UNIVERSITÉ DU QUÉBEC À MONTRÉAL

UTILISATION DE LA CHAÎNE DE BLOCS POUR UN PARTAGE RESPECTUEUX DE LA VIE PRIVÉE DE LA GÉOLOCALISATION

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN INFORMATIQUE

PAR

BOUDERBALA MEROUANE MOHAMED SMAINE

SEPTEMBRE 2025

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES ACRONYMES

ECDSA Elliptic Curve Digital Signature Algorithm

GPS Global Positioning System (Système de positionnement global)

OSM OpenStreetMap

RGPD Règlement Général sur la Protection des Données

TBPF Tolérance aux Pannes Byzantines Pratique (*Practical Byzantine Fault Tolerance*, PBFT)

PdT Preuve de Travail (*Proof of Work*, PoW)

PdE Preuve d'Enjeu (*Proof of Stake*, PoS)

TLS Transport Layer Security

Tor The Onion Router

Remerciements

En premier lieu, je désire exprimer ma gratitude la plus sincère envers Dieu, l'Exalté, pour les nombreuses grâces qu'il a bien voulu me dispenser tout au long de la rédaction de ce travail universitaire.

Je tiens à adresser mes remerciements les plus chaleureux à mes parents. Leur soutien indéfectible, leur amour sans conditions et leurs prières ont été les fondements de mon engagement et de ma résilience face aux défis académiques rencontrés durant cette période d'étude.

Je souhaite également exprimer ma profonde reconnaissance à mon directeur de recherche, Sébastien Gambs, dont l'expertise, la rigueur intellectuelle et l'accompagnement constant ont été des atouts inestimables dans la concrétisation de ce mémoire.

Mes remerciements vont aussi à Didem Demirag, dont l'assistance méthodologique et les conseils avisés m'ont été d'une aide précieuse. Sa contribution a été essentielle pour affiner mon analyse et enrichir le contenu de ce travail.

À tous ceux qui ont touché de près ou de loin à ce travail, que ce soit par leurs conseils, leur soutien ou simplement par leur présence bienveillante, je renouvelle ma gratitude.

RÉSUMÉ

À l'ère numérique, la géolocalisation est essentielle pour des domaines tels que l'urbanisme, les services d'urgence et la gestion des infrastructures. Cependant, son exploitation soulève d'importantes préoccupations en matière de confidentialité et de sécurité des données personnelles. Ce mémoire propose une solution visant à anonymiser et décentraliser les données de géolocalisation pour protéger les utilisateurs tout en maintenant leur utilité. La solution proposée repose sur Hyperledger Fabric, une plateforme chaîne de blocs autorisée, et intègre des identités temporaires, des mécanismes de consensus avancés et une anonymisation intrinsèque des données dès leur collecte. Pour renforcer la confidentialité, des outils comme le réseau Tor et des techniques d'injection de bruit sont intégrés. Une application Android dédiée et un système de visualisation permettent de convertir les transactions de géolocalisation en cartes thermiques et en analyses agrégées, garantissant une exploitation sécurisée des données. Les tests réalisés sur un prototype montrent que le système peut gérer de gros volumes de données tout en respectant les exigences de confidentialité. Les visualisations produites démontrent également l'efficacité de l'anonymisation et la cohérence des données simulées. Cette recherche se distingue par sa capacité à combiner confidentialité, sécurité et performance dans la gestion des données de géolocalisation, ouvrant ainsi de nouvelles perspectives pour des services respectueux de la vie privée à l'ère numérique.

Déclaration sur l'utilisation d'outils d'intelligence artificielle. Pour garantir une présentation claire et fluide de ce mémoire, j'ai utilisé ChatGPT, un outil d'intelligence artificielle développé par OpenAI, comme soutien linguistique. Cet outil a permis d'optimiser la lisibilité de certains passages, que j'avais initialement rédigés, en améliorant leur structure et leur formulation. Il est à noter que toutes les idées et contenus restent entièrement issus de mon travail personnel, et ChatGPT n'a été utilisé que pour des aspects rédactionnels.

CHAPITRE 1
INTRODUCTION

Avec la démocratisation des téléphones intelligents, la génération de données de géolocalisation connaît une croissance exponentielle. Selon un rapport de DOMO (2021), plus de 2,5 quintillions de bytes de données sont créés chaque jour, un chiffre qui souligne la vitesse vertigineuse à laquelle les informations numériques se multiplient à l'ère moderne (24). Ces informations, essentielles mais sensibles, posent un défi majeur en termes de protection de la vie privée, au-delà de leur utilisation dans les applications de navigation courantes, les données de géolocalisation jouent également un rôle crucial dans divers domaines comme l'urbanisme, la gestion des infrastructures et les services d'urgence. Par exemple, elles sont exploitées pour surveiller les flux de mobilité (40), optimiser le trafic (28) et même modéliser des scénarios d'évacuation en cas de catastrophe naturelle (14). Ces multiples usages rendent la gestion sécurisée et respectueuse de la vie privée de ces données encore plus primordiale.

En effet, une mauvaise gestion ou une fuite de ces données peut entraîner des violations significatives de la confidentialité des utilisateurs. Par exemple, une enquête du *New York Times* (65) a révélé en 2019 qu'un jeu de données rassemblant les déplacements de millions de téléphones portables, prétendument anonymisés, pouvait en réalité être recoupé pour identifier et suivre les individus à travers leurs trajets quotidiens. Ce type de détournement de la géolocalisation soulève d'importantes questions éthiques et de sécurité. Par ailleurs, même si elle ne portait pas spécifiquement sur la mobilité, l'affaire Cambridge Analytica (67) révélée en 2018 reste emblématique des abus liés à la collecte et à l'utilisation non autorisée de données personnelles. Dans ce scandale, 87 millions d'utilisateurs de Facebook ont vu leurs informations exploitées pour influencer des campagnes politiques, ce qui a mené à des réactions et des réglementations plus strictes, notamment le Règlement général sur la protection des données (RGPD) de l'UE.

Actuellement, la responsabilité de la collecte et de la sécurisation de ces données incombe principalement aux entreprises, un modèle qui présente des risques de centralisation et de mauvaise utilisation comme discuté par Thompson et Warzel (65). En particulier, les systèmes de gestion de données de localisation actuels centralisent ces informations et les associent à des identités uniques, une pratique qui facilite leur exploitation commerciale. Par exemple, comme discuté par Manjoo (43), cela signifie que les entreprises, comme les plateformes de réseaux sociaux et les services en ligne, collectent et stockent d'énormes quantités de

données personnelles. Cette centralisation des données rend ces entreprises des cibles attractives pour les cyberattaques et augmente le risque de violations de données. De plus, il existe un risque que ces données soient mal utilisées. Ainsi, les détaillants pourraient effectuer des analyses avancées pour comprendre les habitudes d'achat des consommateurs, en associant des données comportementales, y compris celles de localisation, à des identités d'utilisateur uniques. Cette pratique permet un ciblage publicitaire précis basé sur les itinéraires habituels et les lieux fréquemment visités par les consommateurs, augmentant ainsi l'efficacité des campagnes marketing (43).

De plus, cette centralisation soulève des questions éthiques et sécuritaires importantes, notamment en ce qui concerne la revente des données et le risque de fuites en cas de cyberattaques. *L'objectif de la recherche menée dans le cadre de ce mémoire est de proposer une solution visant à renforcer la protection de la vie privée par le biais de l'anonymisation des données de géolocalisation dès leur génération tout en s'assurant que la solution soit facile à utiliser*. Plus précisément, nous nous baserons sur la technologie des chaînes de blocs (*blockchain* en anglais) pour un stockage décentralisé, garantissant ainsi une intégrité élevée des données dans un registre publiquement disponible.

Cette approche cherche plus exactement à répondre à la problématique suivante : comment gérer efficacement les données de géolocalisation sans compromettre la vie privrée des utilisateurs. En anonymisant les données à la source, notre solution réduit le risque d'erreurs humaines et de manipulations malveillantes. Par exemple, dans les approches traditionnelles, les données sont souvent centralisées puis anonymisées *a posteriori* ce qui soulève des risques importants en cas de brèches de sécurité (69; 51). En intégrant l'anonymisation directement dans le processus de génération et de stockage des données, notre solution vise à assurer une protection de la confidentialité de bout-en-bout. De plus, le stockage décentralisé via la chaîne de blocs offre un cadre plus sécurisé, empêchant la centralisation des données par une seule entité et réduisant ainsi les risques de mauvaise gestion.

Pour répondre à ces objectifs, nous avons conçu un système de géolocalisation qui combine un stockage décentralisé, une anonymisation intrinsèque des données et l'exploitation de la chaîne de blocs pour garantir l'intégrité des données. Plus précisément, nous proposons un système décentralisé qui utilise des identités temporaires et le réseau Tor (64) pour protéger l'anonymat des utilisateurs, tout en garantissant la précision et la fiabilité des données de géolocalisation. Le système repose sur une infrastructure Hyperledger Fabric (34), conçue pour gérer efficacement les transactions et les données de localisation de manière sé-

curisée et évolutive. Afin de démontrer son applicabilité, nous avons aussi évalué la performance de notre système à travers des simulations et des tests dans des environnements contrôlés, démontrant ainsi sa capacité à fournir des services de géolocalisation précis tout en préservant la vie privée des utilisateurs.

Au-delà des considérations techniques et de la protection de la vie privée, notre solution ouvre la porte à une multitude de cas d'usage. En effet, la compréhension approfondie des mouvements de foule, permise par une gestion plus respectueuse des données de géolocalisation, pourrait par exemple être utilisée en urbanisme, pour analyser les mouvements de population afin d'aider à planifier et à développer des infrastructures adaptées aux besoins réels des citoyens (40). La gestion du trafic est un autre cas d'usage où notre solution pourrait apporter une valeur ajoutée. Ainsi, des applications existantes, telles que Waze, démontrent déjà l'intérêt d'exploiter les données de localisation pour optimiser les flux de circulation. Toutefois, notre approche, en se passant de l'identification directe des utilisateurs, renforcerait la confidentialité et l'acceptabilité de ces services.

Le plan du mémoire est le suivant. Tout d'abord, le chapitre 2 explore les principes de la technologie chaînes de blocs et son potentiel d'application pour l'amélioration des services de géolocalisation, décrivant ainsi le cadre théorique sur lequel se base la solution proposée. Par la suite, le chapitre 3 se focalise sur l'état de l'art sur les méthodes de collecte de la mobilité en décrivant les travaux existants faisant le pont entre la recherche actuelle et les fondations sur lesquelles cette recherche s'est construite telle que FOAM (26) et le concept de détection participative (*crowdsensing* en anglais). Ce faisant, ce chapitre offre un aperçu des tentatives antérieures pour traiter les défis liés à la géolocalisation précise et respectueuse de la vie privée et permet une comparaison avec l'approche proposée dans ce mémoire. Ensuite, le chapitre 4 décrit de manière détaillée l'architecture du système de géolocalisation décentralisé proposé. En particulier, il explicite les mécanismes de protection de la vie privée et les techniques d'optimisation de performance intégrées, clarifiant le fonctionnement et le rôle de chaque composant du système dans l'ensemble de la solution. Après cela, le chapitre 5 décrit le prototype développé pour valider l'approche proposée, détaillant la mise en œuvre du système, y compris la structure des données, le développement de l'application Android, et l'intégration avec le réseau Hyperledger Fabric. Cette section présente aussi l'évaluation de la performance du système en termes de précision de la géolocalisation, d'efficacité du traitement des données, et de protection de la confidentialité des utilisateurs, fournissant une mesure concrète de l'efficacité de la solution proposée. Enfin, le chapitre 6 conclut ce mémoire en décrivant les travaux futurs ainsi que les possibilités d'amélioration du système.

CHAPITRE 2
NOTIONS PRÉLIMINAIRES EN CHAÎNES DE BLOCS

Ce chapitre se concentre sur les fondements de la technologie des chaînes de blocs. Nous commencerons par présenter son évolution historique, puis nous décrirons les principaux mécanismes de fonctionnement. Enfin, nous proposerons une comparaison générationnelle afin de mettre en évidence les forces et les différences propres à chaque grande étape de cette technologie. Les aspects liés à la géolocalisation seront traités dans le chapitre suivant.

2.1 Historique de la technologie

Dans sa thèse de 1982 intitulée «Systèmes informatiques établis, maintenus et approuvés par des groupes mutuellement suspects», David Chaum a théorisé pour la première fois un protocole de type chaînes de blocs (31). Plus tard, en 1991, Haber et Stornetta ont décrit la technologie cryptographique pour créer un protocole de type chaînes de blocs. Le système proposé permettait des horodatages des documents ne pouvant pas être falsifiés. En 1992, Haber, Stornetta et Dave Bayer avaient amélioré ce système en intégrant des arbres de Merkle (voir Section 2.3] (11), augmentant ainsi considérablement l'efficacité en permettant de compiler plusieurs certificats de documents en un seul bloc. Un fait peu connu est que leur société Surety publie les hachages de ces certificats de documents dans l'hebdomadaire New York Times depuis 1995, étant ainsi la première chaîne de blocs de l'histoire (vic).

Il est intéressant d'avoir cette perspective sur la philosophie originelle de la technologie, qui cherchait à rendre des documents infalsifiables (48), notamment lorsqu'on voit que dans l'imaginaire collectif la genèse de la technologie est fortement liée a Bitcoin (49), qui fut une véritable révolution a été déclenchée avec la création du Bitcoin en 2008 par Satoshi Nakamoto, marquant non seulement la naissance de la première crypto-monnaie, mais aussi la première application à grande échelle de la technologie chaînes de blocs. Ce développement a introduit une nouvelle ère de système financier décentralisé (13), défiant les structures traditionnelles par son fonctionnement autonome sans autorité centrale.

2.2 Principe général

Une chaîne de blocs est un registre numérique décentralisé, distribué et souvent public (49). Cette technologie enregistre les transactions dans des blocs liés et sécurisés à l'aide de la cryptographie, assurant l'intégrité et l'immutabilité des données. Chaque bloc contient un hachage cryptographique du bloc précédent, un horodatage et les données transactionnelles, formant ainsi une chaîne continue et sécurisée. Par exemple, dans une transaction financière, si Alice envoie des bitcoins à Bob, cette transaction sera enregistrée dans un bloc contenant les informations de la transaction, l'horodatage précis de l'événement et un hachage cryptographique reliant ce bloc à tous les blocs précédents. Ce hachage garantit que toute modification apportée à un bloc affecterait l'ensemble des blocs suivants, rendant la manipulation extrêmement difficile et détectable. Cette structure empêche efficacement toute modification rétroactive sans altérer tous les blocs subséquents, garantissant ainsi la transparence et la fiabilité des données enregistrées.

