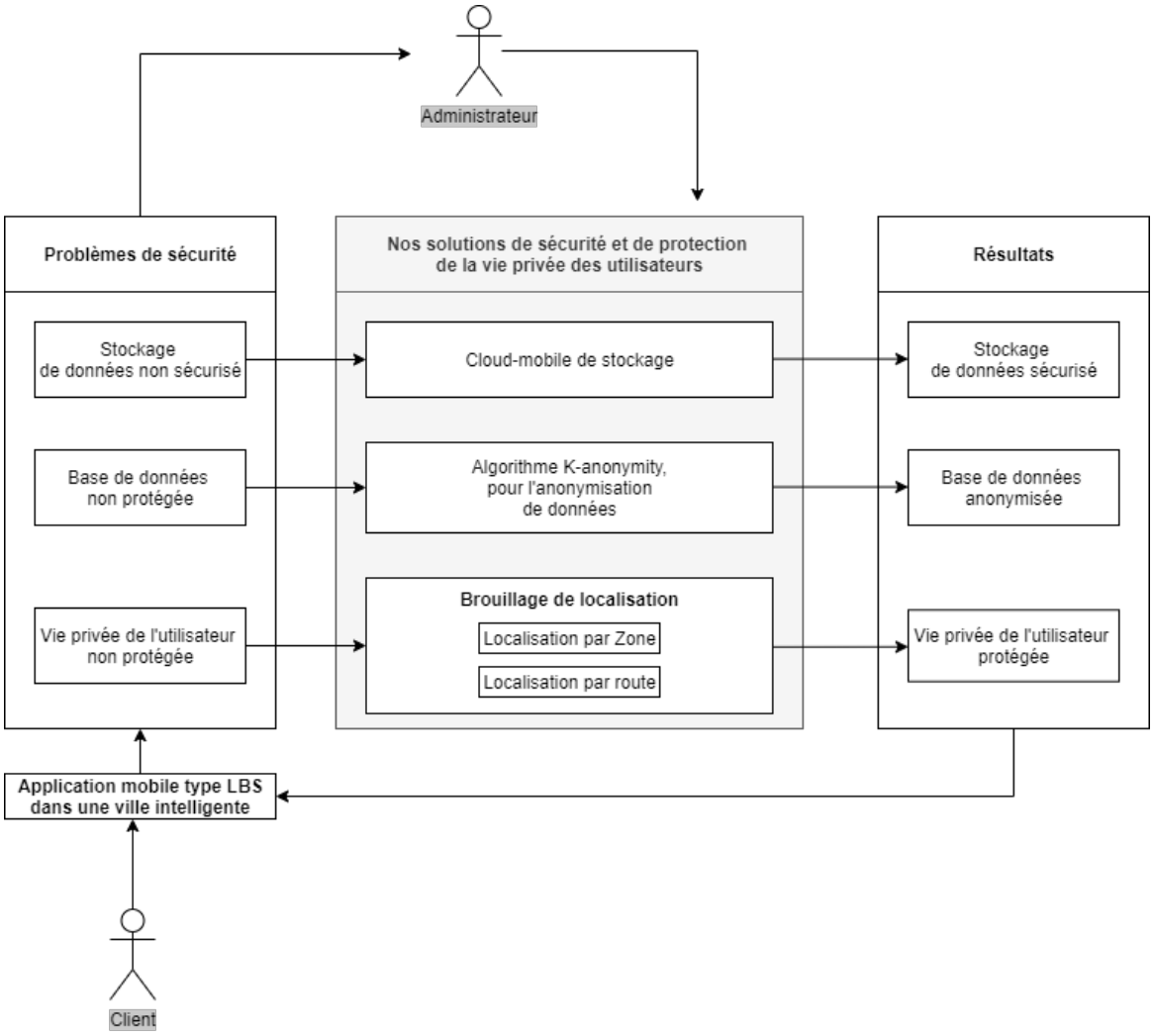

FIGURE 3.6 – Aperçu de notre approche de sécurité pour une application mobile basée sur la géo-localisation

3.3.4 Anonymisation de données

Pour résoudre le problème de préservation de vie privée des utilisateurs lors de partages de données, nous avons choisi l'algorithme K-anonymity, un algorithme d'anonymisation pour ne garder que les données utiles d'un dataset en éliminant les données sensibles. Afin de réaliser une telle opération, le travail sera divisé en plusieurs étapes comme dans le diagramme 3.4.

Analyse de la base de données

Cette étape est considérée comme particulièrement importante et ne peut être négligée pour la sécurité des informations des clients. L'administrateur de sécurité devra faire une analyse approfondie des informations en question afin de souligner les changements qui devront être effectués.

Dans ce qui suit, nous allons prendre un exemple de dataset qui sera analysé de prés afin de bien mettre en avant la technique exercé dans l'application de la solution :

ID	Prénom	Nom	Age	Zipcode	Diagnostic
1	Rodrick	Pittman	28	13053	Maladie cardiaque
2	Hubert	Austin	29	13068	Maladie cardiaque
3	Elwood	Frederick	21	13068	infection virale
4	Marcus	Nguyen	23	13053	infection virale
5	Alexis	Aguilar	50	14853	Cancer
6	Derrick	Hendrix	55	14853	Maladie cardiaque
7	Tamika	Farmer	47	14850	infection virale
8	Stanford	Chen	49	14850	infection virale
9	Brian	Dorsey	31	13053	Cancer
10	Minnie	Mitchel	37	13053	Cancer
11	Nestor	Everett	36	13222	Cancer
12	Cherie	Quinn	35	13068	Cancer

TABLE 3.2 – Dataset d'un hopital fictif sujet d'étude

Pour cette étude, nous avons pris un dataset d'un hôpital fictif 3.2 qui contient toutes les informations de ses patients. Le formulaire qu'un patient doit remplir pour son admission à l'établissement est un formulaire qui contiendra le nom, le prénom, l'âge et le zipcode, étant des informations basiques mais importantes pour cette hôpital. Par ailleurs, le patient qui vient de s'inscrire devra faire une première consultation chez le médecin afin d'obtenir un diagnostic médical initial et l'ajouter également dans les informations du patient en question. Supposons que l'hôpital veut procéder à des statistiques par une seconde partie, elle voudra donc envoyer son dataset après l'avoir sécurisé.

C'est l'administrateur d'anonymisation des données qui effectue cette analyse qui pourrait s'avérer de plus en plus compliquée selon la complexité de la base de données. Il faut, néanmoins, prendre en compte les données des patients à ne pas anonymiser (les données qui vont être partagées par le propriétaire de la base de données).

Le tableau précédent laisse facilement à remarquer que l'ID (étant une information suffisante pour identifier une personne), le nom et le prénom doivent obligatoirement être protégé pour la sécurité du patient (tableau de la figure 3.3) :

ID	Prénom	Nom	Age	Zipcode	Diagnostic
1	Rodrick	Pittman	28	13053	Maladie cardiaque
2	Hubert	Austin	29	13068	Maladie cardiaque
3	Elwood	Frederick	21	13068	infection virale
4	Marcus	Nguyen	23	13053	infection virale
5	Alexis	Aguilar	50	14853	Cancer
6	Derrick	Hendrix	55	14853	Maladie cardiaque
7	Tamika	Farmer	47	14850	infection virale
8	Stanford	Chen	49	14850	infection virale
9	Brian	Dorsey	31	13053	Cancer
10	Minnie	Mitchel	37	13053	Cancer
11	Nestor	Everett	36	13222	Cancer
12	Cherie	Quinn	35	13068	Cancer

TABLE 3.3 – Tableau identifiant l'ID, le nom et le prénom des patients

L'âge, étant un champ trés important pour les deux parties (pour l'hopital et les statistiques), devrait être également protégé, cependant, le résultat manquerait de précision. Il doit être fait en sorte d'obtenir une tranche d'âge afin d'accorder des informations utiles mais qui ne permetterait pas de ré-identifier les patients en question, même cas pour le zipcode 3.4 :

ID	Prénom	Nom	Age	Zipcode	Diagnostic
1	Rodrick	Pittman	28	13053	Maladie cardiaque
2	Hubert	Austin	29	13068	Maladie cardiaque
3	Elwood	Frederick	21	13068	infection virale
4	Marcus	Nguyen	23	13053	infection virale
5	Alexis	Aguilar	50	14853	Cancer
6	Derrick	Hendrix	55	14853	Maladie cardiaque
7	Tamika	Farmer	47	14850	infection virale
8	Stanford	Chen	49	14850	infection virale
9	Brian	Dorsey	31	13053	Cancer
10	Minnie	Mitchel	37	13053	Cancer
11	Nestor	Everett	36	13222	Cancer
12	Cherie	Quinn	35	13068	Cancer

TABLE 3.4 – Tableau montrant les tranches d'âge et la partie commune du zipcode

Le dernier champ devra alors être gardé afin de procéder aux statistiques. La prochaine étape consistera à appliquer l'algorithme k-anonymity pour anonymiser les données identifiés sensibles.