Chaque participant du réseau possède une copie complète de la chaîne de blocs et peut vérifier et auditer les transactions de manière indépendante. Par exemple, dans le cas d'une transaction entre Alice et Bob, tout participant peut consulter et valider cette transaction sans avoir besoin de faire confiance à une autorité centrale. Ceci permet de réduire les coûts associés à l'audit et à la vérification, car les mécanismes cryptographiques intégrés assurent l'exactitude des données. Les données présentes dans la chaîne de blocs sont gérées de manière autonome à l'aide d'un réseau pair-à-pair et d'un serveur d'horodatage distribué. Les transactions sont authentifiées par un mécanisme de consensus distribué (18), où les participants (ou « nœuds ») du réseau vérifient et valident chaque transaction. Ce processus est alimenté par des incitations personnelles, telles que les récompenses pour les mineurs en cas de preuve de travail (*Proof Of Work* en anglais) ou les gains pour les validateurs en cas de preuve d'enjeu (*Proof Of Stake* en anglais). Ces mécanismes encouragent les participants à agir honnêtement pour le bénéfice du réseau.

La décentralisation et la transparence inhérentes à la technologie chaîne de blocs assurent que même si certains nœuds du réseau sont compromis, la majorité des nœuds honnêtes maintient l'intégrité et la sécurité du registre, ce qui favorise un environnement fiable et sécurisé pour les transactions et les enregistrements de données.

2.3 Arbre de Merkle et hachage de blocs

Les arbres de Merkle et le hachage de blocs (45) sont fondamentaux pour le fonctionnement de la chaîne de blocs pour garantir l'intégrité et la sécurité des données. Dans ces systèmes, les transactions au sein d'un bloc sont hachées et organisées dans un arbre de Merkle, une structure qui résume et permet de vérifier efficacement de grands ensembles de données (voir figure 3.1). Le hachage final, connu sous le nom de racine de Merkle, représente toutes les transactions et est stocké dans l'en-tête du bloc. Chaque bloc est également identifié par un hachage unique, généré à partir de l'en-tête du bloc, qui inclut la racine Merkle. Ce hachage de blocs relie chaque bloc à son prédécesseur, formant ainsi la chaîne de blocs. Toute modification dans une transaction modifie la racine Merkle et le hachage du bloc, signalant une falsification.

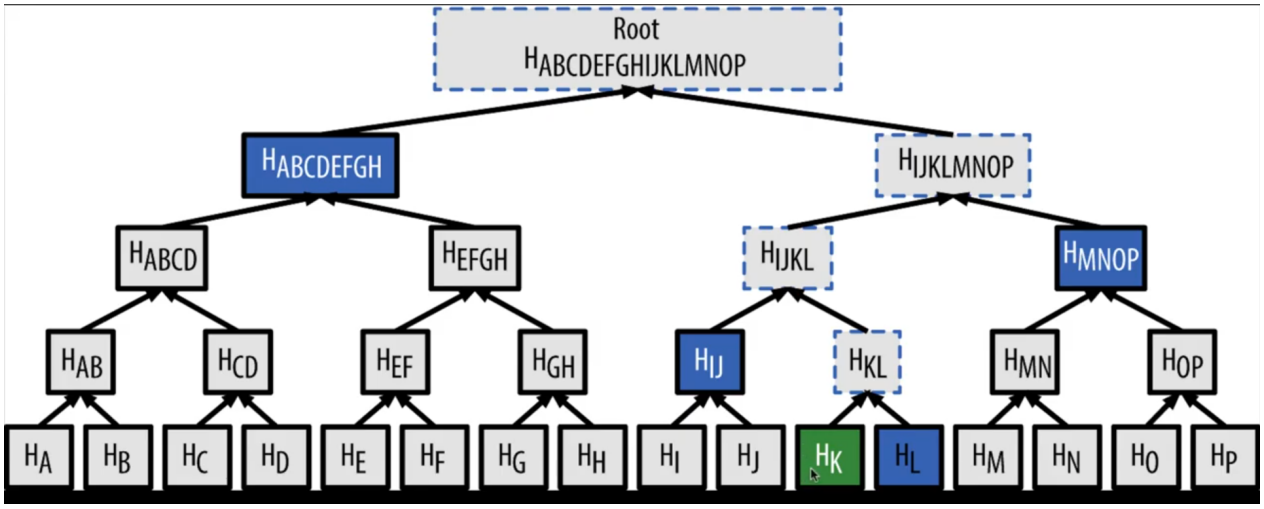

Figure 2.1 – Structure d'un arbre de Merkle. Image adaptée de (19).

Cette image représente un arbre de Merkle, où chaque feuille de l'arbre représente un hachage cryptographique d'un petit ensemble de données, comme une transaction. Plus précisément, chaque lettre (HA, HB, etc.) symbolise les hachages des données individuelles, formant les feuilles de l'arbre. Les combinaisons de ces hachages, comme HAB ou HCD, sont obtenues en concaténant deux hachages enfants et en calculant leur hachage commun. Le hachage final, HABCDEFGH, représente la combinaison de tous les hachages sous-jacents, et la racine (Root HABCDEFGHJKLM) est le hachage final de l'arbre, utilisé pour vérifier l'intégrité de toutes les données sous-jacentes dans l'arbre.

Pour illustrer ce qui est inclus dans les blocs, considérons un bloc typique dans une Chaîne de blocs comme Bitcoin. Chaque bloc contient :

— Un numéro de version pour suivre les modifications du protocole.

— La racine de Merkle, qui est un hachage de toutes les transactions dans le bloc.
— Le hachage du bloc précédent, qui relie ce bloc à la chaîne.
— Un horodatage, indiquant quand le bloc a été miné.
— La difficulté cible, qui est une valeur définissant l'objectif à atteindre pour que le hachage du bloc soit considéré comme valide. Elle détermine, par exemple dans Bitcoin, le nombre de zéros initiaux requis dans le hachage. Plus la difficulté est élevée, plus la recherche d'un hachage valide est coûteuse en calcul.
— Un nonce, qui est une valeur que les mineurs modifient à chaque tentative afin que le hachage de l'en-tête du bloc respecte la condition imposée par la difficulté. Autrement dit, le nonce correct est celui qui permet d'obtenir un hachage inférieur à la difficulté cible.
— Une liste de transactions, qui sont les enregistrements des mouvements de cryptomonnaie entre adresses.

Cette chaîne de hachages d'un bloc à l'autre garantit l'intégrité de la structure, la rendant immuable et sécurisée. Ce système permet une vérification rapide des transactions et une protection robuste contre la manipulation des données, essentielles au maintien de la confiance dans les systèmes basés sur la chaîne de blocs, d'autant plus qu'il se repose sur le fait que les réseaux sont décentralisés et contiennent un nombre important de copies du registre au sein de plusieurs nœuds différents.

2.4 Problème du général byzantin et mécanismes de consensus

Le problème du général byzantin est un problème classique en calcul distribué (39), particulièrement pertinent pour comprendre les mécanismes derrière la technologie des chaînes de blocs. Ce problème illustre les difficultés inhérentes à l'obtention d'un consensus dans un système informatique distribué doté de composants peu fiables. Dans sa formulation originale, ce problème est présenté comme un groupe de généraux, chacun commandant une partie de l'armée byzantine, devant se mettre d'accord sur un plan de bataille unifié. Cependant, ils ne peuvent communiquer que par messagers, de plus un ou plusieurs généraux pourraient être des traîtres essayant d'empêcher les généraux loyaux de parvenir à un accord. Le défi est de trouver un algorithme qui permet aux généraux fidèles de se mettre d'accord sur un plan, même en présence de traîtres.

Afin de traduire cela dans le contexte de la technologie, les nœuds du réseau représentent les généraux. Chaque nœud dispose d'une copie du registre et participe au processus de consensus. De la même ma-

nière que les généraux ne peuvent pas se fier à tous les messagers, les nœuds d'un réseau ne peuvent pas intrinsèquement faire confiance à toutes les informations reçues. Ainsi, certains nœuds peuvent être malveillants, avec l'intention de perturber le réseau ou de réaliser une double dépense d'une cryptomonnaie, ce qui était une préoccupation majeure lors de la première mise en œuvre grand public avec Bitcoin. Ces nœuds malveillants pourraient tenter de tricher en fournissant des informations divergentes à différents nœuds loyaux, créant ainsi des informations contradictoires dans le registre. Le cœur du problème des généraux byzantins dans les chaînes de blocs est d'atteindre un consensus (c'est-à-dire que tous les nœuds s'accordent sur l'état du registre) malgré la présence de ces nœuds malveillants. Cette uniformité du registre est essentielle au bon fonctionnement de la technologie. Les technologies des chaînes de blocs abordent ce défi grâce à différents protocoles de consensus (10), dont trois fondamentaux que nous détaillerons ci-après.

Tout d'abord, la *preuve de travail* (PdT) (49) est un mécanisme de consensus employé par des cryptomonnaies telles que Bitcoin. Il exige des participants, dénommés mineurs, de résoudre des puzzles cryptographiques complexes pour valider les transactions et créer de nouveaux blocs. Par exemple, un puzzle typique consiste à trouver un nonce, un nombre arbitraire, qui, lorsqu'il est combiné avec les données du bloc et haché à l'aide de l'algorithme SHA-256, produit un hachage commençant par un certain nombre de zéros. Supposons qu'un bloc contienne les données de transaction $T1, T2, T3$ et que son en-tête inclue un hachage précédent H_{prev}. Le mineur ajuste le nonce N jusqu'à ce que le résultat de la fonction $H(H_{prev} + T1 + T2 + T3 + N)$ commence par, par exemple, 4 zéros (0000...). Cela garantit la sécurité en rendant les modifications rétrospectives presque impossibles sans recalculer toute la chaîne. Ce processus assure la sécurité du réseau, mais est énergivore en raison de la puissance de calcul nécessaire pour résoudre les puzzles. La sécurité de la PdT repose sur le fait qu'altérer une partie de la chaîne de blocs nécessiterait de recalculer non seulement le bloc modifié, mais aussi tous les blocs subséquents, rendant la fraude peu pratique et donc un puissant moyen de dissuasion.

La *preuve d'enjeu* (PdE) (16) marque un progrès vers une efficacité énergétique et une durabilité environnementale. Dans la PdE, les validateurs sont sélectionnés pour créer un nouveau bloc basé sur la quantité de monnaie qu'ils possèdent et sont prêts à mettre en jeu. Par exemple, dans Ethereum 2.0 (15), les validateurs doivent verrouiller un minimum de 32 ETH dans un contrat intelligent pour participer. Lorsqu'un bloc est proposé, un comité de validateurs est sélectionné de manière pseudo-aléatoire. Chaque validateur du comité vote pour approuver ou rejeter le bloc. Si un validateur propose un bloc invalide ou agit de manière

malveillante, il risque de perdre une partie ou la totalité de sa mise (*slashing*). Cette méthode diminue la quantité d'effort de calcul nécessaire, puisque les validateurs sont choisis en fonction de critères tels que leur participation, plutôt que leur capacité à résoudre des puzzles. La PdE offre également une sécurité robuste, car les validateurs ayant un intérêt financier direct dans le réseau, un comportement malveillant peut entraîner la perte de leur mise, constituant ainsi une dissuasion financière contre la malhonnêteté.

La *tolérance pratique aux fautes byzantines* (TFBP) est un mécanisme de consensus développé pour pallier les limites de la tolérance aux fautes byzantines traditionnelle dans les réseaux distribués. Conçu pour être efficace dans des systèmes asynchrones tels que les chaînes de blocs, où le délai de livraison des messages n'est pas garanti, TFBP est idéal pour les chaînes de blocs privées où les identités des participants sont connues et fiables (voir section 2.5). Le processus de consensus en TFBP implique plusieurs étapes d'échanges de messages entre les nœuds pour vérifier l'authenticité et l'ordre des transactions. Un nœud est désigné comme le «primaire», les autres agissant comme «secondaires». Plus précisément, le primaire propose l'ordre des transactions, et les nœuds secondaires vérifient et approuvent cet ordre. La TFBP se distingue par sa capacité à offrir un haut débit et une faible latence dans le traitement des transactions, la rendant adaptée aux applications exigeant une gestion efficace et fiable des transactions.

Il convient de noter qu'au-delà de ces approches tolérantes aux fautes byzantines comme TFBP, d'autres types de protocoles de consensus existent et sont particulièrement utilisés dans le contexte des blockchains permissionnées telles qu'Hyperledger Fabric. Par exemple, des protocoles plus légers comme *Solo*, principalement employés lors des phases de développement, ou encore des protocoles basés sur l'élection d'un leader comme *Raft*, adoptés en production, seront décrits plus en détail dans la Section 2.8.

2.5 Types de chaînes de blocs

Les *chaînes de blocs publiques* sont des plateformes entièrement ouvertes et décentralisées, accessibles à quiconque souhaite y adhérer et y participer. Bitcoin et Ethereum (15) en sont des exemples notables. Elles fonctionnent sans autorité centrale, chaque participant ayant des droits égaux pour valider les transactions, participer au processus de consensus et consulter l'ensemble des transactions enregistrées. Ces systèmes se distinguent par leur sécurité et leur transparence, puisque toutes les transactions sont publiquement accessibles sur le réseau. Des mécanismes de consensus comme la PdT (49) ou PdE (16) sont utilisés pour assurer l'intégrité et la confiance au sein du réseau. Toutefois, ces réseaux peuvent souffrir de vitesses de transaction réduites et d'une consommation d'énergie importante, particulièrement ceux reposant sur la

PdT.

À l'opposé, les *chaînes de blocs privées* sont gérées par une entité unique. Hyperledger Fabric (7) et Corda (13) sont des exemples de tels systèmes. Hyperledger Fabric se distingue par sa modularité et son support pour des contrats intelligents complexes dans différents langages, tandis que Corda est conçu pour des transactions bilatérales optimisées, particulièrement adaptées aux applications financières où la confidentialité des transactions est primordiale. L'accès à ces réseaux est restreint et l'entité en charge contrôle le processus de consensus. Elles offrent ainsi une meilleure efficacité en termes de vitesse de transaction et d'évolutivité par rapport aux chaînes de blocs publiques, tout en garantissant un haut niveau de confidentialité pour les transactions. Elles conviennent en particulier aux entreprises et organisations désireuses de bénéficier des avantages de la technologie sans rendre leurs données accessibles au public. Leur principal inconvénient réside dans leur centralisation, ce qui va à l'encontre du principe de décentralisation de la chaîne de blocs.

Enfin, les *chaînes de blocs de consortium* offrent un compromis entre les systèmes publics et privés. Plus précisément, il s'agit de solutions semi-décentralisées gérées par plusieurs organisations collaborant au sein d'une même plateforme. Ce type de chaînes de blocs vise à atteindre un équilibre entre l'efficacité et le contrôle caractéristiques des systèmes privés et la confiance et la transparence propres aux systèmes publics. Elles sont particulièrement adaptées aux projets collaboratifs nécessitant un partage sécurisé de données entre diverses entités. R3 et l'Energy Web Foundation (46) sont des exemples de cette catégorie, mais également Hyperledger Fabric, qui, bien qu'étant utilisable en tant que chaîne de blocs privée, peut être également configuré pour agir comme une chaîne de blocs de consortium. Le principal défi de ces systèmes réside dans l'établissement d'un modèle de gouvernance convenant à toutes les parties prenantes. Le choix de Hyperledger Fabric pour ce projet repose sur plusieurs considérations clés. Contrairement à Ethereum, Hyperledger Fabric offre une architecture modulaire, permettant une personnalisation adaptée aux besoins spécifiques du projet. De plus, son approche basée sur des réseaux autorisés garantit un contrôle accru sur les participants, ce qui est crucial pour des applications nécessitant un partage de données sécurisé entre entités fiables. Enfin, l'absence de mécanismes de crypto-monnaie réduit les coûts opérationnels, rendant Hyperledger Fabric plus pratique pour des environnements dans lesquels les ressources sont limitées.

Ci-après, nous procéderons à une comparaison générationnelle entre trois exemples de chaînes de blocs : celle sous-jacente à Bitcoin, Ethereum et Hyperledger Fabric. Ces trois solutions représentent différentes approches en matière de type de chaîne de blocs (publique, privée ou de consortium) et de mécanismes de

consensus. Cette comparaison nous permettra de présenter progressivement les fonctionnalités spécifiques offertes par chacune de ces technologies, tout en illustrant comment elles s'alignent avec les principes fondamentaux de sécurité, de décentralisation et de transparence que nous avons discutés précédemment.

2.6 Chaîne de blocs sous-jacente à Bitcoin

Conçu en 2008, le Bitcoin (49) est apparu comme un concept révolutionnaire qui remettait en question les visions conventionnelles de la monnaie électronique. Il a été diffusé par le biais d'un article rédigé par une personne ou groupe connu sous le pseudonyme de Satoshi Nakamoto. L'objectif affiché de Bitcoin est d'être une nouvelle forme de système financier décentralisé, libre du contrôle d'une seule entité ou d'un seul gouvernement. Une caractéristique clé du Bitcoin est son offre limitée à 21 millions d'unités, imitant la rareté des métaux précieux et justifiant son surnom d'or digital. Cette rareté soutient sa valeur en offrant une protection contre l'inflation et la dévaluation monétaire, problèmes courants dans les systèmes monétaires traditionnels.

Dans le cas d'une transaction Bitcoin où Bob veut envoyer un bitcoin, il initie la transaction depuis son portefeuille numérique vers l'adresse du portefeuille d'Alice, qui est ensuite diffusée sur le réseau Bitcoin. Les mineurs valident la transaction en résolvant des puzzles cryptographiques par une *PdT*. Ce processus assure la sécurité du réseau grâce à l'effort de calcul nécessaire pour résoudre les puzzles. La transaction, avec d'autres, est combinée en un bloc, résumée par un arbre de Merkle et ajoutée au registre. Ce processus, qui prend environ dix minutes, est transparent et sécurisé grâce à l'usage de la PdT. Ce mécanisme garantit que chaque transaction est vérifiable et immuable, empêchant toute falsification ou double dépense. Cependant, cette sécurité a un coût : le débit de Bitcoin est inférieur à celui des opérations bancaires traditionnelles, ne traitant qu'approximativement 4,6 transactions par seconde. Cela s'explique par la nature énergivore et chronophage de la PdT, où la validation d'un bloc nécessite des ressources computationnelles significatives.

2.7 Ethereum

Ethereum (16) représente une évolution de la technologie, introduite par son créateur, Vitalik Buterin, en tant que plate-forme non seulement pour les transactions de crypto-monnaie, mais aussi pour l'exécution de contrats intelligents (15). Ces contrats intelligents sont des programmes qui exécutent automatiquement les termes d'un contrat lorsque des conditions prédéfinies sont remplies, ouvrant de nombreuses possibi-

lités d'utilisation. Par exemple, un contrat de gestion de fiducie peut être utilisé pour sécuriser les transactions entre un acheteur et un vendeur, en libérant les fonds uniquement lorsque les conditions de la transaction sont remplies. Pour éviter que des contrats mal écrits ou malveillants n'épuisent les ressources du réseau en s'exécutant indéfiniment, Ethereum utilise un mécanisme de frais appelé «gaz». Le gaz est une unité qui mesure la quantité de travail nécessaire pour exécuter des opérations sur le réseau Ethereum et chaque opération d'un contrat intelligent a un coût en gaz associé. Lorsqu'un utilisateur souhaite exécuter un contrat, il doit spécifier une limite de gaz et payer en Ether pour ce gaz. Le système de gaz garantit que les contrats intelligents ne peuvent pas monopoliser les ressources du réseau, car l'exécution s'arrêtera automatiquement lorsque la limite de gaz est atteinte ce qui permet de maintenir la sécurité et la stabilité du réseau, tout en incitant les développeurs à écrire des contrats intelligents efficaces et optimisés.

De plus, «The Merge» d'Ethereum fait référence à la transition du réseau Ethereum de son mécanisme de consensus initial basé sur la PdT vers un mécanisme de PdE. Cette évolution, achevée en 2022, vise à améliorer l'efficacité énergétique et le passage a l'échelle du réseau. Dans le modèle PdT, les mineurs résolvent des puzzles cryptographiques pour valider les transactions et sécuriser le réseau, un processus énergivore alors que la PdE sélectionne les validateurs en fonction de la quantité de cryptomonnaie qu'ils détiennent et mettent en jeu. Ce changement réduit considérablement la consommation d'énergie d'Ethereum, améliore sa sécurité et prépare le terrain pour de futures mises à niveau, telles que le *sharding* ou «fragmentation», qui consiste à diviser la chaîne de blocs en plusieurs segments, appelés fragments, chacun capable de traiter des transactions de manière indépendante. Cette approche améliore le passage à l'échelle en permettant le traitement parallèle des données, qui augmentera la capacité et la vitesse des transactions sur le réseau.

2.8 Hyperledger fabric

Les coûts associés à Ethereum, tels que les frais de gaz ou les engagements dans le système de PdE, peuvent ne pas être alignés avec les objectifs ou les ressources d'un projet. Toutefois, les avantages de la chaîne de blocs (tels que sa décentralisation et sa sécurité) peuvent demeurer des propriétés intéressantes. Hyperledger Fabric (34) représente une approche nuancée et flexible, conçue spécifiquement pour des environnements autorisés (ou permissionnés), qui se distingue des chaînes de blocs publiques comme Ethereum ou Bitcoin en raison de sa modularité, de ses protocoles de consensus personnalisables et de son architecture orientée entreprise.

Contrairement aux chaînes publiques, Hyperledger Fabric repose sur des identités numériques robustes

gérées par des Autorités de Certification (AC) pour authentifier les participants. Chaque organisation participante dispose d'un certificat numérique émis par une AC, garantissant que seuls les acteurs autorisés peuvent accéder au réseau et y effectuer des transactions. Cette approche offre un contrôle granulaire sur les permissions et améliore la sécurité globale du système en minimisant les risques d'accès non autorisé. De plus, Hyperledger Fabric est construit autour de canaux, un concept permettant de partitionner les données au sein du réseau. Plus précisément, un canal est un sous-réseau privé auquel seuls certains participants peuvent accéder et les transactions au sein d'un canal sont visibles uniquement par ses membres, offrant une confidentialité accrue sans sacrifier la transparence générale. Dans le cadre de notre projet, cette fonctionnalité est exploitée pour créer des canaux d'ingestion, périphériques et globaux (voir section 5.4), permettant une gestion localisée et sécurisée des données tout en garantissant une cohérence globale. Un autre aspect clé de Hyperledger Fabric est son architecture modulaire. Cette modularité s'exprime dans la capacité à personnaliser les protocoles de consensus, les langages pour les contrats intelligents (tels que Go, Java ou Node.js) et les bases de données utilisées pour stocker l'état du registre. Par exemple, Hyperledger Fabric prend en charge LevelDB et CouchDB comme options de bases de données pour le stockage des états, offrant une flexibilité en fonction des exigences du projet.

Le modèle de transaction de Fabric se distingue par son découplage en trois étapes intégrées : les transactions sont d'abord *proposées* par les clients en interagissant avec les *chaincodes*, puis elles sont *validées* par les pairs afin de vérifier qu'elles respectent les politiques du réseau et ne violent pas les conditions du contrat intelligent, avant d'être finalement *engagées* sous forme de blocs dans le registre distribué. Ce pipeline transactionnel renforce l'intégrité des données en s'assurant que seules les transactions conformes parviennent jusqu'à l'enregistrement final. Hyperledger Fabric s'appuie donc sur ce pipeline transactionnel, mais son comportement dépend fortement du protocole de consensus choisi. Comme introduit dans la section 2.4, plusieurs familles de protocoles existent : les mécanismes tolérants aux fautes byzantines (comme PBFT/TBFT), les solutions plus légères comme Solo utilisées en phase de développement, ou encore Raft, basé sur l'élection de leaders, privilégié pour les environnements de production. Cette intégration directe de différents protocoles illustre la modularité de Fabric et son adaptation à divers contextes applicatifs.

En résumé, les principales caractéristiques d'Hyperledger Fabric sont les suivantes :

— *Enregistrements immuables et de confiance*. Hyperledger Fabric, tout comme Bitcoin, respecte le principe de l'immuabilité. Ainsi, chaque transaction enregistrée sur le registre est permanente et immuable. Cette fonctionnalité garantit un niveau élevé de confiance et de fiabilité dans les données

partagées sur le réseau, ce qui est crucial pour les applications nécessitant des enregistrements infalsifiables.

— *Décentralisation dans un contexte différent*. Bien que Fabric fonctionne comme un réseau autorisé, ce qui implique un contrôle sur les participants et leurs permissions, il reste partiellement décentralisé. Le contrôle est distribué entre les entités autorisées, évitant les points uniques de défaillance tout en garantissant une sécurité et une transparence accrues.

— *Transparence encadrée*. Hyperledger Fabric fournit un environnement transparent à tous les participants autorisés grâce à son architecture de chaîne de blocs autorisée. Dans ce système, bien que l'accès soit restreint aux participants pré-approuvés (contrairement aux chaînes publiques comme Ethereum ou Bitcoin), chaque transaction enregistrée est visible par tous les membres autorisés du réseau. Cette transparence interne est renforcée par l'utilisation de canaux, qui permettent de segmenter les données accessibles par différents groupes de participants en fonction de leur rôle et de leurs besoins. Ainsi, même si Hyperledger Fabric n'est pas public dans le sens où n'importe qui pourrait y participer, il garantit que les participants autorisés partagent une vision commune de l'état du registre. Cette capacité à vérifier et à auditer indépendamment les transactions dans un environnement restreint, mais partagé renforce la confiance entre les parties prenantes, car chacune sait que les transactions sont enregistrées de manière immuable et accessible, selon des règles définies collectivement. Cette approche est particulièrement efficace dans les consortiums ou les réseaux d'entreprises.

— *Usage public dans un environnement contrôlé*. Hyperledger Fabric, bien que souvent utilisé dans des environnements privés ou en consortium, peut également être configuré pour un usage public décentralisé. Dans ce cadre, il offre une plate-forme sur laquelle différentes organisations ou entités peuvent collaborer et effectuer des transactions de manière transparente, sécurisée et immuable, sans avoir besoin d'une chaîne de blocs publique.

Dans cette recherche, notre choix s'est orienté vers Hyperledger Fabric pour plusieurs raisons cruciales. Tout d'abord, Hyperledger Fabric est reconnu pour sa diversité de protocoles de consensus qui joue un rôle essentiel dans le fonctionnement de celui-ci, assurant ainsi la fiabilité et la cohérence du réseau. Le mécanisme de TBPF (18), en particulier, est conçu pour contrer les fautes byzantines, qui représentent des défaillances système où les nœuds peuvent se comporter de manière malveillante ou imprévisible. Dans le cadre de Hyperledger Fabric, TBPF contribue à atteindre un consensus même dans des conditions défavorables, garantissant que tous les nœuds participants s'accordent sur l'état du registre et préservant de ce fait l'intégrité et la cohérence des données. En complément de TBPF, Hyperledger Fabric supporte

également d'autres protocoles de consensus comme Solo (7) et Raft (52). Le protocole de consensus Solo est l'option la plus simple et est généralement utilisé pour les environnements de développement et de test. Il repose sur un seul ordonnanceur, ce qui signifie qu'un seul nœud est responsable de la création et de l'organisation des blocs avant de les diffuser aux autres nœuds du réseau. Ce mode de fonctionnement est idéal pour les petites équipes ou les tests locaux, car il minimise la complexité et les ressources nécessaires pour faire fonctionner un réseau Fabric. Cependant, ce modèle n'est pas conçu pour des environnements de production, puisqu'il constitue un point de défaillance unique. Ainsi, si le nœud ordereur tombe en panne, tout le réseau devient inopérant, ce qui rend Solo inadapté pour les réseaux nécessitant une haute disponibilité et une tolérance aux pannes.

Le protocole de consensus Raft est une méthode plus robuste et adaptée aux environnements de production. Il s'agit d'un protocole de consensus réparti, où plusieurs nœuds ordereurs participent à la gestion des blocs. Contrairement à Solo, Raft ne repose pas sur un seul nœud, mais sur un ensemble de nœuds qui collaborent pour élire un leader responsable de l'organisation des blocs. Si le leader échoue, un nouveau leader est automatiquement élu parmi les nœuds restants, garantissant ainsi la continuité du service. Raft est conçu pour offrir une tolérance aux pannes, une haute disponibilité et une meilleure évolutivité que Solo, ce qui en fait un choix préféré pour les déploiements de production où la résilience et la performance sont cruciales.