K-Anonymity

Passons maintenant à l'application de l'algorithme. Il faut, avant toute chose, bien modéliser la procédure en considérant l'algorithme en question comme une boîte noire. Cela permet d'avoir une vision sur les paramètres en entrée et en sortie (figure 3.7) :

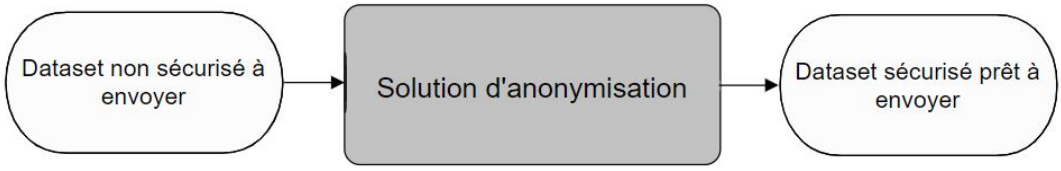

FIGURE 3.7 – Anonymiser les données sensibles

Nous allons maintenant considéré notre solution, c'est-à-dire, l'algorithme K-anonymity. L'algorithme 4 se compose de plusieurs concepts et procédures tels que :

— l**-diversity** : La l-diversité est une extension de la k-anonymité, ça signifie qu'il existe au moins l ligne(s) où la valeur des champs est différente.
— **Suppression** : ce concept consiste à supprimer la valeur d'un champ et la remplacer par une " * ".
— **Généralisation** : Procédure qui consiste à regrouper plusieurs valeurs d'une colonne dans un ensemble de valeurs.
— **Valeur formaté** : une valeur formaté est une valeur qui suit certaines règles de format comme par exemple le numéro de téléphone (exemple : +213 558 33 54 74) ou encore l'état civil (exemple : marié, divorcé ...).

Algorithme 4 : Algorithme de procédure d'anonymisation

```
Var :
    D : Dataset en entrée;
    K : Valeur de k-anounymity;
    L : Valeur de l-diversity;
Début
    lire(D);
    Dataminig(D);
    while (K n'est pas confirmé) do
        if (valeur non formaté) then
            Suppression();
        else
            Généralisation();
    while (L n'est pas confirmé) do
        if (valeur non formaté) then
            Suppression();
        else
            Généralisation();
Fin
```

Le tableau montré dans la partie précédente va être utilisé pour démontrer son déroulement afin d'anonymiser les données sensibles. Pour procéder à l'opération d'anonymisation, les informations sensibles non utilisables des patients tirées de l'analyse de la base de donnée vont être anonymisées (tableau 3.5) :

ID	Prénom	Nom	Age	Zipcode	Diagnostic
1	Rodrick	Pittman	28 [20-30]	13053	Maladie cardiaque
2	Hubert	Austin	29	13068	Maladie cardiaque
3	Elwood	Frederick	21	13068	infection virale
4	Marcus	Nguyen	23	13053	infection virale
5	Alexis	Aguilar	50 [40-60]	14853	Cancer
6	Derrick	Hendrix	55	14853	Maladie cardiaque
7	Tamika	Farmer	47	14850	infection virale
8	Stanford	Chen	49	14850	infection virale
9	Brian	Dorsey	31 [20-30]	13053	Cancer
10	Minnie	Mitchel	37	13053	Cancer
11	Nestor	Everett	36	13222	Cancer
12	Cherie	Quinn	35	13068	Cancer

TABLE 3.5 – Tableau d'identification des champs sensibles à anonymiser

Le plus important au cours de cette étape d'anonymisation est de s'assurer que le dataset présente le cas de L-diversité, un concept qui signifie qu'il y a au moins L champs

dans le dataset qui ne peuvent être identifié. Le tableau passera par une seconde analyse après l'anonymisation afin de vérifier si la L-diversité est bien respecté.

Maintenant que les données sensibles ont été repéré, il est possible de passer à l'étape de l'anonymisation. Comme il a été vu précédemment, il existe deux types d'anonymisation dans le K-anonymity :

La procédure est donc simple, remplacer l'ID, nom et prénom des patients par des " * " et l'âge et le zipcode par des ensembles de valeurs moins reconnaissables (tableau 3.6) :

ID	Prénom	Nom	Age	Zipcode	Diagnostic
*	*	*	[20-30]	13*	Maladie cardiaque
*	*	*	[20-30]	13*	Maladie cardiaque
*	*	*	[20-30]	13*	infection virale
*	*	*	[20-30]	13*	infection virale
*	*	*	[40-50]	14*	Cancer
*	*	*	[40-50]	14*	Maladie cardiaque
*	*	*	[40-50]	14*	infection virale
*	*	*	[40-50]	14*	infection virale
*	*	*	[30-40]	13*	Cancer
*	*	*	[30-40]	13*	Cancer
*	*	*	[30-40]	13*	Cancer
*	*	*	[30-40]	13*	Cancer

TABLE 3.6 – Dataset de l'hopital anonymisé

Analyse post-application

Le dataset anonymisé doit répondre au principe de la l-diversité, le sujet le plus important de l'algorithme. Le principe est simple ; en prenant n'importe quelle ligne du dataset, nous devrions obtenir l ligne qui contiennnent les mêmes valeurs. La différence devrait se remarquer uniquement au niveau du diagnostic médical. Pour ce faire, l'administrateur de la solution d'anonymisation prend, par exemple, la cinquième ligne du dataset, il est simple de remarquer que les 3 prochaines lignes contiennent les mêmes valeurs dans leurs champs mis à part le diagnostic. Cela signifie que si il essaye d'identifier une personne qui a un âge entre 40 et 50 ans, il ne pourra identifier son diagnostic médical car ces caractéristiques correspondent à 4 champs différents (tableau 3.7) :

ID	Prénom	Nom	Age	Zipcode	Diagnostic
*	*	*	[20-30]	13*	Maladie cardiaque
*	*	*	[20-30]	13*	Maladie cardiaque
*	*	*	[20-30]	13*	infection virale
*	*	*	[20-30]	13*	infection virale
*	*	*	[40-50]	14*	Cancer
*	*	*	[40-50]	14*	Maladie cardiaque
*	*	*	[40-50]	14*	infection virale
*	*	*	[40-50]	14*	infection virale
*	*	*	[30-40]	13*	Cancer
*	*	*	[30-40]	13*	Cancer
*	*	*	[30-40]	13*	Cancer
*	*	*	[30-40]	13*	Cancer

TABLE 3.7 – Tableau montrant les résultats de l'anonymisation

3.3.5 Le Cloud-mobile

Le cloud computing mobile utilise le cloud computing pour fournir des applications aux appareils mobiles. Ces applications mobiles peuvent être déployées à distance en utilisant des outils de vitesse, de flexibilité et de développement. Les applications cloud mobiles peuvent être créées ou révisées rapidement à l'aide des services cloud. Ils peuvent être livrés à de nombreux appareils différents avec différents systèmes d'exploitation, tâches informatiques et stockage de données. Ainsi, les utilisateurs peuvent accéder à des applications qui ne pourraient autrement pas être prises en charge. Quelques avantages dans l'utilisation de cloud-mobile :

— **Management :** Les professionnels de l'informatique ne disposent pas toujours des ressources nécessaires pour gérer les applications. Les fournisseurs de cloud contribuent à les aider à cette tâche.
— **Le partage de ressources :** Les applications mobiles qui s'exécutent sur le cloud ne sont pas contraintes par les ressources de stockage et de traitement d'un appareil. Les processus gourmands en données peuvent s'exécuter dans le cloud.
— **Sécurité :** La protection des données confidentielles est une préoccupation à tous les niveaux : pour les utilisateurs, les appareils et en ce qui concerne l'intégration dans d'autres systèmes.