Le choix du mécanisme de consensus approprié dépend étroitement des besoins spécifiques de chaque application. Si TBPF est optimal pour les environnements nécessitant une forte résilience aux fautes, Raft offre une solution plus simple et plus efficiente pour les réseaux moins exposés aux actes malveillants. En résumé, l'intégration de ces mécanismes variés de consensus enrichit l'adaptabilité d'Hyperledger Fabric, lui permettant de couvrir un large éventail de cas d'usage, des réseaux exigeant une haute sécurité et une grande tolérance aux fautes, jusqu'à des systèmes plus épurés et performants. Cette polyvalence représente un avantage majeur pour les projets cherchant des solutions de consensus spécifiques afin de répondre à des besoins uniques, soulignant ainsi pourquoi Hyperledger Fabric est l'approche choisie pour notre projet.

2.9 Fragmentation

La fragmentation (ou *sharding*) est une technique de partitionnement de bases de données élaborée pour booster le passage à l'échelle et la performance des réseaux de chaînes de blocs. Traditionnellement, chaque nœud doit non seulement traiter l'intégralité des transactions, mais également maintenir une copie com-

plète de toute la chaîne. Cette exigence peut devenir un frein significatif à la mise à l'échelle et à la performance du réseau à mesure que le volume de transactions croît. En fragmentant le réseau en fragments plus maniables, ou «shards», chaque portion gérant une fraction des transactions et une part de l'état global de la chaîne de blocs, la fragmentation permet une exécution simultanée de multiples transactions. Cet éclatement augmente de manière notable la capacité et la rapidité du système.

Le fonctionnement de la fragmentation se base sur plusieurs principes clés. Le réseau est d'abord divisé en plusieurs fragments, chaque fragment opérant un sous-ensemble distinct des transactions et des états de la chaîne de blocs, agissant en effet comme une mini-chaîne indépendante avec son propre journal et ses propres nœuds de validation. Cette répartition des charges permet aux transactions d'être traitées en parallèle, réduisant ainsi le temps de traitement général et augmentant le débit du réseau. Toutefois, pour que le réseau global reste cohérent et que les interactions entre comptes ou contrats intelligents sur différents fragments soient possibles, un mécanisme de communication inter-fragments est nécessaire. Cela implique l'utilisation de protocoles spécifiques qui autorisent l'échange d'informations et la validation des transactions s'étendant sur plusieurs fragments. Parmi ces protocoles, on peut citer :

— *Protocoles de communication inter-chaînes (PCIC)*. Ces protocoles permettent la validation et l'exécution des transactions entre différents fragments en assurant la cohérence globale de la chaîne.
— *Transactions inter-chaînes atomiques (TICA)*. Ils garantissent l'atomicité des transactions inter-fragments, où toutes les étapes doivent être validées ou annulées ensemble, préservant ainsi l'intégrité des données.
— *Preuves d'état Merkle (PEM)*. Ils sont utilisés pour vérifier l'état d'un fragment spécifique lorsqu'une transaction dépend d'une interaction entre fragments.
— *Protocole Interledger (PIL)*. Bien qu'initialement conçu pour connecter différentes chaînes de blocs, ce protocole peut être adapté pour faciliter les interactions inter-fragments dans un réseau fragmenté.

Ces mécanismes assurent une interaction fluide entre les fragments tout en maintenant la cohérence et la sécurité des données sur l'ensemble du réseau. Chaque fragment adopte son propre mécanisme de consensus pour valider ses transactions, pouvant être identique ou différent de celui employé par la chaîne de blocs principale, selon les besoins spécifiques du fragment.

Les avantages de la fragmentation sont multiples. Elle permet au système d'accroître son nombre de transactions par seconde sans exiger une augmentation proportionnelle de la puissance de calcul ou de la bande

passante réseau de chaque nœud. En facilitant le traitement parallèle des transactions, la fragmentation peut considérablement accélérer la vitesse de traitement sur le réseau et rendre l'ensemble du système plus efficient en termes de ressources, puisque les nœuds ne sont pas contraints de stocker l'intégralité de la chaîne ou de traiter toutes les transactions. Néanmoins, elle introduit aussi des défis spécifiques. La sécurité est une préoccupation majeure, car chaque fragment, étant potentiellement moins sécurisé qu'une chaîne de blocs unifiée du fait du nombre réduit de nœuds nécessaires pour réaliser une attaque, peut présenter des vulnérabilités supplémentaires. La gestion des transactions qui impliquent plusieurs fragments requiert des mécanismes sophistiqués pour garantir leur cohérence et leur atomicité, soulignant la complexité de ces opérations. En outre, assurer un équilibrage de charge optimal entre les fragments, nécessitant des algorithmes dynamiques pour ajuster la répartition des transactions et des états avec le temps, représente un autre défi.

CHAPITRE 3
ÉTAT DE L'ART SUR LES MÉTHODES DE COLLECTE DES DONNÉES DE MOBILITÉ

Ce chapitre a pour objectif de présenter et d'analyser les technologies et protocoles existants qui ont un lien direct avec les problématiques de géolocalisation et de confidentialité des données. En particulier, nous survolerons d'abord les fondements des technologies de géolocalisation, notamment le GPS, pour ensuite introduire des systèmes de collecte de données de mobilité et de géolocalisation. Enfin, nous y explorons deux approches possibles pour améliorer la collecte et l'authentification des données de géolocalisation dans un cadre décentralisé le protocole FOAM, qui propose une solution décentralisée pour les services géospatiaux, et la détection participative, qui exploite les capacités collectives des utilisateurs pour la collecte de données à grande échelle. En examinant ces technologies, nous mettrons en lumière leurs forces, leurs limites et leur potentiel d'intégration dans un cadre décentralisé comme celui des chaînes de blocs. Cette analyse servira de préambule à la discussion de la solution proposée dans les chapitres suivants.

3.1 Données de mobilité et GPS

Les données de mobilité font référence aux informations générées par les déplacements des individus ou des objets à travers l'espace. Elles peuvent inclure des informations telles que la vitesse, la direction, les points d'arrêt et les trajectoires empruntées, où une trajectoire désigne une suite ordonnée d'observations géographiques horodatées d'un même individu ou objet $\tau = \{(lat_i, lon_i, t_i)\}_{i=1..n}$. Ces données sont essentielles pour diverses applications, allant de la gestion du trafic (28) à la planification urbaine (40), en passant par la personnalisation des services mobiles (54). La géolocalisation désigne la capacité de déterminer la position géographique d'un individu ou d'un objet en temps réel, généralement à l'aide de technologies comme le GPS, les réseaux Wi-Fi ou les signaux de téléphonie mobile. Elle est au cœur de nombreux services modernes, tels que la navigation, les applications de partage de position ou encore la surveillance environnementale.

Le *Global Positioning System* (GPS) (gps) a été développé par le département de la défense des États-Unis à l'origine pour des applications militaires avant d'être ouvert pour un usage civil. Cette technologie repose sur une constellation de 24 satellites en orbite, disposés de manière à assurer qu'au moins quatre d'entre eux soient visibles depuis n'importe quel point sur la surface de la terre à tout moment. Les récepteurs GPS, qui équipent aujourd'hui principalement nos téléphones intelligents, calculent leur position en captant les

signaux émis par ces satellites. Ces signaux incluent des informations essentielles telles que la position exacte du satellite, son orbite, ainsi que des horodatages précis. Ainsi, grâce à la trilatération des données recueillies de plusieurs satellites, le récepteur peut déterminer avec précision sa position géographique. En effet, le récepteur mesure le temps de propagation des signaux entre lui et plusieurs satellites pour en déduire des distances. Ces distances correspondent aux côtés de triangles géométriques qui, lorsqu'ils sont combinés, permettent de calculer la position précise du récepteur. Cette distinction est importante : contrairement à la triangulation, qui repose sur la mesure d'angles, la localisation par GNSS est bien une trilatération. Le même principe est d'ailleurs utilisé pour la localisation basée sur les signaux cellulaires ou Wi-Fi, où l'on se fonde également sur les distances déduites des temps de transmission plutôt que sur des angles.

Le GPS fait partie des principaux systèmes globaux de navigation par satellite (GNSS) existants, aux côtés de GLONASS (Russie) (glo), Galileo (Union européenne) (esa) et Beidou (Chine) (bei). Bien qu'ils reposent sur des principes similaires de trilatération par signaux satellites, ces systèmes sont indépendants les uns des autres et utilisent chacun leur propre constellation. Chaque système présente ses spécificités et, selon la région du monde, certains peuvent offrir une couverture ou une précision supérieure. L'intégration et l'utilisation combinée de ces différents systèmes peuvent potentiellement améliorer la fiabilité et la précision des services de géolocalisation.

La précision offerte par le GPS et les autres systèmes GNSS est généralement évaluée entre 3 à 10 mètres pour les applications civiles, cette variabilité étant influencée par divers facteurs tels que la configuration spatiale des satellites, les interférences atmosphériques, ou encore les obstacles bloquant directement les signaux (comme les bâtiments en milieu urbain dense) (6). Des technologies d'augmentation, à l'instar du WAAS (*Wide Area Augmentation System*) et des récepteurs à double fréquence, ont été développées pour raffiner cette précision à un niveau centimétrique (33). Ces améliorations, bien qu'initialement destinées aux usages professionnels et scientifiques, démontrent l'innovation croissante dans le domaine de la géolocalisation pour répondre aux exigences croissantes en matière de précision. Cependant, pour certains projets, cette technologie est délaissée au profit d'outils jugés plus précis ou plus polyvalents. Prenons l'exemple de FOAM (26), décrit ci-dessous. Dans le contexte de ce projet, le GPS est remplacé par l'utilisation d'équipements de radiofréquence spécifiques, malgré un investissement initial plus important qui, en théorie, devrait être amorti par les utilisateurs. Dans d'autres contextes, le choix se porte sur la *détection participative*, un mode de collecte de données où plusieurs utilisateurs recueillent, partagent et agrègent les

informations issues de leur environnement. Ce principe a été largement adopté pour le déploiement d'applications de suivi de la COVID-19 ces dernières années, tirant parti du grand nombre d'appareils mobiles connectés et de leur répartition géographique.

Dans cette recherche, nous choisissons de recourir au GPS même si d'autres solutions, potentiellement plus précises, existent pour les raisons suivantes. La première repose sur la mission fondamentale du projet : améliorer la confidentialité des données de géolocalisation. Dans ce cadre, une précision de trois à dix mètres est considérée comme un compromis acceptable, offrant un niveau d'utilité satisfaisant pour l'utilisateur final sans pour autant nécessiter des positions géographiques extrêmement précises. Le projet s'appuie sur une carte de points répartis dans une zone donnée pour conserver des repères communs et élaborer des modèles de mouvement.

Ainsi, la précision du GPS n'est cruciale qu'en termes de détermination du point de référence le plus proche et de suivi des changements de position, pour lesquels la précision actuelle du GPS est largement suffisante. La seconde raison concerne la familiarité des utilisateurs avec le GPS. Avec la généralisation des téléphones intelligents, le GPS est devenu un outil quotidien. Il paraît donc plus pragmatique d'intégrer notre solution à ce système déjà bien établi et largement adopté, en minimisant les modifications et sans altérer l'expérience utilisateur, plutôt que d'espérer un basculement vers une technologie de rupture. Ainsi, nous croyons que même si la protection de la vie privée est primordiale, elle ne suffit pas toujours à motiver un changement. De ce fait, moins nous imposons de modifications et d'impacts perturbateurs sur les technologies en place, plus nous augmentons nos chances de succès.

3.2 Approches centralisées pour la capture de la mobilité

Les données de mobilité peuvent être collectées par une variété de moyens, incluant, mais ne se limitant pas, aux applications de navigation. Ainsi, elles peuvent également être extraites des métadonnées de photos (72), déduites à partir des adresses IP (29), ou encore estimées grâce à la trilatérations des signaux de réseaux cellulaires et Wi-Fi (41). Cette diversité de méthodes de collecte soulève des questions importantes concernant la précision des données ainsi que la protection de la vie privée des individus. Ainsi, la prédominance de grands acteurs technologiques tels que Google et Apple met en lumière les enjeux associés au consentement des utilisateurs et à la sécurité des données personnelles.

3.2.1 Google Maps

Google Maps est une solution de cartographie et de navigation largement adoptée, en partie grâce à son intégration native dans Android, le système d'exploitation dominant pour les téléphones intelligents. Elle est utilisée par plus d'un milliard d'utilisateurs uniques chaque mois (60). Google Maps propose une variété de services, notamment la navigation en temps réel, les informations sur le trafic, les vues de rue et satellite, ainsi que les listes d'entreprises (58). Cependant, le modèle centralisé de Google pour la gestion des données de localisation a été critiqué pour son manque de transparence. En 2018, Google a révélé une faille de sécurité dans Google+ qui avait exposé les données personnelles de 500 000 utilisateurs (42). De plus, la *Australian Competition and Consumer Commission* (*ACCC*) a poursuivi Google pour avoir induit les consommateurs en erreur quant à la collecte de leurs données de localisation (**?** 44). Malgré la désactivation de l'option «Historique des positions», Google continuait à collecter ces données lorsque l'activité Web et les applications étaient activées. Cette affaire a souligné la nécessité de pratiques plus transparentes et de consentement éclairé dans la gestion des données personnelles.

3.2.2 Apple Maps

Apple Maps, intégré aux appareils iOS, représente une alternative populaire à Google Maps. Apple met l'accent sur la protection de la vie privée des utilisateurs, mais certaines fonctionnalités, comme le service «Emplacements significatifs», ont suscité des préoccupations. Ce service enregistre automatiquement les lieux fréquemment visités pour offrir des services personnalisés, tels que des suggestions de navigation vers des lieux importants comme le domicile ou le lieu de travail. Cependant, en 2020, une étude a révélé qu'Apple continuait à collecter certaines données de localisation même lorsque les services de localisation étaient désactivés (53). Bien que ces pratiques visent à améliorer l'expérience utilisateur, elles ont renforcé les critiques concernant le manque de transparence dans la gestion des données personnelles. Apple a également été confronté à des incidents de confidentialité, comme le bug FaceTime de 2019, qui permettait aux utilisateurs d'espionner d'autres personnes avant que l'appel ne soit accepté (12).

3.2.3 Enjeux communs de vie privée associés à Google Maps et Apple Maps

Les deux solutions partagent des enjeux communs liés à leur nature centralisée. Le stockage massif des données de localisation sur leurs serveurs centralisés les expose à des risques accrus de violation de données. Des incidents, tels que la faille de Google+ ou le bug FaceTime d'Apple, montrent que même les géants

technologiques ne sont pas à l'abri de vulnérabilités. De plus, leurs pratiques de collecte de données, parfois perçues comme intrusives, posent des défis éthiques, notamment en matière de consentement éclairé et d'utilisation des données personnelles. Ces enjeux soulignent l'importance de développer des alternatives décentralisées qui offriraient un meilleur équilibre entre fonctionnalité, sécurité et respect de la vie privée.

3.2.4 Open Street Map

OpenStreetMap (OSM) est un projet collaboratif qui ambitionne de devenir l'équivalent de Wikipédia pour les données géolocalisées (**?**). Il offre pour cela un accès à un logiciel permettant d'éditer des cartes, les utilisateurs peuvent ajouter des routes, des chemins, des bâtiments, et d'autres points d'intérêt, en utilisant des images satellites, des relevés GPS personnels, ou d'autres sources de données, ainsi que l'enrichissement de cartes grâce à des tags. En mars 2024, le projet comptait plus de 10,5 millions d'utilisateurs enregistrés et s'appuyait aussi sur des données officielles fournies par divers gouvernements. OSM s'est révélé extrêmement précieux lors de crises humanitaires, comme le tremblement de terre en Haïti en 2010, où il a permis de créer la carte numérique la plus détaillée d'Haïti à l'époque en seulement deux jours, devenant un outil indispensable pour toutes les ONG participant à l'effort d'aide humanitaire. La gratuité et l'aspect collaboratif d'OSM représentent à la fois ses principaux atouts et ses limites. En effet, si cette approche permet une réactivité et une utilité remarquables en situation d'urgence, l'absence de soutien financier et matériel important rend OSM moins séduisant comparé aux solutions commerciales plus étoffées.

Les projets comme Google Maps et Apple Maps bénéficient de vastes ressources financières, d'équipes de développement dédiées, et d'une infrastructure matérielle robuste, ce qui leur permet d'offrir des fonctionnalités avancées telles que la navigation en temps réel, les mises à jour fréquentes, et l'intégration de données provenant de multiples sources (trafic en temps réel, données commerciales, etc.). En revanche, OSM repose principalement sur une communauté de bénévoles pour la collecte et la mise à jour des données, ce qui peut entraîner des disparités en termes de couverture géographique et de qualité des informations. Ainsi, bien que les contributions bénévoles soient souvent rapides et réactives, elles peuvent aussi être inégales, particulièrement dans les régions moins peuplées ou moins accessibles où la collecte de données peut être sporadique. De plus, l'absence de financement centralisé signifie qu'OSM doit généralement compter sur des dons et des subventions pour maintenir ses serveurs et ses infrastructures, ce qui peut limiter sa capacité à rivaliser avec les mises à jour et les innovations continues des solutions commerciales.

3.2.5 Maps.Me

Maps.Me (61) est une application de navigation mobile réputée pour ses fonctionnalités hors ligne, offrant la possibilité d'accéder aux cartes et à la navigation sans nécessiter de connexion Internet. Initialement développée par une entreprise danoise avant d'être acquise par le groupe Mail.Ru, Maps.Me vise comme base utilisateur les voyageurs ainsi que les personnes vivant dans des zones dans lesquelles la connectivité Internet est restreinte. L'application s'alimente en données via OpenStreetMap, ce qui lui assure un accès à des informations cartographiques à la fois complètes et actualisées. L'un des principaux atouts de Maps.Me réside dans son orientation vers les fonctionnalités hors ligne. Ainsi, en plus de simplifier la vie des utilisateurs dans les régions isolées ou avec un accès Internet limité, faire le calcul en local est positif pour la vie privée, car les données de localisation ne sont pas constamment transmises à des serveurs distants. Bien que l'accent mis sur le hors-ligne réduise la nécessité d'une connexion de données constante, les inquiétudes liées au stockage des données de localisation et à la vie privée des utilisateurs demeurent pertinentes.

En résumé, l'approche prise par Maps.Me combine le développement centralisé de l'application avec une collecte décentralisée des données, lui conférant une place particulière dans l'écosystème des solutions de navigation GPS. Cependant, malgré ses nombreux avantages, Maps.Me fait face à des défis pour concurrencer des applications plus établies et dépendantes du réseau Internet. Ainsi, contrairement à Google Maps et Apple Maps, qui offrent des informations en direct sur le trafic, les accidents ou les fermetures de routes, Maps.Me est limité par son fonctionnement hors ligne. Bien que pratique pour les zones à faible connectivité, cette limitation réduit son attrait pour les utilisateurs qui privilégient des mises à jour en temps réel pour optimiser leurs déplacements. De plus, Maps.Me ne dispose pas des vastes ressources financières et techniques de ses concurrents. Cela limite sa capacité à intégrer des fonctionnalités avancées, comme l'intelligence artificielle pour améliorer les itinéraires ou les suggestions personnalisées, et à rivaliser avec la fréquence et la précision des mises à jour offertes par Google et Apple. Enfin, sa dépendance aux données d'OpenStreetMap, bien que bénéfique pour sa couverture géographique et la fréquence de ses mises à jour, signifie aussi qu'elle hérite des limitations inhérentes au contenu généré par les utilisateurs, des défis similaires à ceux rencontrés par OSM.

3.2.6 Waze

Waze (62) est présenté par ses créateurs comme une application de navigation communautaire et s'est imposé comme l'un des rares concurrents sérieux des grandes entreprises technologiques dans le domaine

du GPS, avant son acquisition par Google en juin 2013. Cette application permet à ses utilisateurs de signaler de nombreuses informations liées au trafic, qu'il s'agisse d'accidents, d'embouteillages ou de contrôles de police. Ces rapports, combinés aux informations officielles fournies par les gouvernements pour les systèmes de gestion du trafic (tel que Bison Futé en France), sont utilisés pour offrir des itinéraires et des mises à jour de trafic en temps réel.

En 2014, Waze a lancé le programme «Citoyen Connecté», qui établit un échange bidirectionnel de données entre Waze et ses partenaires, incluant plus de 450 gouvernements et municipalités à travers le monde. Cependant, l'application a soulevé certaines préoccupations concernant l'anonymat de ses utilisateurs, notamment en raison de l'emploi de systèmes d'incitation (59). En effet, les utilisateurs accumulent des points pour leur profil en rapportant des incidents de trafic, ce qui soulève des questions sur la protection de la vie privée et la sécurité des données personnelles.

3.2.7 Résumé des enjeux principaux des systèmes existants

Dans le secteur de la géolocalisation, la question du consentement éclairé est cruciale avant même la collecte des données. Souvent, les utilisateurs ne sont pas pleinement conscients, ou ne sont informés que de manière complexe à travers des politiques de confidentialité obscures, sur la façon dont leurs données de localisation sont collectées et utilisées. Cette ambiguïté dans l'obtention du consentement des utilisateurs engendre d'importants enjeux éthiques et de confidentialité. Ensuite, au stade de la collecte, la précision et la fiabilité des données géolocalisées sont primordiales. La qualité de ces données peut varier selon la technologie employée, comme le GPS, le Wi-Fi ou les réseaux cellulaires, et est fréquemment affectée par des facteurs environnementaux. Par ailleurs, la tendance à la surcollecte de données, bien qu'utile pour des analyses fines, soulève des préoccupations quant à la protection de la vie privée et au risque de mauvais usage de ces informations sensibles.

Une fois les données récoltées, la manière dont elles sont stockées et exploitées devient essentielle. Sécuriser les données de géolocalisation stockées représente un défi constant, les violations de données pouvant révéler des informations personnelles sensibles. De surcroît, le partage de ces données avec des tiers ou leur utilisation pour des objectifs allant au-delà des attentes initiales, tels que la publicité ciblée ou le profilage des utilisateurs, intensifie les problèmes de confidentialité. Pour pallier ces enjeux, le secteur s'aligne sur certaines normes et pratiques opérationnelles, telles que l'ISO/IEC 27001 pour la gestion de la sécurité de l'information et l'ISO/IEC 27701 pour la gestion des informations personnelles (37).

Application concrète des normes ISO au contexte mobilité. Dans le cas particulier des données de mobilité, les référentiels ISO/IEC 27001 et 27701 se traduisent par une série de mesures appliquées tout au long de la chaîne de traitement. Dès la collecte, le principe de minimisation prévaut : seuls les capteurs strictement nécessaires sont activés, la fréquence d'échantillonnage est réduite, et tout traitement à risque doit faire l'objet d'une analyse d'impact (DPIA/EFVP) accompagnée d'un consentement explicite et journalisé de l'utilisateur. Lors de la transmission, la confidentialité et l'intégrité sont assurées par le chiffrement en transit (TLS), complété par une authentification forte des clients et une journalisation rigoureuse des accès. Ces principes se prolongent dans la phase de stockage et de traitement, où les traces sont chiffrées au repos, les identifiants sont séparés des trajectoires, et les accès sont strictement limités selon le principe du « besoin de savoir », le tout appuyé par une traçabilité renforcée à travers des journaux immuables.

Au-delà de la gestion interne, les normes encadrent également les relations avec les tiers. Tout partage de données implique des accords contractuels précisant finalités et durées, complétés par des tests systématiques d'anonymisation et de ré-identification ainsi que par des mécanismes d'agrégation. La fin du cycle de vie des données est tout aussi encadrée, avec des politiques de rétention limitées dans le temps, une purge automatique et la garantie d'un droit à l'effacement documenté. ISO/IEC 27701 apporte enfin une extension essentielle en définissant les rôles et registres de traitement (responsable, sous-traitant), en cartographiant les catégories spécifiques de données (trajectoires, points d'arrêt) et en établissant des procédures précises pour répondre aux droits des personnes concernées.

Ces normes exigent des organisations qu'elles adoptent des pratiques positives, notamment :

- La mise en place de contrôles techniques pour protéger les données, comme le chiffrement ou l'accès restreint aux systèmes sensibles.
- L'évaluation régulière des risques pour identifier et atténuer les vulnérabilités potentielles.
- La sensibilisation et la formation des employés à la sécurité et à la confidentialité des données.
- La documentation rigoureuse des processus pour assurer une traçabilité et une conformité accrues.

Cependant, certaines pratiques comme l'anonymisation des données, autrefois perçues comme un standard de vie privée, sont de plus en plus contestées (51). Ainsi, des recherches récentes ont montré que même des données anonymisées peuvent être ré-identifiées en croisant plusieurs ensembles de données, compromettant ainsi la confidentialité initialement garantie. Par exemple, des techniques d'apprentissage automatique ou des bases de données accessibles publiquement peuvent être utilisées pour associer des identifiants uniques à des données prétendument anonymes. Le renforcement des législations sur la vie

privée telles que le RGPD (Règlement général sur la protection des données) accentue la protection des données personnelles et du consentement des utilisateurs (66).

Ces nouvelles normes et régulations sont essentielles pour répondre aux défis liés à la transparence du consentement, à la précision et à la protection des données collectées, ainsi qu'à leur sécurité et à leur exploitation éthique. Elles visent à garantir une gestion des données qui respecte à la fois les droits des utilisateurs et les exigences réglementaires, tout en limitant les risques d'abus ou de mauvaises pratiques dans les systèmes de géolocalisation.

3.3 FOAM

FOAM (26) est un protocole développé pour les marchés de données décentralisés construit sur Ethereum. Son but est de faciliter les services géospatiaux à travers la coordination «d'ancres de zone», qui sont des balises radio conçues pour authentifier et prouver la présence d'un utilisateur à un emplacement spécifique à un moment donné. Ces ancres utilisent la synchronisation d'horloge byzantine tolérante aux fautes (23) 2.4 pour maintenir des horloges précises. Le protocole incite à l'emploi des ancres de zone par le biais de jetons FOAM. Ces jetons jouent un double rôle. Tout d'abord, les jetons sont déposés en tant que collatéral pour renforcer la sécurité du système. Ce mécanisme limite les comportements malveillants, car les validateurs risquent de perdre leur garantie en cas de fraude avérée. Ensuite, les validateurs, dont l'identité est connue, utilisent des balises radio pour effectuer des trilatérations et des mesures anti-fraude qui authentifient les emplacements dans les zones définies. En retour, ils reçoivent des frais de transaction et des jetons échangeables comme récompense. Ce modèle économique encourage la participation active des utilisateurs et favorise l'expansion du réseau.

3.3.1 Preuves de localisation respectueuses de la vie privée

Au-delà de fournir des preuves de localisation, FOAM ouvre la voie à des systèmes décentralisés où la confidentialité est une priorité. Les preuves de localisation respectueuses de la vie privée (*privacy-preserving location proofs*) sont un domaine en plein essor.

Modèle générique d'une PPLP. Un schéma typique comporte : (1) un *prouver* (appareil) qui capte un signal de présence (GNSS, RF, Wi-Fi ou ancre FOAM) ; (2) un *vérificateur* qui valide la preuve ; (3) un *transcript* signé liant {identité éphémère, lieu/temps} à l'ancre. Pour préserver la vie privée, on remplace la divulgation brute

par : *(a) signatures)* pour l'authenticité ; *(b) engagements/hachages)* pour lier sans révéler ; *(c) preuves à divulgation nulle de connaissance (ZK)* pour prouver "j'étais dans la zone Z à t" sans révéler la coordonnée exacte ; *(d) agrégation/anonymisation)* pour limiter l'inférence.

Des variantes existent qui permettent de prouver qu'un individu ou un appareil était présent à un endroit précis à un moment donné, sans révéler davantage d'informations personnelles. Parmi les architectures notables, on peut citer ZKP-LP (*Zero-Knowledge Proof Location Proofs*) qui se base sur des preuves à divulgation nulle de connaissance pour valider la présence sans révéler la localisation exacte (71) ainsi que CrowdBLPS qui utilise la chaîne de blocs pour sécuriser les preuves de localisation tout en garantissant l'anonymat des utilisateurs (73). Ces modèles apportent des solutions aux problèmes de confidentialité souvent associés aux technologies géospatiales, en particulier celles basées sur des infrastructures centralisées. Dans ce chapitre, nous *survolons* ces approches ; la conception précise utilisée par notre solution est détaillée plus loin (voir sections 4.3.1 et 5.2).

3.3.2 Coordonnées spatiales cryptographiques et géo-hache

Le géo-hache (ou *geohash*) est une méthode de codage de coordonnées géographiques en une chaîne de caractères alphanumériques (50). Chaque géo-hache représente une zone géographique spécifique et permet de diviser un espace en sous-segments de plus en plus précis. La procédure pour construire un géo-hache à partir d'une coordonnée géographique est la suivante :

1. **Encodage binaire des coordonnées.** Les coordonnées de latitude et de longitude sont encodées séparément en utilisant un format binaire. La plage initiale est fixée à $[-90, 90]$ pour la latitude et $[-180, 180]$ pour la longitude. Les coordonnées sont ensuite divisées récursivement par un processus de dichotomie, assignant un bit un si la valeur est dans la moitié supérieure de la plage actuelle et un bit zéro sinon.
2. **Entrelacement des bits.** Les bits des coordonnées de latitude et de longitude sont entrelacer pour créer une séquence binaire unique.
3. **Conversion en chaîne alphanumérique.** La séquence binaire résultante est ensuite convertie en une chaîne de caractères en utilisant un ensemble prédéfini de symboles, généralement une base-32 standard. Par exemple, la coordonnée $(48, 8566, 2, 3522)$ pourrait être encodée en géo-hache sous une forme telle que `u09tunq`.