Il en existe bien d'autres, mais nous avons essayé de nous focaliser sur les plus importants d'entre eux.

3.3.6 Brouillage et localisation

Ces techniques de sécurité concerne les applications LBS (applications qui apportent des services basé sur la localisation de l'utilisateur). Il existe plusieurs techniques. Pour bien les expliquer, nous allons considéré une application quelconque avec un système de localisation. Quand l'utilisateur donnera l'autorisation d'accès au GPS, sa localisation sera présentée comme illustré sur la figure 3.8 :

FIGURE 3.8 – Localisation exacte d'un utilisateur dans la ville de Hadjout, Tipaza

Le but est de sécuriser cette localisation tout en permettant l'offre de service. Pour cela, nous avons étudié deux techniques : la localisation par zone et la localisation par route.

Localisation par zone

Dans le cas de ville intelligente, généralement, les applications LBS obtiennnent les informations de localisation GPS de leurs clients seulement pour leur proposer des services, par exemple, à proximité de leur localisation actuelle ou autour de leur domicile. C'est pour cela qu'il est possible d'utiliser la première technique choisie pour ce projet qui est la localisation par zone afin de permettre à l'application d'accorder les services qu'elle propose tout en sécurisant la situation actuelle de l'utilisateur, comme le montre la figure 3.9) :

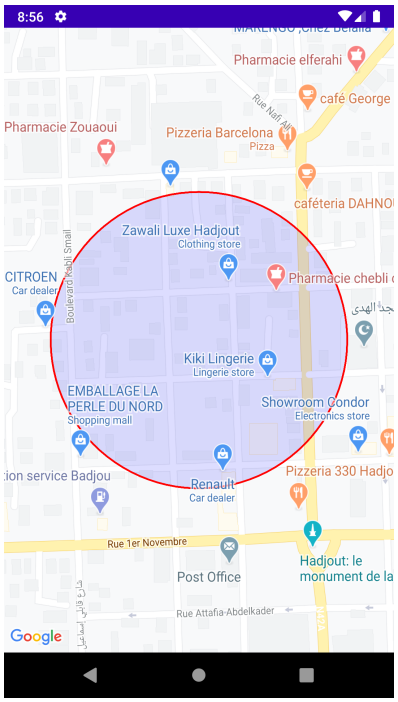

FIGURE 3.9 – Localisation sécurisé d'un utilisateur dans la ville de Hadjout, Tipaza

Grâce à cette technique, l'utilisateur de l'application est protégé contre les personnes malveillantes car sa position actuelle n'est pas communiqué. Néanmoins, il faut faire en sorte que l'utilisateur peut quand même profiter de l'offre de service de l'application en question. La précision de localisation sera réduite ou augmentée au dépend du service que l'application a à offrir en retour. Il sera alors difficile de savoir où se trouve exactement l'utilisateur, ce qui le satisfait lui et l'administrateur de l'application. Pour les applications qui localisent des points de localisation autour de l'utilisateur, cette solution est adéquate pour la protection de la vie privée des utilisateurs.

Localisation par route

Dans la technique de la localisation pas route, nous allons prendre le même exemple. Le contexte de l'application, par contre, change pour un autre et concerne maintenant de localiser les routes autour de l'utilisateur comme dans la figure 3.10 :

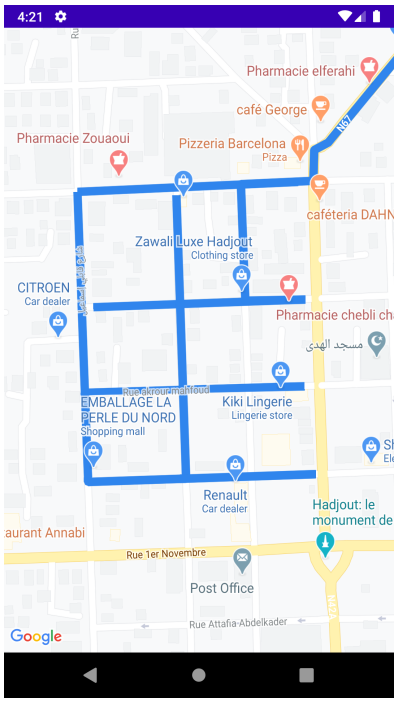

FIGURE 3.10 – Localisation de route à proximité de l'utilisateur dans la ville de Hadjout, Tipaza

Grâce à ce système de localisation par route, la localisation exacte de l'utilisateur n'est pas communiqué directement, mais plutôt les routes avoisinantes de son emplacement actuel. Cela permet de travailler sur des applications qui localisent les voies routières ou, pour encore aller plus loin, les évacuations d'eau ou d'électricité pour le signalement de panne par exemple.

3.4 Conclusion

Durant ce chapitre, nous avons, tout d'abord, étudier l'environnement de notre étude et explorer les besoins fonctionnels et non fonctionnels de notre clients à travers les schémas et les diagrammes concernant les différents acteurs. Nous avons visité, juste après, les techniques d'anonymisation de données, de stockage sécurisé et de protection de vie privée des citoyens des villes intelligentes et qui, par la suite, ont été choisi pour la résolution du problème de sécurité des applications mobiles basées LBS. Au cours du prochain chapitre, nous allons implémenté une application, analysé ses faiblesses et la sécuriser en appliquant notre solution pour la protection des utilisateurs.

Chapitre 4

Implémentation

4.1 Introduction

Au cours de ce chapitre, nous allons décrire notre implémentation de la solution proposée sur une application mobile. Ce chapitre va introduire, tout d'abord, le matériel, les outils et les logiciels utilisés pour la réalisation du projet. Par ailleurs, il va traiter également du développement d'une application mobile de type LBS et l'analyse de ses faiblesses afin de pouvoir la sécuriser. Ensuite, il va décrire la ou les solutions de sécurité choisies pour l'application mobile concernée et montrera les résultats obtenues après la mise en place de la solution.

4.2 Environnement de développement

Pour la procédure de développement de ce projet, nous avons travaillé avec deux ordinateurs portables :

— MSI, Windows 10 Pro, Intel® Core™ i7-4720HQ CPU @ 2.60 GHz 2.60 GHz, RAM 16Go, carte graphique NVidia GeForce GTX 965M.
— HP, Windows 8.1, Intel® Core™ i7-5500U CPU @ 2.40 GHz 2.40 GHz, RAM 8Go, carte graphique NVidia GeForce 820M.

Pour le test des résultats au cours de l'implémentation, deux smart-phones distincts ont été utilisé afin de s'assure du bon fonctionnement de la solution proposée :

— Samsung Galaxy S8+, Android version 9, Processeur Octuple coeur 4x2.3 GHz, RAM 6Go.
— OPPO A1K, Android version 9, Processeur Octuple coeur 2.0 GHz, RAM 2Go.

Les plateformes IDE utilisées durant ce projet :

— **Android Studio :** l'environnement de développement integrée (IDE) officiel du système d'exploitation Android de Google designé spécialement pour le développement d'application mobile en Java pour les fonctionnalités et XML pour le design. Il a été utilisé pour le développement et la sécurisation de l'application LBS développé au cours du projet.

— **FireBase :** Plateforme développé par Google. Elle permet de développer des applications web et mobiles. Peut être utilisé par exemple pour la construction de base de données. Services utilisés pour la mise en place de la base de données et l'authentification des utilisateurs de l'application que nous avons développé.
— **IntelliJ IDEA :** Plateforme IDE utilisé pour le développement de logiciels informatiques en langage Java. Exploité pour la lecture d'un logiciel de K-anonymity précédemment développé en 2004.

4.3 Base de données FireBase par Google

Firebase est un ensemble de services d'hébergement pour n'importe quel type d'application (Android, iOS, Java ...). Il propose d'héberger des bases de données en temps réel, du contenu, de l'authentification sociale (Google, Facebook, Twitter et Github), des notifications, ou encore différents services. Parmi les services les plus intéressants :

Base de données FireBase RealTime

La base de données Firebase Realtime est une base de données hébergée dans le cloud. Les données sont stockées au format JSON et synchronisées en temps réel avec chaque client connecté. En créant des applications Android ou autres, tous les clients partagent une instance de base de données en temps réel et reçoivent automatiquement des mises à jour avec les données les plus récentes. La sécurité peut être géré directement à partir de la plateforme de Firebase à travers des règles de sécurité pour définir quel utilisateur a accès à quelles données et comment la base de données doit être structuré. Les règles de sécurité sont enregistrées avec la base de données FireBase RealTime dans les serveurs de Google.

Authentification FireBase

L'authentification FireBase (figure 4.1 offre des services faciles à utiliser et des bibliothèques interface utilisateur prêtes pour authentifier les utilisateurs dans une application. Au moment de l'inscription, elle peut permettre l'authentification à travers différentes parties tierces (Google, Facebook, Twitter ou GitHub) et accorde à l'utilisateur un identifiant qui permet à la base de données de lui accorder l'accès aux données qui lui sont permis.

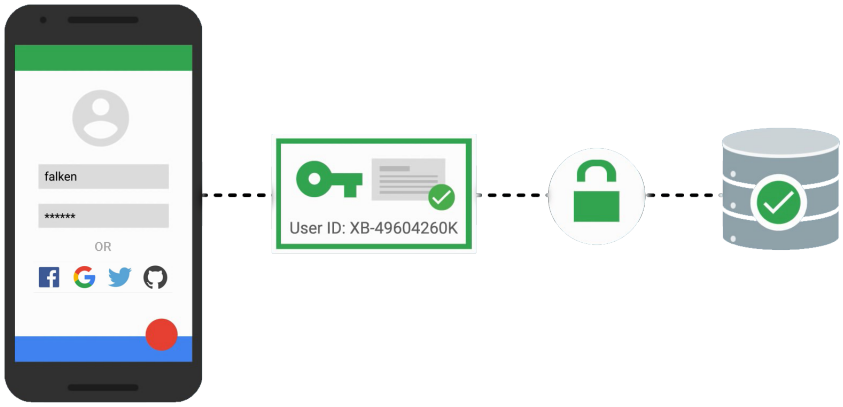

FIGURE 4.1 – Authentification FireBase [24]

Stockage Cloud FireBase

FireBase propose un cloud où les données d'application tels que les photos et les vidéos sont stockés en toute sécurité afin qu'elles puissent être partagés avec n'importe qui. Les transferts sont, bien évidemment, réalisés à travers une connexion sécurisée. Il propose aussi une architecture exa-octet, ce qui signifie que la quantité de fichiers qui peut être stockée augmente en conséquence pour qu'il y ait toujours suffisamment d'espace de stockage.

4.4 Description du cas d'étude NirBy : Implémentation d'une application LBS

Afin d'appliquer notre solution et bien mettre en avant les solutions étudiés au cours des chapitres précédents, nous avons développé une application que nous avons nommé NirBy (voir capture 4.2) pour réaliser les tests de sécurité.

4.4.1 Architecture de l'application

Cette application travaille avec le service basé sur la localisation (LBS). C'est une application mobile qui permet de situer dans la carte du monde, des lieux autour de la localisation actuelle de son utilisateur. Elle permet facilement de localiser les hôpitaux, les écoles, les lieux de prière ... etc. L'utilisateur devra d'abord s'enregistrer pour profiter de ces fonctionnalités.

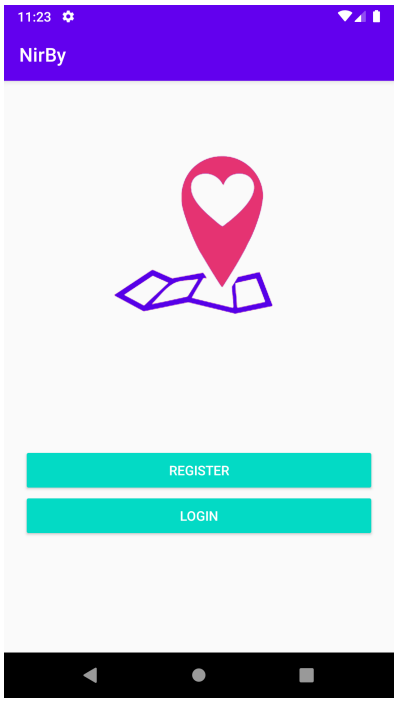

FIGURE 4.2 – Page d'accueil de l'application du cas d'étude

Base de données

Pour la mise en place de la base de données, la plateforme FireBase a été utilisé. FireBase propose tout un tas de fonctionnalités. Pour la base de données, nous avons utilisé une base de données FireBase RealTime qui se constitue d'un fichier JSON des utilisateurs de l'application qui se met à jour en temps réel. Le fichier JSON est organisé comme suit :

— La photo de couverture du profil.
— L'email.
— La photo de profil.
— Dernière localisation (latitude, longitude).
— Le nom.
— Le numéro de téléphone.
— Les préférences (dépend de la recherche effectué à travers l'application).
— L'identifiant.

Architecture globale de l'application

L'application se compose de plusieurs activités reliés entre elle, elle est simple à utiliser et facile à prendre en main. Au cours des lignes suivantes, nous allons détailler ses fonctionnalités qui nous intéressent le plus :

— **Inscription :** L'utilisateur peut créer un compte à travers une adresse mail valide et un mot de passe privée. Il pourra, dans ce cas, se connecter sur son profil à tout moment (voir figure 4.3).

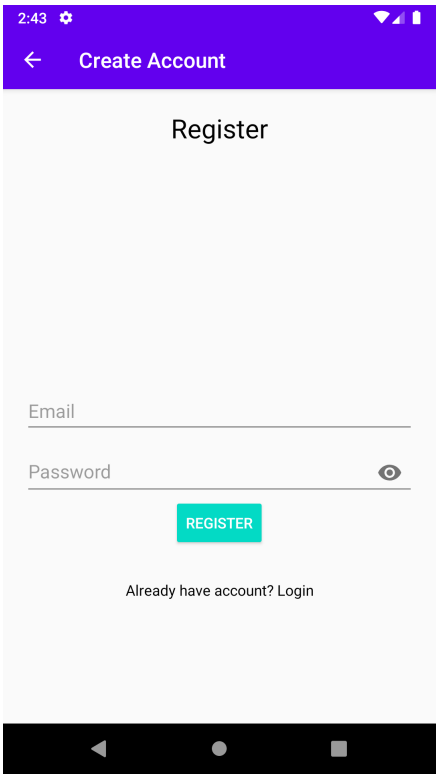

FIGURE 4.3 – Page d'inscription dans "NirBy"

— **Connexion :** L'application offre à l'utilisateur la possibilité de se connecter, soit, avec son compte Nirby qu'il a créé sur l'application ou alors son compte Gmail lié directement à l'application avec le service FireBase Authentication (voir figure 4.4).

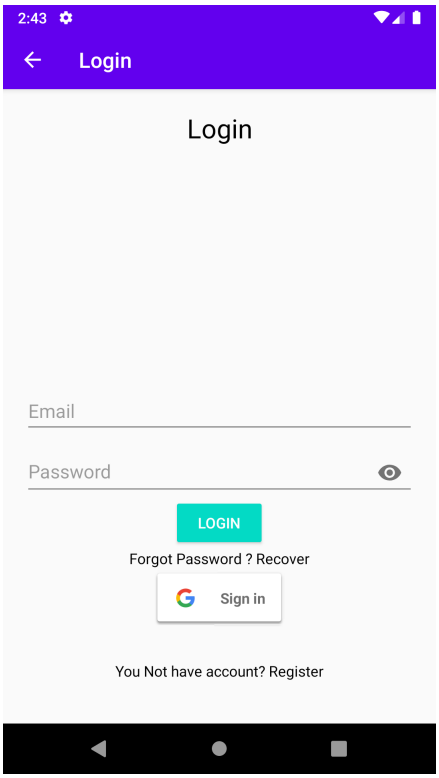

FIGURE 4.4 – Fenêtre de connexion dans "NirBy"

— **Consultation et modification du profil :** L'utilisateur peut consulter son profil, modifier sa photo de couverture et sa photo de profil. Au moment de l'ajout d'une nouvelle photo, elle sera ajouté au stockage Cloud de FireBase. Cela permet d'afficher le profil de l'utilisateur correctement même s'il se connecte dessus à partir d'un nouvel appareil (voir figure 4.5).