Il est à noter que, puisque la plage de longitude est deux fois plus large que celle de latitude, l'entrelace-

ment binaire induit une résolution deux fois moins fine en longitude. Cette approximation est généralement considérée comme acceptable dans l'usage courant du géo-hache, mais elle ne concerne pas directement notre solution, qui repose sur les géoadresses, qui sont des points d'ancrage fixes. Les géo-haches sont largement utilisés pour indexer et rechercher efficacement des données géographiques dans des systèmes distribués.

Les coordonnées spatiales cryptographiques, qui associent un géo-hache représentant une localisation physique à une adresse Ethereum, représentent une notion innovante et sont comparables à la solution que nous envisageons, la différence principale étant que notre solution utilise les intersections du réseau routier comme points d'ancrage fixes plutôt que des points générés dynamiquement par les utilisateurs. Le système de visualisation de FOAM, utilisant une application web pour interpréter les données stockées sur la chaîne de blocs et les présenter dans un format facilement exploitable, est également une caractéristique intéressante.

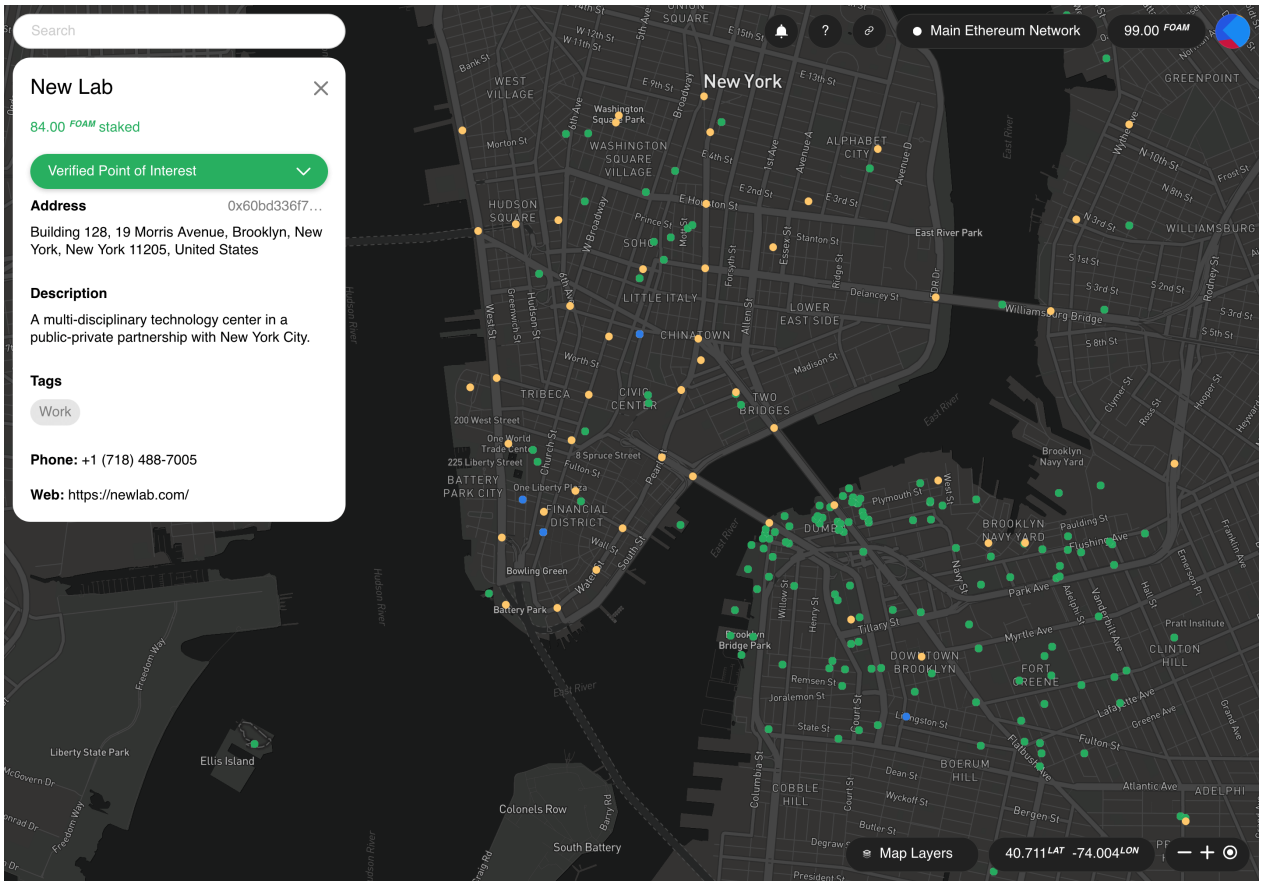

Figure 3.1 – Carte de la solution FOAM.

3.4 Détection participative

La détection participative (ou *crowdsensing* en anglais), tire parti du potentiel collectif des individus, généralement via leurs appareils mobiles, pour collecter et partager des données dans un large éventail d'applications (73). Cette technologie vise à utiliser l'ubiquité des dispositifs intelligents pour amasser des données volumineuses fournies volontairement par les participants. Ces données proviennent des divers capteurs intégrés dans les appareils mobiles contemporains, comme le GPS, accéléromètres et caméras, permettant ainsi la collecte d'informations géospatiales, environnementales et urbaines.

Le fonctionnement de la détection participative repose sur le partage volontaire des données de capteurs des appareils des utilisateurs, soit vers un système centralisé, comme des serveurs gérés par une entité unique (28), soit distribué à travers des réseaux décentralisés, où les données sont partagées directement entre les participants ou via des nœuds intermédiaires sur un réseau pair-à-pair (40). Ce partage peut être actif, avec des utilisateurs soumettant consciemment des informations, ou passif, où les applications récoltent des données en arrière-plan. Ces informations sont ensuite agrégées et analysées pour en extraire des connaissances. Dans le cas centralisé, les données sont collectées et agrégées sur des serveurs pour analyse, tandis que dans un système distribué, les participants collaborent pour traiter et valider les données localement ou au sein de petits clusters, réduisant ainsi les risques d'un point de défaillance unique.

Les principaux avantages de la détection participative incluent sa capacité à recueillir des données à grande échelle de manière économique et avec une finesse de détail importante. Ceci permet une surveillance en temps réel et la création de bases de données haute résolution qui peuvent servir à diverses fins telles que la gestion du trafic, la planification urbaine et la surveillance environnementale. Par exemple, OpenStreetMap (section 3.2.4) est une base de données géospatiale collaborative qui fournit des informations de haute précision sur les routes, bâtiments et autres points d'intérêt, souvent enrichie par des contributions participatives (32). Un autre exemple est la base de données PlanetScope, qui offre des images satellitaires à haute résolution et est utilisée pour le suivi environnemental, la gestion des ressources naturelles et la cartographie urbaine (36). D'autres points forts comprennent sa capacité à étendre facilement le système pour y inclure plus de participants, augmentant ainsi le volume et la diversité des données collectées ainsi que la réduction du besoin d'infrastructures dédiées, puisque la détection participative s'appuie sur des appareils personnels déjà en circulation. De plus, elle encourage la participation et la sensibilisation des communautés, les participants contribuant activement à la collecte de données pour des causes communes.

Cependant, cette méthode soulève aussi des défis, notamment en termes de confidentialité. Ainsi, la collecte et le traitement des données personnelles posent d'importants enjeux de sécurité, nécessitant des mesures pour protéger les informations sensibles tout en fournissant des données utiles. Par exemple, dans OpenStreetMap, une des mesures de sécurité mises en place est la pseudonymisation des utilisateurs, où les contributions sont associées à des pseudonymes plutôt qu'à des identités réelles, réduisant ainsi les risques de divulgation de données personnelles. De plus, le chiffrement TLS (*Transport Layer Security*) est utilisé pour sécuriser les communications entre les utilisateurs et les serveurs (32), empêchant ainsi l'interception des données lors de leur transmission. Par ailleurs, Xiong et co-auteurs (2014) ont montré que la confidentialité différentielle est une approche qui peut être appliquée pour limiter les risques de ré-identification des utilisateurs (70). Cette technique introduit du bruit aléatoire dans les données collectées afin de garantir qu'aucune information individuelle ne puisse être isolée tout en permettant des analyses agrégées fiables. Cette méthode est particulièrement utile pour les ensembles de données massifs où des attaques de ré-identification sont possibles

De plus, il y a un enjeu de confidentialité à cause de la variabilité de la précision des données recueillies, qui est influencée par la qualité des capteurs et les conditions de collecte. Quercia et co-auteurs (2011), par exemple, ont montré comment des techniques d'obfuscation comme l'ajout de faux emplacements, la généralisation des coordonnées à une zone plus large, ou encore la rotation des points de localisation peuvent être utilisées pour protéger les informations de localisation tout en partageant des données précises (56). Cependant, Pyrgelis et ses co-auteurs (2017) ont démontré que même avec des méthodes d'anonymisation et d'obfuscation, les données agrégées peuvent encore être vulnérables à des attaques de ré-identification, en particulier lorsque les adversaires disposent d'informations auxiliaires (55). Ils ont illustré comment des modèles d'apprentissage automatique et des bases de données publiques peuvent être exploités pour identifier à nouveau des utilisateurs dans des ensembles de données agrégées, soulignant ainsi la nécessité d'approches encore plus robustes. Il est également important de noter que pour que l'approche fonctionne, il est nécessaire de développer des mécanismes incitatifs efficaces pour motiver une participation active et régulière. Par exemple, Waze, une application de navigation basée sur la détection participative (section 3.2.6], incite ses utilisateurs à signaler des incidents de trafic en leur attribuant des points et des badges qui renforcent leur statut au sein de la communauté.

Enfin, en intégrant les mécanismes de consensus pour la vérification et l'anonymisation des données, il est possible d'assurer une sécurité renforcée des données de localisation dans des environnements distribués.

Bano et ses co-auteurs (2019) ont proposé une approche intégrant des mécanismes de consensus pour améliorer la sécurité et la confidentialité dans des environnements distribués (10). Ils décrivent comment des consensus distribués, comme la *tolérance pratique aux fautes byzantines* (TFBP) ou la Preuve d'enjeu (PdE), peuvent être utilisés pour valider les données et garantir leur immuabilité tout en respectant la confidentialité des participants. Ces mécanismes introduisent également des incitations économiques qui encouragent un comportement honnête au sein du réseau.

Un exemple de systèmes mêlant la détection participative et la chaîne de blocs pour préserver la confidentialité des utilisateurs dans les systèmes de localisation est CrowdBLPS (*Crowdsensing Blockchain-based Location Privacy-Preserving System*) (73). CrowdBLPS est un système qui combine les chaînes de blocs et des mécanismes cryptographiques pour garantir la confidentialité des participants tout en validant des preuves de localisation rassemblées à l'aide de la détection participative. Le fonctionnement du système repose sur les éléments suivants :

1. **Enregistrement des données de localisation.** Les participants collectent des preuves de localisation à l'aide de leurs appareils mobiles. Ces données sont ensuite anonymisées en générant des identités temporaires. Ce mécanisme est comparable à celui que nous introduirons dans notre solution, où des identités jetables sont utilisées pour renforcer l'anonymat des utilisateurs (*cf.* section 4.3.1).
2. **Validation par consensus.** Les preuves de localisation sont enregistrées sur la chaîne de blocs. Les nœuds participants valident ces données via un mécanisme de consensus, tel que le *TFBP* ou une variante adaptée.
3. **Mécanismes cryptographiques utilisés.** Le *chiffrement homomorphe* permet d'effectuer des calculs sur des données chiffrées, garantissant ainsi la confidentialité des participants pendant le traitement des preuves tandis que les *preuves à divulgation nulle de connaissance (ZKP)* assurent que les participants peuvent prouver la validité de leur localisation sans révéler de détails sensibles.
4. **Applications pratiques.** Les données validées peuvent être utilisées pour des services de gestion urbaine, des analyses de trafic, ou des applications sociales anonymisées tout en préservant la confidentialité des utilisateurs.

En s'appuyant sur la chaîne de blocs, CrowdBLPS propose une solution de rechange décentralisée et sécurisée aux systèmes traditionnels de localisation et de navigation, souvent appelés « applications GPS » (comme Google/Apple Maps), qui exploitent les signaux du GPS ou d'autres GNSS pour fournir des services de cartographie et d'itinéraire. Ce type de système ouvre des perspectives intéressantes pour des applications comme la gestion urbaine, les services géolocalisés et les réseaux sociaux anonymisés basés sur la

localisation. En surmontant ces défis, particulièrement concernant la confidentialité et la qualité des données, la détection participative pourrait significativement enrichir la robustesse et l'utilité des systèmes de géolocalisation basés sur les chaînes de blocs. Ainsi, bien qu'elle ne fasse pas actuellement partie de notre solution, du fait que nous nous basons sur le GPS ainsi qu'un ensemble de points de références sans utiliser les autres solutions de capture généralement mise à profit pour la détection participative tels que les accéléromètres ou les réseaux Wifi ou Bluetooth , son ajout futur reste une possibilité pour augmenter la fiabilité du système.

CHAPITRE 4
SOLUTION PROPOSÉE : LOC[ATIONC]HAIN

4.1 Introduction

Dans ce chapitre, nous présentons notre solution à la problématique de la confidentialité et de l'efficacité dans les services de géolocalisation. LoChain, conçu sur le principe de la décentralisation, vise à offrir une alternative robuste et sécurisée aux modèles centralisés actuellement dominants dans l'industrie. Plus précisément, en exploitant les technologies de chaînes de blocs, nous proposons une architecture qui non seulement protège la vie privée des utilisateurs à travers des mécanismes tels que les identités jetables et l'injection de bruit, mais qui améliore également la précision et la fiabilité des services de géolocalisation. Ce système repose sur deux piliers fondamentaux : l'application de géoadresses pour uniformiser et anonymiser les données de localisation et une approche décentralisée qui renforce la sécurité et la transparence du traitement des données.

De plus, l'idée de développer un module complémentaire plutôt que de tenter de lancer une nouvelle application GPS autonome émerge d'une évaluation pratique du paysage actuel du marché. En effet, des acteurs majeurs comme Google et Apple ont consolidé leurs positions en proposant des services complets de cartographie et de navigation profondément intégrés dans leurs solutions qui jouent des rôles clés dans la vie numérique quotidienne. Ces plates-formes sont devenues les outils de navigation par défaut pour des centaines de millions de personnes, ancrées dans l'écosystème des appareils mobiles et soutenues par une collecte approfondie de données, la familiarité des utilisateurs et la confiance dans la marque.