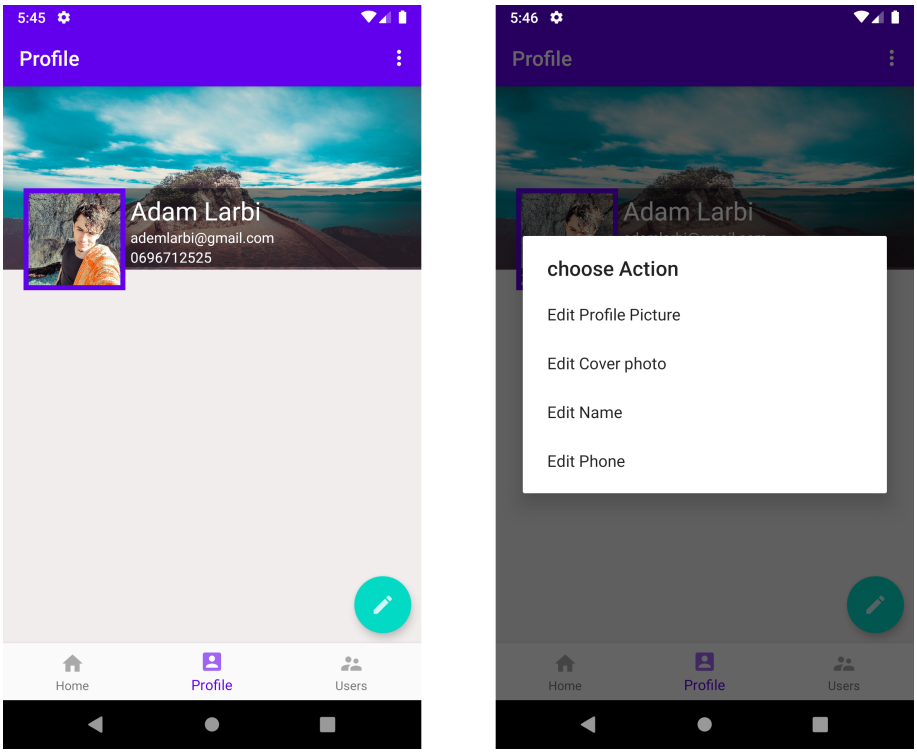

FIGURE 4.5 – Fenêtre de consultation (à gauche) et modification (à droite) du profil d'utilisateur

— **Recherche de lieux intéressants :** Pour profiter du service que l'application a à offrir, l'utilisateur devra sélectionner un type de lieu (centre de santé, lieu de restauration, école, ...etc) qu'il souhaite rejoindre à travers une liste de sélection. Ensuite, il devra appuyer sur le bouton "Show on map" qui lui permettra de consulter les lieux les plus proches de sa localisation exacte (voir figure 4.6).

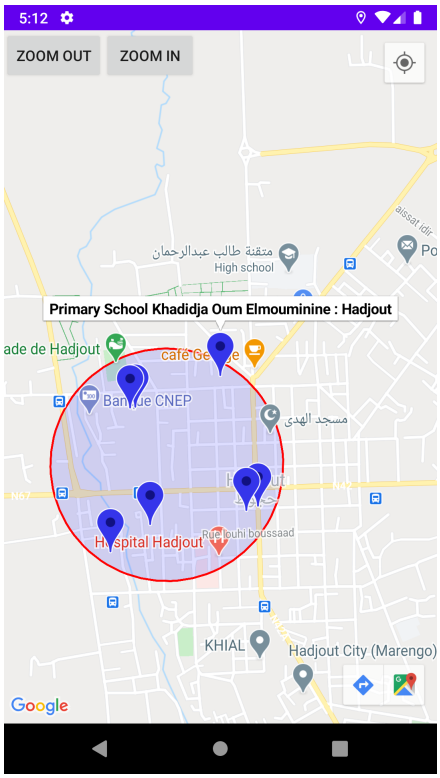

FIGURE 4.6 – Consultation des écoles à coté de l'utilisateur dans un rayon de 300 mètres

— **Consultation des utilisateurs de l'application :** L'application offre une interface où l'utilisateur peut voir tous les utilisateurs qui ont créé un compte dans l'application et font confiance à ses services. Cette fonctionnalité met en avant le système de partage d'informations de FireBase en affichant chaque utilisateur qui crée un nouveau compte dans l'application (voir figure 4.7).

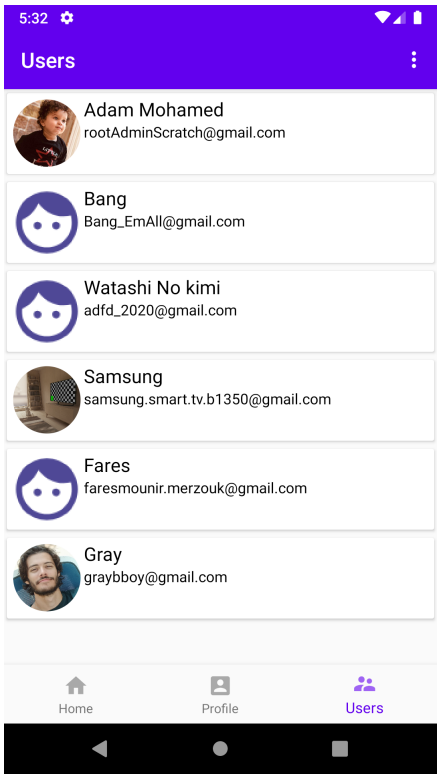

FIGURE 4.7 – Consultation des utilisateurs enregistrés dans l'application "NirBy"

4.5 Analyse de sécurité de l'application LBS et mise en avant de ses problèmes de sécurité

Le cas d'étude ci-dessus représente un cas typique d'application LBS au sein d'une ville intelligente. Pour le stockage de données des utilisateurs de l'application, nous avons utilisé, comme la plupart des développeurs débutant, une base de donnée mySQL connecté à l'application, ce qui s'avère, effectivement, un choix très médiocre par rapport à la sécurité et la confidentialité de données. Ensuite, on remarque que, concernant l'affichage des lieux autour de l'utilisateur, la sécurité reste à désirer. Certes, il est plus simple de situer les lieux par rapport à notre localisation d'un point A à un point B. Le problème demeure dans le fait que la localisation exacte de l'utilisateur est enregistré dans la base de données à chaque fois qu'il se connecte à l'application et cela risque de l'exposer inutilement. Pour éviter cela, il est important de la sécuriser pour protéger la vie privée des citoyens de la ville intelligente.

4.6 Application de notre solution de sécurité

Au cours de cette section, nous allons passer à la sécurisation de l'application LBS "NirBy" et afficher la différence post-application de la solution.

4.6.1 Mise en place de la base de données dans le cloud

Pour la sécurisation des données de notre base de données, nous avons - comme cité précédemment - utilisé le cloud-mobile de FireBase afin de stocker les données en toute

sécurité. Nous avons également utilisé le service FireBase Realtime Database pour stocker les données réelles de l'utilisateur afin de lui sauvegarder sa dernière utilisation de l'application.

Nous avons, par ailleurs, utilisé le service d'authentification de FireBase pour authentifier les utilisateurs dans NirBy et nous leur avons accordé la possibilité de s'authentifier avec leur compte Gmail pour une plus grande simplicité.

4.6.2 Anonymisation des données

Dans cette partie du travail, nous avons généré automatiquement un dataset qui pourrait nous aider au cours du développement et l'amélioration de l'application "NirBy". Ce dataset contient les informations des utilisateurs d'une application de recommandation d'évènements culturelles, artistiques ou musicales partout en Algérie. Les utilisateurs sont répertoriés de la façon suivante :

— L'identifiant de l'utilisateur (unique).
— L'e-mail.
— L'âge.
— Le sexe.
— Le numéro de téléphone.
— La latitude.
— La longitude.
— Le profil (suite de mots clé utilisés pour la localisation d'événements en relation avec les préférences des utilisateurs).

ID	Mail	Age	Sexe	N° Tel	Latitude	Longitude	Profil
202008292 10689949	teste_9 @ce-rist.com	19	Femme	+213 559 78 82 60	36.649428	3.167672	Ingenerie, Sante, Voyage, Bienfaisance, Religion, Ftness, Sante-mentale, Yoga, Medecine-naturelle, Escalade, Voyages, Randonnee, Visite, Camping, Environnement ...
202008311 14560891	teste_61 @ce-rist.com	20	Femme	+213 782 98 30 47	36.649470	3.167970	Droit, Religion, conference.
202008311 14568805	teste_32 @ce-rist.com	21	Femme	+213 559 09 70 48	36.645146	2.996532	Education, Commerce, Cinema, Bienfaisance, Religion, Science-Fiction, Action, Aventure, Crime, Animation, Super-Heros, Environnement, Nettoyage, concours, Bienfaisance
202008311 14658993	teste_100 @ce-rist.com	19	Homme	+213671 84 84 38	36.644681	2.994414	Sante, Commerce, Tourisme, Loisirs, Voyage, Bienfaisance, Religion, Jeux, Photographie
202008311 14687902	teste_65 @ce-rist.com	20	Femme	+213 561 52 46 50	36.646936	2.990459	Ingenerie, Bienfaisance, Pauverete
202008311 14859917	teste_70 @ce-rist.com	24	Femme	+213 782 51 16 33	36.514396	2.412761	Culture, Ingenerie
202008311 14915835	teste_42 @ce-rist.com	26	Femme	+213 561 75 82 79	36.513510	2.417878	Ingenierie, Music, Voyage, Religion, Chaabi, Classique, Visite, Camping, concours, cercl
202008311 15159856	teste_49 @ce-rist.com	20	Homme	+213 774 70 01 46	36.518765	2.418167	Ingenerie, Bienfaisance, Religion, Environnement, Nettoyage, concours, cercle, conference, charite
202008311 15535802	teste_31 @ce-rist.com	23	Femme	+213 541 01 52 15	36.730022	2.077191	Sante, Sante, Sante-mentale
202008311 15548719	teste_4 @ce-rist.com	25	Homme	+213 671 67 30 08	36.737821	2.070106	Culture, Ingenerie, Ingenerie, Music, Cinema, Sport, Rai, Classique, Traditionnel, Jazz, Mystere, Historique, Football
202008311 15502800	teste_30 @ce-rist.com	27	Homme	+213 555 65 30 84	36.733504	2.072602	Ingenierie, Music, Voyage, Fetes, Rai, Traditionnel, Visite, Randonnee, Camping

TABLE 4.1 – Dataset généré automatiquement

La première chose qu'on remarque dans un dataset comme celui-la est la précision avec laquelle les utilisateurs sont décrits. Entre de mauvaises mains, ce dataset pourrait causer des dégâts critiques aux utilisateurs et à leurs données sensibles. Dans une ville intelligente, il est plus primordial de gérer ce genre de risques afin d'instaurer une relation de confiance avec les citoyens. Afin de réaliser une telle tâche, nous avons utilisé notre solution de k-anonymity pour la protection des données des utilisateurs tout en gardant un dataset final exploitable pour l'obtention de résultats. Par rapport à notre cas d'étude, les données dont on a le plus besoin sont la localisation des utilisateurs, leur tranche d'âge, le sexe et leur profil, les champs restants pourront être anonymisés pour la protection des données sensibles des utilisateurs.

Nous allons donc procéder, d'abord, à la suppression de l'identifiant (sachant qu'il est unique), l'e-mail et le numéro de téléphone. Pour l'âge, nous comptons généraliser les champs et créer des tranches d'âge comme telles : de 0 à 17 ans, de 18 à 34, de 35 à 59 et enfin de 60 et plus. Nous avons laissé le sexe et le profil comme ils sont, vu leur importance dans le partage de données. Quant à la localisation actuelle des utilisateurs, nous avons décidé d'utiliser la généralisation. Le résultat obtenu sera donc représenté par une immense grille où les utilisateurs pourront se localiser dans les points de croisements de traits horizontaux et de traits verticaux.

Le tableau suivant montre une portion du résultat de l'anonymisation du dataset de

l'application de recommandation.

ID	Mail	Age	Sexe	N° Tel	Latitude	Longitude	Profil
*	*	18-34	Femme	*	36.64	3.16	Ingenerie, Sante, Voyage, Bienfaisance, Religion, Ftness, Sante-mentale, Yoga, Medecine-naturelle, Escalade, Voyages, Randonnee, Visite, Camping, Environnement ...
*	*	18-34	Femme	*	36.64	3.16	Droit, Religion, conference.
*	*	18-34	Femme	*	36.64	2.99	Education, Commerce, Cinema, Bienfaisance, Religion, Science-Fiction, Action, Aventure, Crime, Animation, Super-Heros, Environnement, Nettoyage, concours, Bienfaisance
*	*	18-34	Homme	*	36.64	2.99	Sante, Commerce, Tourisme, Loisirs, Voyage, Bienfaisance, Religion, Jeux, Photographie
*	*	18-34	Femme	*	36.64	2.99	Ingenerie, Bienfaisance, Pauverete
*	*	18-34	Femme	*	36.51	2.41	Culture, Ingenerie
*	*	18-34	Femme	*	36.51	2.41	Ingenierie, Music, Voyage, Religion, Chaabi, Classique, Visite, Camping, concours, cercl
*	*	18-34	Homme	*	36.51	2.41	Ingenerie, Bienfaisance, Religion, Environnement, Nettoyage, concours, cercle, conference, charite
*	*	18-34	Femme	*	36.73	2.07	Sante, Sante, Sante-mentale
*	*	18-34	Homme	*	36.73	2.07	Culture, Ingenerie, Ingenerie, Music, Cinema, Sport, Rai, Classique, Traditionnel, Jazz, Mystere, Historique, Football
*	*	18-34	Homme	*	36.73	2.07	Ingenierie, Music, Voyage, Fetes, Rai, Traditionnel, Visite, Randonnee, Camping

TABLE 4.2 – Dataset anonymisé

4.6.3 Brouillage et localisation par zone

Nous allons passer maintenant à la sécurisation du système de localisation de l'application LBS avec la technique de localisation par zone (cité dans la section 3.3.6). Le code source de base de l'application LBS "NirBy" localise les utilisateurs, dessine un cercle autour d'eux sur la map avec un rayon de 300 mètres (fixé par notre code source) pour localiser les endroits intéressants. Le listing 4.8 indique une partition du code source par rapport au système de localisation avant l'application de notre solution de sécurité. Notre objectif est de, premièrement, brouiller la localisation exacte de l'utilisateur et, deuxièmement, déplacer le centre du cercle pour rendre difficile la localisation des utilisateurs tout en s'assurant que l'utilisateur reste toujours à une distance au rayon du nouveau cercle généré par la solution.

```
LatLng currLatLng = new LatLng(currrentLocation.getLatitude()
                                , currrentLocation.getLongitude());

// set current location on cloud

HashMap < String, Object > result = new HashMap < > ();
result.put("lastLoaction", currLatLng);
databaseReference.child(user.getUid()).updateChildren(result)
.addOnSuccessListener(new OnSuccessListener < Void > ()
{
    @Override
    public void onSuccess(Void aVoid) {}
})
.addOnFailureListener(new OnFailureListener()
{
    @Override
    public void onFailure(@NonNull Exception e) {}
});

showNearBy(currLatLng);

MarkerOptions markerOptions = new MarkerOptions()
                                .position(currLatLng)
                                .title("I am Here");
mMap.addMarker(markerOptions);

setCircelCenter(currLatLng);
mMap.addCircle(new CircleOptions()
                .center(currLatLng)
                .radius(300)
                .strokeColor(Color.RED)
                .fillColor(0x220000FF)
                .strokeWidth(5));
mMap.moveCamera(CameraUpdateFactory.newLatLngZoom(currLatLng, 15));
```

FIGURE 4.8 – Code source localisant l'utilisateur avec sa localisation exacte dans l'application "Nirby"