Il serait donc déraisonnable de tenter d'être un nouvel acteur sur le marché, car une concurrence directe avec les géants établis nécessite non seulement d'égaler, mais aussi de dépasser leurs offres en termes de précision des données, d'expérience utilisateur et de fonctionnalités supplémentaires, ce qui nécessite des ressources et une innovation substantielle. De plus, la nature fermée de ces plates-formes signifie qu'il y a peu de place pour les développeurs externes pour ajouter des fonctionnalités directement à ces applications, ce qui limite le potentiel d'intégration et d'amélioration. Il est également important de souligner que les problématiques que nous tentons d'aborder dans ce projet n'ont pas de lien avec la manière dont les applications GPS fonctionnent, mais bien avec la manière dont les donnéessont générées et stockées plus tard dans le processus.

En concevant un système qui fonctionne avec les applications GPS existantes, l'extension peut introduire de nouvelles fonctionnalités ou améliorer celles existantes sans obliger les utilisateurs à abandonner leurs plateformes préférées. Cette approche tire parti de l'utilisation généralisée et de l'infrastructure des services existants tout en s'attaquant aux lacunes spécifiques ou aux problèmes de confidentialité que ces plateformes pourraient ne pas résoudre entièrement. Néanmoins, il est important de préciser que dans notre cas précis, l'extension sert plutôt à ajouter une seconde manière de traiter les données qui fonctionne en parallèle des solutions existantes et ne cherche pas forcément à les remplacer.

4.2 Principes fondamentaux

4.2.1 Utilisation de géoadresses

Nous appelons *géoadresse* un identifiant discret et stable qui représente un *point d'ancrage fixe* du réseau routier (par exemple une intersection ou un rond-point), et qui sert d'abstraction uniforme pour des positions lat./lon. voisines. Une fonction d'assignation $g : \mathbb{R}^2 \to \mathcal{A}$ traduit chaque observation (lat, lon) en une géoadresse correspondante, laquelle est ensuite intégrée dans l'ensemble public S de points d'ancrage partagés. Les géoadresses sont utilisées dans LoChain pour (1) *uniformiser* les données de localisation hétérogènes et (2) *renforcer la confidentialité* en évitant l'exposition de coordonnées exactes (la méthode d'implémentation est explicitée plus en détails dans la section 5.2),

Dans notre approche, les géoadresses représentent des lieux physiques fixes (ex : intersections du réseau routier). Elles doivent donc répondre aux critères suivants :

- **Uniformité et comparabilité.** Chaque géoadresse correspond à un point d'ancrage unique et stable, utilisable de la même manière par toutes les applications et tous les utilisateurs. Cette uniformité garantit que deux observations différentes situées à proximité soient traduites en la même géoadresse.
- **Granularité contrôlée.** Dans notre prototype, la maille spatiale est fixée à l'avance, ce qui assure un niveau constant de généralisation.
- **Confidentialité.** Un tableau de correspondance $(lat, lon) \mapsto$ géoadresse est maintenu. Ce tableau permet aux applications clientes d'assigner automatiquement une géoadresse en fonction de la localisation courante, tout en évitant de manipuler ou transmettre les coordonnées exactes dans les transactions.
- **Interopérabilité.** Le tableau de correspondance S et la fonction d'assignation g sont documentés,

assurant une utilisation cohérente et reproductible par différents acteurs du réseau.

La variabilité des données de géolocalisation, même entre individus se déplaçant dans le même espace géographique, peut entraver considérablement l'analyse et l'application de ces informations. Par exemple, les utilisateurs naviguant dans les rues d'une ville à l'aide de Google/Apple Maps peuvent générer des données de localisation légèrement différentes en fonction de la précision du GPS de leur appareil, des itinéraires spécifiques qu'ils empruntent ou même du côté de la rue utilisé, comme illustré dans la figure 4.1 où chaque couleur sur le graphique correspond au trajet d'un utilisateur. Cette disparité dans la collecte de données devient un obstacle à la création de solutions basées sur la géolocalisation cohérentes et universellement applicables, mais elle pose également des risques potentiels pour la confidentialité des utilisateurs, permettant dans certains cas l'inférence de l'identité des individus. En effet, des études ont démontré que les données de localisation peuvent être utilisées pour identifier des individus, même lorsqu'elles sont agrégées ou anonymisées.

Zandbergen (2009) a illustré dans son étude comment les données GPS, croisées avec des informations contextuelles ou personnelles, peuvent permettre d'identifier des personnes dans un environnement urbain dense (72). En analysant les données GPS provenant de smartphones, il a démontré que des schémas d'activité spécifiques, tels que les visites répétées à des adresses résidentielles ou professionnelles, permettent de déduire l'identité d'un individu avec une grande précision. Ces observations mettent en évidence les limites des techniques traditionnelles d'anonymisation et soulignent la nécessité de méthodes supplémentaires pour protéger la vie privée. De plus, Pyrgelis et collaborateurs (2017) ont montré que l'agrégation de données de localisation peut révéler des schémas individuels, permettant ainsi la ré-identification même dans des contextes de grande échelle (55). Ces risques mettent en lumière l'importance d'une approche de géo-anonymisation, pour réduire la probabilité d'inférence d'identité.

Pour répondre à ces défis, la méthode des géoadresses traduit les données de géolocalisation variables en un ensemble uniforme d'adresses comme illustré dans la figure 4.2 où chaque point représente une géoadresse, c'est-à-dire un point d'ancrage fixe du réseau routier (intersection/rond-point) auquel les observations GPS voisines sont rattachées. Cette opération de généralisation standardisée réduit la fragmentation des données, facilitant ainsi leur analyse et renforçant la confidentialité des utilisateurs. En focalisant l'attention sur l'analyse de modèles de mobilité collective plutôt que sur le suivi des mouvements individuels, les géoadresses permettent une meilleure compréhension des flux de déplacement, essentielle pour des applications telles que la planification urbaine et la gestion du trafic.

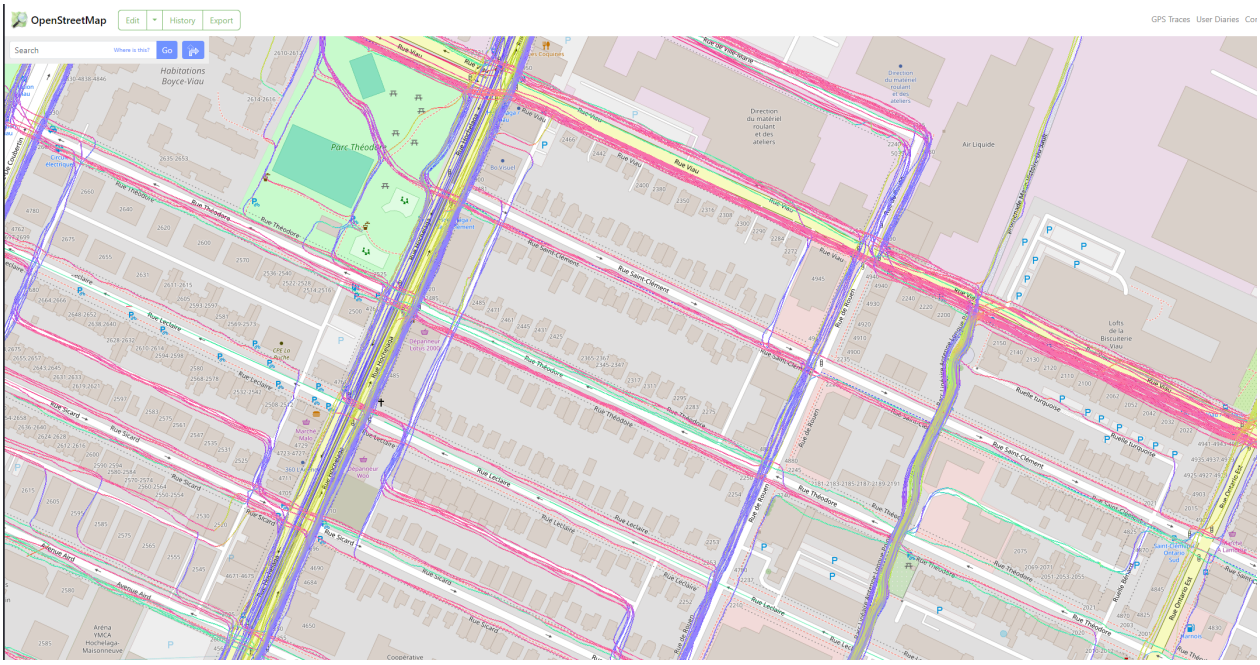

Figure 4.1 – Traces GPS publiques sur OSM.

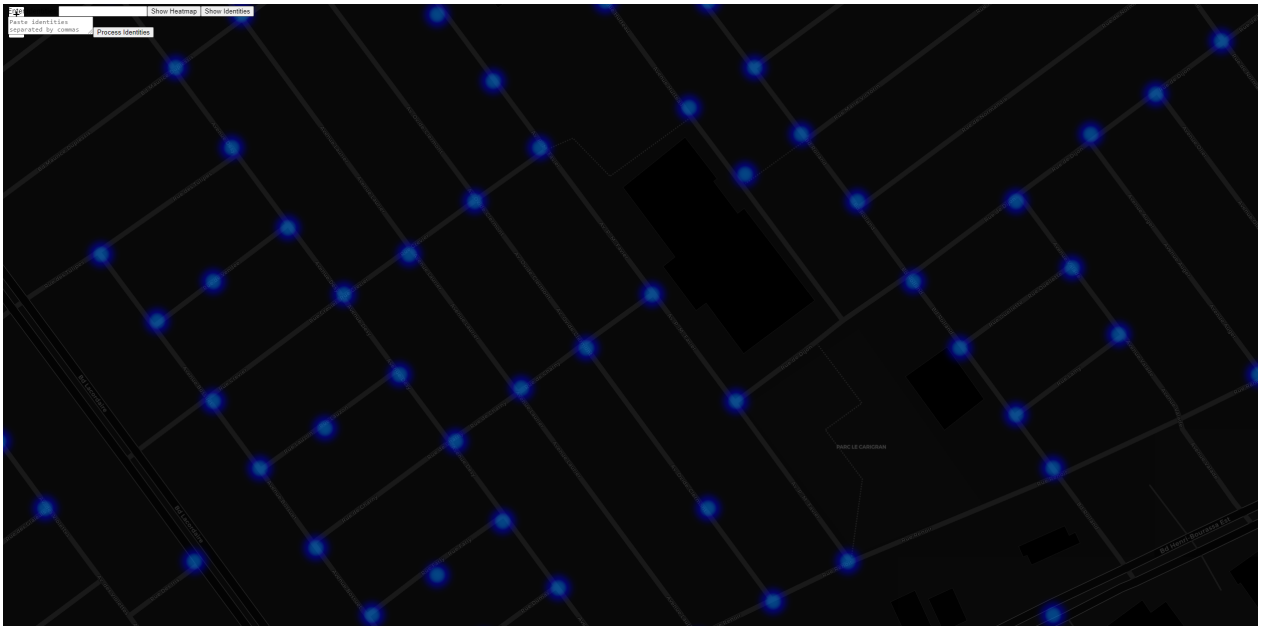

Figure 4.2 – Géoadresses LoChain : chaque point est un point d'ancrage fixe (intersection/rond-point) servant d'abstraction uniforme pour des observations GPS voisines.

L'intégration des géoadresses dans un système de géolocalisation offre l'avantage de simplifier l'analyse des données en rendant les points de référence uniformes et comparables, même entre différents dispositifs.

De plus, cette méthode atténue les risques de ré-identification en regroupant les informations individuelles dans des ensembles agrégés et anonymisés.

4.2.2 Décentralisation

La décentralisation dans le contexte de la gestion des données de géolocalisation représente une transition majeure par rapport aux modèles traditionnels centralisés, où les données sont collectées, stockées et traitées par une seule entité. Ces modèles, adoptés par des géants de la technologie, centralisent les données sur des serveurs qu'elles contrôlent entièrement, conférant à ces entités la responsabilité exclusive de protéger et de gérer ces informations sensibles. Cependant, plusieurs cas documentés de violations de données ou d'abus ont mis en lumière les vulnérabilités de ces systèmes. Par exemple, Google a été impliqué dans des affaires concernant l'exploitation non autorisée de données de localisation (44), et Apple a été critiqué pour la collecte de données liées à son service «Emplacements significatifs», même lorsque les services de localisation étaient désactivés (53). Ces incidents soulignent les limites des systèmes centralisés, notamment leur dépendance à des pratiques internes souvent opaques.

En outre, bien que certaines de ces données soient anonymisées, des recherches ont montré que ces anonymisations restent vulnérables à la réidentification. Par exemple, De Montjoye et ses collaborateurs (2013) ont démontré qu'en croisant seulement quatre points de localisation anonymes avec des données auxiliaires, il est possible de réidentifier jusqu'à 95% des individus dans un ensemble de données anonymisées (20). Ces résultats mettent en évidence les risques associés à la centralisation des données de localisation, même lorsque des protections superficielles sont mises en place.

La décentralisation introduit un paradigme dans lequel les données ne sont pas stockées en un seul endroit, mais sont réparties sur un réseau distribué, fréquemment en utilisant la technologie des chaînes de blocs. Ce modèle réduit considérablement les risques associés aux points de défaillance uniques, car les données sont redondantes et disséminées sur plusieurs nœuds du réseau. De plus, dans un système décentralisé, les mécanismes d'anonymisation et de chiffrement sont intrinsèquement intégrés, offrant une sécurité renforcée sans dépendre de politiques internes spécifiques. Par exemple, les chaînes de blocs publiques telles qu'Ethereum implémentent des mécanismes cryptographiques permettant une immutabilité des données tout en préservant la confidentialité des utilisateurs (68). Cette approche renforce la transparence et limite les possibilités d'abus, car aucune entité unique ne contrôle l'intégralité des données.