Pour appliquer notre solution de brouillage de la localisation, nous avons considéré le paramètre du rayon de la terre (R) et l'avons estimé à 6378 kilomètres (valeur approximative). La localisation actuelle de l'utilisateur a été représenté en lattitude (LT) et longitude (LG). Par la suite, nous avons généré deux nombres aléatoires, dx et dy entre le décalage maximum (300 mètres) et le décalage minimum (-300 mètres). Ensuite, nous avons calculé la latitude (LTD) et longitude (LGD) du décalage (la partie du code source 4.9 et équations 4.1 et 4.2) :

$$LTD = LT + \frac{dy}{R} \times \frac{\Pi}{180} \tag{4.1}$$

$$LGD = LG + \frac{\frac{dx}{R} \times \frac{\Pi}{180}}{\cos(LT \times \frac{\Pi}{180})} \tag{4.2}$$

```
LatLng currLatLng = new LatLng(currrentLocation.getLatitude()
                               ,currrentLocation.getLongitude());
double latitude = currLatLng.latitude;
double longitude = currLatLng.longitude;
double decalage = 0.3 ; // in kilometer

float r_earth = 6378; // r_earth is approximately 6378 km.

double max = decalage * 1000;
double min = decalage * (-1000);
double dy = (Math.random() * (max - min + 1) + min) / 1000;
double dx = (Math.random() * (max - min + 1) + min) / 1000;

double diag = Math.sqrt(Math.pow(decalage, 2) * 2);
double new_latitude  = latitude  + (dy / r_earth) * (180 / pi);
double new_longitude = longitude + (dx / r_earth) * (180 / pi)
                                 / Math.cos(latitude * Math.PI / 180);

LatLng newLatLng = new LatLng(new_latitude, new_longitude);

setPROXIMITY_RADIUS(diag * 1000); // in meter
setCircelCenter(newLatLng);

// set current location on cloud

HashMap<String, Object> result = new HashMap<>();
result.put("lastLoaction", newLatLng);
databaseReference.child(user.getUid()).updateChildren(result)
.addOnSuccessListener(new OnSuccessListener<Void>()
{
    @Override
    public void onSuccess(Void aVoid)
    {
    }
})
.addOnFailureListener(new OnFailureListener()
{
    @Override
    public void onFailure(@NonNull Exception e)
    {
    }
});

showNearBy(newLatLng);

mMap.addCircle(new CircleOptions()
                .center(newLatLng)
                .radius(diag * 1000)
                .strokeColor(Color.RED)
                .fillColor(0x220000FF)
                .strokeWidth(5));
mMap.moveCamera(CameraUpdateFactory.newLatLngZoom(newLatLng, 15));
```

FIGURE 4.9 – Code source de localisation brouillé de l'application "NirBy"

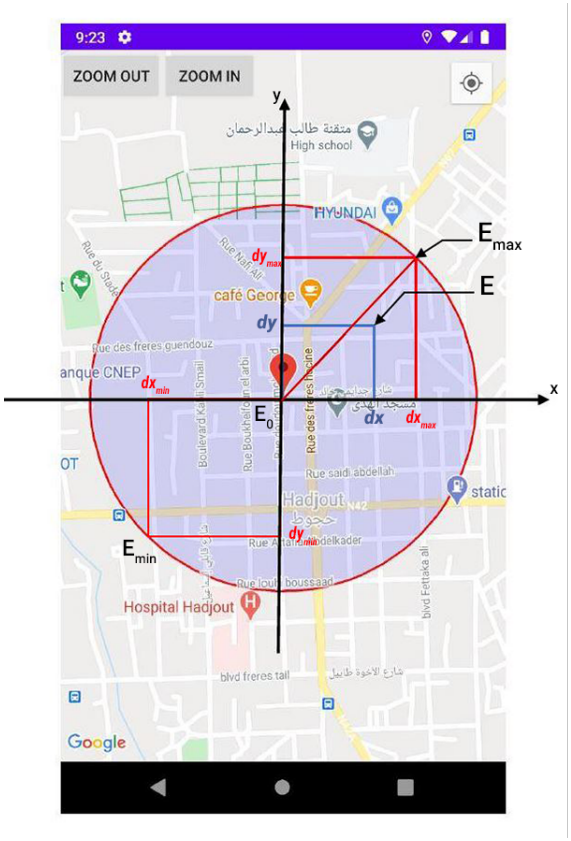

FIGURE 4.10 – Figure géométrique de système de décalage de la localisation

La figure 4.10 explique le résultat obtenue avec l'application de la solution, sachant que la valeur :

— E_0 **:** Représente la localisation réelle de l'utilisateur de l'application.

— E **:** Représente la nouvelle localisation de l'utilisateur (générée avec dx et dy aléatoires).

— E_{max} **:** la plus grande distance possible entre E_0 et E (pareil pour E_{min}).

Afin que E_0 soit toujours à l'intérieur du nouveau cercle, le rayon a été recalculé et représente maintenant la diagonale entre E_0 et E_{max} (voir l'équation 4.3, application du théorème de Pythagore).

$$distance(E_0, E_{max}) = \sqrt{(distance(E_0, dx_{max})^2 + distance(E_0, dy_{max})^2} \tag{4.3}$$

La figure 4.11 montre le résultat obtenue après l'application de la solution.

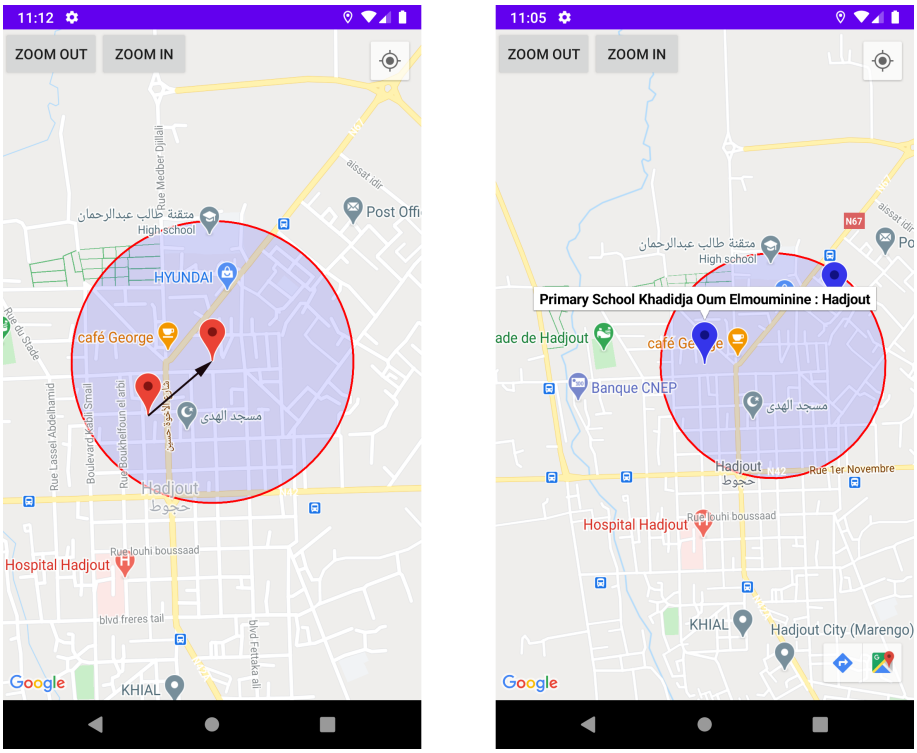

FIGURE 4.11 – Figure géométrique de système de décalage de la localisation

4.7 Conclusion

Dans ce chapitre, nous avons présenté l'application LBS qui a servi de cas d'étude et mis en avant ces différentes fonctionnalités, la composition de sa base de données et son offre de service à travers la localisation. Nous avons, par la suite, étudié ses faiblesses en terme de sécurité concernant la localisation de ses utilisateur et le stockage des données. Après cela, nous avons modifié l'application afin de la sécuriser et accorder la préservation de la vie privée des utilisateurs avec la protection de la localisation avec le système de localisation par zone. Dans la dernière partie du chapitre, un dataset non sécurisé a été pris en tant que cas d'étude et a été sécurisé avec l'algorithme K-anonymity pour protéger les données personnelles des utilisateurs, tout en gardant son utilité par rapport aux parties intéressées.

Conclusion générale

Le concept de ville intelligente est un domaine en cours de développement et qui, dans une décennie, peut être généralisé partout dans le monde. Dans la première partie de notre sujet, nous avons étudié les villes intelligentes, déterminer son architecture et ses différentes caractéristiques afin de déterminer l'univers de ce travail. Nous avons, par la suite, un deuxième chapitre qui introduit les smart-phones et met en avant les techniques de collecte de données des utilisateurs à travers différents services et applications mobiles. L'étude a révélé trois problèmes de sécurité : la sur-collecte de données, l'exposition de la vie privée des citoyens et, enfin, l'exposition de leurs données sensibles. Nous avons estimé plus qu'important de sécuriser ces trois problèmes de confidentialité et de sécurité afin d'assurer la sécurité des citoyens dans une ville intelligente. C'est pour cela que nous avons enchaîné, avec l'étude des solutions de sécurité des données sensibles des citoyens et de techniques de stockage sécurisé ultérieurement proposés. Cette recherche nous a permis de nous inspirer afin de proposer des solutions de sécurité pour ces problèmes de sur-collecte et de stockage de données sensibles.

Dans le chapitre conception, nous avons décrit avec précision : de la proposition de stockage dans un cloud-mobile pour le problème de sécurité des données, vers la sécurisation de la localisation des applications mobiles de type LBS pour le problème d'exposition de la vie privée des gens, en passant par l'algorithme de K-anonymity qui permet d'anonymiser les datasets pour garder les données des citoyens protégées.

Cette conception nous a permis d'entamer le chapitre d'implémentation au cours duquel nous avons développé une application LBS nommée NirBy et mis en avant quelques services et fonctionnalités. Nous l'avons développé avec une base de données sécurisée par un cloud-mobile que Google propose à travers sa plate-forme FireBase. Nous avons mis en avant les problèmes de sécurité cités ultérieurement dans l'application LBS NirBy et avons opté pour la meilleure solution à appliquer pour la sécurisation de la localisation. Nous avons, donc, implémenté la solution de localisation par zone et avons démontré les résultats obtenues. Pour finir, afin de démontrer les possibilités d'anonymisation de l'algorithme K-anonymity pour la sécurisation des données sensibles des utilisateurs, nous avons généré, automatiquement, un dataset contenant les données privées et la localisation de citoyens d'une ville intelligente fictif et l'avons sécurisé en protégeant les données privées et en généralisant la localisation afin que la vie privée des citoyens reste anonyme. Comme perspective de notre travail, nous souhaitons, dans un premier temps, appliquer notre solution sur des applications LBS opérationnelles avec des données réelles et effectuer une évaluation de notre solution mais aussi de travailler avec des données de datasets et procéder à leur anonymisation. Nous pensons également améliorer notre procédure d'anonymisation et utiliser d'autres algorithmes plus récents car l'algorithme K-anonymity se limite à un certain degré de complexité de dataset.

Références

[1] Yibin LI, Wenyun DAI, Zhong MING et Meikang QIU. « Privacy protection for preventing data over-collection in smart city ». In : *IEEE Transactions on ComputersZ* 65, 5, 2016, p. 1339-1350.

[2] Rida KHATOUN et Sherali ZEADALLY. « Cybersecurity and privacy solutions in smart cities ». In : *IEEE Communications Magazine* 55, 3, 2017, p. 51-59.

[3] LU TAN et NENG WANG. « Future internet : The Internet of Things ». In : *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*. T. 5. IEEE, 2010, p. 376-380.

[4] Lei CUI, Gang XIE, Youyang QU, Longxiang GAO et Yunyun YANG. « Security and privacy in smart cities : Challenges and opportunities ». In : *IEEE access* 6, 2018, p. 46134-46145.

[5] Marco CENTENARO, Lorenzo VANGELISTA, Andrea ZANELLA et Michele ZORZI. « Long-range communications in unlicensed bands : The rising stars in the IoT and smart city scenarios ». In : *IEEE Wireless Communications* 23, 5, 2016, p. 60-67.

[6] Taewoo NAM et Theresa A PARDO. « Conceptualizing smart city with dimensions of technology, people, and institutions ». In : *Proceedings of the 12th annual international digital government research conference : digital government innovation in challenging times*. Digital Government Research Center, 2011, p. 282-291.

[7] Trevor BRAUN, Benjamin CM FUNG, Farkhund IQBAL et Babar SHAH. « Security and privacy challenges in smart cities ». In : *Sustainable cities and society* 39, 2018, p. 499-507.

[8] Benjamin CM FUNG, Thomas TROJER, Patrick CK HUNG, Li XIONG, Khalil AL-HUSSAENI et Rachida DSSOULI. « Service-oriented architecture for high-dimensional private data mashup ». In : *IEEE Transactions on Services Computing* 5, 3, 2011, p. 373-386.

[9] Pierangela SAMARATI et Latanya SWEENEY. « Protecting privacy when disclosing information : k-anonymity and its enforcement through generalization and suppression ». In : 1998.

[10] Cynthia DWORK. « Differential privacy Proceedings of the 33rd International Conference on Automata, Languages and Programming. 1–12 ». In : *Google Scholar Google Scholar Digital Library Digital Library*, 2006.

[11] Mourjo SEN, Anuvabh DUTT, Shalabh AGARWAL et Asoke NATH. « Issues of privacy and security in the role of software in smart cities ». In : *2013 International Conference on Communication Systems and Network Technologies*. IEEE, 2013, p. 518-523.

[12] Adrien BARTOLI, Juan HERNÁNDEZ-SERRANO, Miguel SORIANO, Mischa DOHLER, Apostolos KOUNTOURIS et Dominique BARTHEL. « Security and privacy in your smart city ». In : *Proceedings of the Barcelona smart cities congress*. T. 292. 2011, p. 1-6.

[13] Sidra IJAZ, Munam Ali SHAH, Abid KHAN et Mansoor AHMED. « Smart cities : A survey on security concerns ». In : *International Journal of Advanced Computer Science and Applications* 7, 2, 2016, p. 612-625.

[14] Norman SADEH, Jason HONG, Lorrie CRANOR, Ian FETTE, Patrick KELLEY, Madhu PRABAKER et Jinghai RAO. « Understanding and capturing people's privacy policies in a mobile social networking application ». In : *Personal and Ubiquitous Computing* 13, 6, 2009, p. 401-412.

[15] Sunny CONSOLVO, Ian E SMITH, Tara MATTHEWS, Anthony LAMARCA, Jason TABERT et Pauline POWLEDGE. « Location disclosure to social relations : why, when, & what people want to share ». In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2005, p. 81-90.

[16] APPTHORITY. « App Reputation Report, An Appthority Publication ». In : 2014. www.cs.otago.ac.nz/cosc345/resources/AppReputationReportSummer14.pdf.

[17] Gang PAN, Guande QI, Wangsheng ZHANG, Shijian LI, Zhaohui WU et Laurence Tianruo YANG. « Trace analysis and mining for smart cities : issues, methods, and applications ». In : *IEEE Communications Magazine* 51, 6, 2013, p. 120-126.

[18] David J HAND et Niall M ADAMS. « Data Mining ». In : *Wiley StatsRef : Statistics Reference Online*, 2014, p. 1-7.

[19] Pablo A PÉREZ-MARTÍNEZ et Agusti SOLANAS. « W3-privacy : the three dimensions of user privacy in LBS ». In : *12th ACM Int'l. Symp. Mobile Ad Hoc Networking and Computing*. 2011.

[20] Roger DINGLEDINE, Nick MATHEWSON et Paul SYVERSON. *Tor : The second-generation onion router*. Rapp. tech. Naval Research Lab Washington DC, juin 2004.

[21] Masayuki ABE et Tatsuaki OKAMOTO. « Provably secure partially blind signatures ». In : *Annual International Cryptology Conference*. Springer, 2000, p. 271-286.

[22] Vipul GOYAL, Omkant PANDEY, Amit SAHAI et Brent WATERS. « Attribute-based encryption for fine-grained access control of encrypted data ». In : *Proceedings of the 13th ACM conference on Computer and communications security*. 2006, p. 89-98.

[23] Zhibin ZHOU et Dijiang HUANG. « Efficient and secure data storage operations for mobile cloud computing ». In : *2012 8th international conference on network and service management (cnsm) and 2012 workshop on systems virtualiztion management (svm)*. IEEE, 2012, p. 37-45.

[24] GOOGLE. « Firebase by google ». In : 2012. `https://firebase.google.com`.