Un avantage supplémentaire des systèmes décentralisés est leur résilience accrue face aux cyberattaques. Dans un réseau décentralisé, un attaquant doit compromettre la majorité des nœuds pour obtenir un contrôle significatif, ce qui est considérablement plus complexe que de cibler un serveur centralisé. Par exemple, Nakamoto (2008) a démontré que la sécurité d'un réseau basé sur la preuve de travail (PdT) dépend de la puissance totale de calcul des participants honnêtes, rendant les attaques coûteuses et peu probables à grande échelle (49). Cependant, la décentralisation n'est pas exempte de défis. Elle introduit une complexité technique notable dans la conception et la maintenance des réseaux. Par exemple, des problèmes d'évolutivité peuvent émerger lorsque le nombre de participants ou le volume de données dépasse la capacité du réseau à les gérer efficacement (17). En outre, il est essentiel de garantir que ces systèmes restent intuitifs et accessibles pour l'utilisateur final, afin de favoriser leur adoption à grande échelle.

4.3 Brique fondamentales

4.3.1 Identités jetables

Cette section présente les principales briques de LoChain — identités jetables, preuves de localisation, routage via Tor, géo-pools et perturbation de la localisation — en détaillant pour chacune les objectifs, les mécanismes et garanties de sécurité, ainsi que les choix d'implémentation retenus dans le prototype.

Les identités jetables introduisent une méthode de renouvellement régulier des identifiants numériques des utilisateurs au sein des services de géolocalisation, s'inspirant directement des principes de confidentialité et de sécurité propres à la technologie chaîne de blocs, plus précisément de Bitcoin où il est parfois recommandé d'utiliser une adresse par transaction (9). Cette approche génère des identifiants temporaires pour un nombre limité de transactions ou une période définie, après lesquels ces identités sont remplacées par de nouvelles, sans aucun lien direct avec les identifiants numériques permanents de l'utilisateur.

L'objectif principal de la mise en œuvre d'identités jetables est d'améliorer la confidentialité des utilisateurs. En effet, dans les services de géolocalisation traditionnels, les mouvements d'un utilisateur sont souvent liés à un identifiant unique et persistant (tel qu'un compte ou un identifiant d'appareil). Ce lien permet de suivre et de profiler les utilisateurs au fil du temps, ce qui peut éventuellement entraîner des enjeux en termes de vie privée. Les identités jetables cherchent à prévenir cela en garantissant que toute tentative de construction du profil de mouvement d'un utilisateur serait limitée à la durée de vie d'une identité unique et temporaire, ce qui réduit considérablement les risques de l'utilisation des données collectées à des fins

de suivi ou de profilage invasif.

Les différentes étapes du processus des identités jetables sont les suivantes :

1. **Génération d'identité.** A chaque utilisation du service, les utilisateurs reçoivent une identité temporaire unique, similaire à une adresse publique en termes de chaînes de blocs. Cette identité ne révèle pas les identifiants personnels permanents de l'utilisateur.
2. **Représentation de la transaction.** Les mouvements de l'utilisateur sont représentés par des transactions liées à cette identité jetable, utilisant une liste de géoadresses pour cartographier l'espace géographique parcouru. Une transaction est générée chaque fois qu'une nouvelle adresse la plus proche est identifiée lors du déplacement.
3. **Élimination et renouvellement de l'identité.** Après un certain nombre de transactions ou une période spécifique, l'identité temporaire actuelle est abandonnée. Une nouvelle identité est générée pour les futures activités, empêchant ainsi la corrélation des données collectées sur différentes périodes avec un seul utilisateur.

Contrairement au système classique d'Hyperledger Fabric, qui utilise des autorités de certification (*Certificate Authorities* en anglais) pour attribuer et gérer les identités des nœuds du réseau, une approche alternative est adoptée pour la gestion des identités temporaires des utilisateurs. Cette solution est conçue pour éviter de surcharger les nœuds du réseau tout en maintenant un haut niveau de confidentialité et de sécurité.

Les nœuds du réseau, qui appartiennent aux organisations constituant le système, continuent de gérer leurs identités via le mécanisme standard des AC de Fabric. Cependant, pour les identités temporaires des utilisateurs, les nœuds génèrent des *pools d'identités disponibles*, constitués d'adresses standard de chaînes de blocs. Le processus fonctionne comme suit. Tout d'abord, des *contrats intelligents* spécifiques sont déployés sur le réseau et accessibles via les plateformes mobiles des utilisateurs. Lorsqu'un utilisateur souhaite obtenir une identité temporaire, il exécute le contrat intelligent à partir de son appareil mobile pour sélectionner une identité disponible dans le pool. Une paire de clés cryptographiques est alors générée localement sur l'appareil de l'utilisateur, en utilisant l'algorithme ECDSA (*Elliptic Curve Digital Signature Algorithm*), choisi pour sa compatibilité standard et son efficacité. Par la suite, l'utilisateur transmet l'identité sélectionnée ainsi que la clé publique générée localement au système, tout en conservant la clé privée sur son appareil pour garantir la sécurité et la confidentialité. Cette architecture décentralisée permet de déléguer efficacement la gestion des identités temporaires aux utilisateurs finaux, réduisant ainsi la charge sur les nœuds du

réseau. Elle assure également que les identités temporaires restent anonymes et sécurisées, en cohérence avec les principes de confidentialité et de minimisation des données.

Les identités jetables présentent plusieurs avantages, notamment une amélioration significative de la confidentialité, en protégeant les utilisateurs contre le suivi et le profilage à long terme grâce au renouvellement périodique des identifiants. L'ensemble du processus, de la génération à l'élimination des identités, est sécurisé par des méthodes cryptographiques, assurant la protection de la vie privée des utilisateurs, telles que :

- *Paires de clés asymétriques (ECDSA)*. Chaque identité jetable est associée à une paire de clés publique/privée générée localement sur l'appareil de l'utilisateur. La clé privée n'est jamais transmise et sert exclusivement à signer les transactions, tandis que la clé publique correspondante est utilisée par le réseau pour vérifier ces signatures.
- *Hachage cryptographique*. Pour anonymiser les transactions associées à ces identités, rendant impossible le retour à des identifiants d'origine.
- *Signatures numériques (ECDSA)*. Ces signatures sont utilisées pour garantir l'authenticité et l'intégrité de chaque transaction. Ce mécanisme s'appuie sur la paire de clés asymétriques associée à l'identité jetable.

Notons que la confidentialité en transit des transactions est assurée séparément par l'utilisation de TLS, tandis que la paire de clés ECDSA sert exclusivement aux signatures numériques et à la vérification des transactions. De plus, cette approche soutient le principe de minimisation des données, en ne collectant que les informations strictement nécessaires sans conserver d'identifiants à long terme. Cependant, la mise en œuvre de ce système n'est pas sans défis. Il exige une gestion avancée pour assurer des transitions d'identité fluides,et il est essentiel de maintenir une expérience utilisateur sans compromis tout en intégrant ces mesures de sécurité, ce qui peut être complexe. Cette gestion implique :

- *Uniformité des traces et indépendance des identités*. Il s'agit de garantir que les identités jetables produisent des traces de taille uniforme, tout en veillant à ce qu'il soit impossible de faire le lien entre deux identités successives. Cette séparation stricte renforce la confidentialité des utilisateurs et complique la corrélation entre les périodes d'activité.
- *Gestion des clés cryptographiques*. Il s'agit d'éviter les collisions et garantir que les paires de clés soient générées de manière sécurisée et distribuées efficacement.
- *Expérience utilisateur intuitive*. Pour cela, le mécanisme est intégré dans les services existants sans alourdir l'expérience utilisateur.

Spécification technique (condensée). Les identités jetables reposent sur un socle de primitives cryptographiques standardisées : la signature ECDSA sur la courbe P-256 (secp256r1), le hachage SHA-256 et un générateur pseudo-aléatoire sécurisé fourni par l'OS (CSPRNG) pour la génération des clés. Le cycle de vie de chaque identité suit un mécanisme clair : une paire de clés (sk, pk) est générée localement, puis associée à une identité jetable id_{tmp} définie avec des bornes d'usage $(N_{\max}, T_{\max})$. Lorsque ces bornes sont atteintes (par quota ou par temps), l'identité est désactivée côté appareil : la clé privée et le préfixe id_{tmp} ne sont plus réutilisés, et une nouvelle identité est créée. Les données et clés de l'ancienne identité restent toutefois archivées localement pour consultation, tandis que les transactions déjà soumises demeurent immuables sur le registre.

Chaque transaction est sérialisée sous la forme :

$$m = \langle id_{\mathrm{tmp}}, pk, geo, ts, nonce, meta \rangle$$

où geo désigne la géoadresse, ts l'horodatage, $nonce$ un entier strictement croissant par identité, et *meta* un champ optionnel (par exemple `noise_flag`). L'identifiant de transaction est alors calculé comme $txid = H(m)$. Avant envoi, le client signe le message via $\sigma = \mathrm{ECDSA.Sign}(sk, H(m))$, puis transmet (m, σ) au réseau. Les pairs vérifient la validité en contrôlant la signature $\mathrm{ECDSA.Verify}(pk, H(m), \sigma)$ ainsi que la monotonie du *nonce*. Enfin, la gestion des clés suit le principe de minimisation : la clé privée reste confinée dans le Keystore ou l'enclave sécurisée de l'appareil, tandis que seules les clés publiques et id_{tmp} sont visibles sur le réseau. Il n'y a aucune réutilisation de pk entre identités successives.

Propriétés garanties. Le mécanisme proposé assure d'abord l'uniformité des traces : les bornes $(N_{\max}, T_{\max})$ et la maille spatiale fixe imposent que chaque identité génère des trajectoires de taille comparable, ce qui limite les possibilités d'inférence basées sur la longueur des traces. À cela s'ajoute l'indépendance stricte entre identités successives : chaque couple (id_{tmp}, pk) est unique à une période donnée et aucune métadonnée persistante n'est transmise lors du passage à une nouvelle identité. Cette séparation est renforcée par le routage via Tor, qui dissocie les adresses IP des identités numériques.

La robustesse de la gestion des clés constitue une autre garantie importante. La génération est réalisée localement avec un CSPRNG, le secret n'est jamais exporté, et les identités désactivées à la rotation ou à l'expiration ne sont pas réutilisées, ni au niveau des clés privées (sk) ni des clés publiques (pk). L'espace de 256 bits rend par ailleurs les collisions négligeables. Enfin, le système prévient les attaques par rejeu et assure l'ordre des transactions grâce à la vérification d'un *nonce* strictement croissant par identité, couplée

au rejet automatique des doublons de $txid$.

À ce stade, nous pouvons résumer l'architecture spécifique aux identités jetables, indépendamment des autres composants du système (comme les géo-pools, décrits plus loin). Elle inclut :

- Une *application mobile utilisateur* qui génère localement une paire de clés asymétriques (ECDSA). La clé privée reste sur l'appareil tandis que la clé publique est communiquée au réseau.
- Un *contrat intelligent de gestion des identités* qui attribue une identité jetable depuis un pool prédéfini. Il associe cette identité à la clé publique fournie par l'utilisateur.
- Les *nœuds du réseau Fabric* qui enregistrent les identités jetables et utilisent les clés publiques pour vérifier les signatures associées aux transactions.

Ce mécanisme garantit que les utilisateurs conservent le contrôle exclusif de leurs clés privées, que l'assignation des identités est décentralisée, et que la vérification des transactions est rendue possible par la publication des clés publiques.

4.3.2 Preuves de localisation

Dans le domaine des services de géolocalisation, les preuves de localisation (*Location Proofs* en anglais) désignent des mécanismes permettant de certifier qu'un individu ou un appareil s'est trouvé à un endroit donné à un moment précis. Ces preuves, qui combinent des données de position et des signatures cryptographiques, sont utilisées dans des applications variées allant de la validation de présence dans des zones réglementées à des systèmes incitatifs basés sur la localisation. Le défi principal réside dans l'équilibre entre l'intégrité de la preuve (pour garantir son authenticité et empêcher les falsifications) et la confidentialité (pour éviter que cette preuve ne compromette l'identité ou les mouvements de l'utilisateur). Les approches modernes, notamment celles reposant sur la cryptographie avancée et les chaînes de blocs, permettent de relever ce défi en proposant des solutions sécurisées et respectueuses de la vie privée (30).

Dans le contexte des services de géolocalisation, l'intégrité et l'authenticité des données sont cruciales, non seulement pour l'exactitude des services offerts, mais aussi pour la confidentialité et la sécurité des utilisateurs. Notre système proposé intègre une fonction de preuve de localisation, développée pour authentifier chaque donnée de géolocalisation sans pour autant compromettre l'anonymat des utilisateurs. Le processus repose sur les étapes suivantes :

1. Lorsqu'un utilisateur entre dans une zone couverte par notre système, une identité jetable est gé-

nérée pour l'utilisateur. Cette identité est utilisée pour enregistrer les déplacements sous forme de transactions anonymes.

2. Chaque transaction est associée à des informations clés, notamment, (1) l'identité jetable utilisée pour signer la transaction, (2) l'horodatage indiquant le moment exact de la transaction, (3) les coordonnées géographiques sous forme de géoadresse, garantissant une granularité uniforme et une abstraction suffisante pour préserver la vie privée et (4) un identifiant unique de transaction généré pour éviter toute duplication ou falsification.
3. Une signature cryptographique est appliquée à chaque transaction à l'aide de la clé privée associée à l'identité jetable. Cette signature garantit l'authenticité des données, prouvant qu'elles ont été générées par une identité légitime du système, ainsi que la non-répudiation, empêchant l'utilisateur de nier l'origine de la transaction.

Le système utilise la cryptographie asymétrique pour sécuriser les transactions. Plus précisément, une signature basée sur l'algorithme ECDSA (*Elliptic Curve Digital Signature Algorithm*) est employée. Ce choix est motivé par son efficacité, sa compatibilité avec les systèmes décentralisés tels que les chaînes de blocs, et sa robustesse face aux attaques cryptographiques connues (38). Chaque transaction est signée avec une clé privée, tandis que la clé publique correspondante, liée à l'identité jetable, est utilisée pour valider la signature.

Une fois signées, les transactions sont consignées sur la chaîne de blocs, garantissant un registre immuable qui offre plusieurs garanties :

— *Transparence*. Toute entité autorisée peut vérifier l'intégrité des données sans accéder aux informations personnelles des utilisateurs.
— *Immutabilité*. Les transactions enregistrées ne peuvent pas être altérées ou supprimées, ce qui renforce la confiance dans le système.
— *Confidentialité*. Les identités jetables et les géoadresses utilisées dans les transactions ne révèlent pas d'informations permettant d'identifier l'utilisateur.

Le mécanisme de renouvellement des identités jetables, comme décrit précédemment, joue un rôle central dans la préservation de la vie privée. En rompant périodiquement la chaîne de corrélation entre les transactions, il est considérablement plus dur de reconstruire des profils de mouvement à long terme. Chaque nouvelle identité jetable introduit une discontinuité, compliquant le suivi ou le profilage invasif.

En conclusion, nous pensons que l'approche proposée répond aux exigences modernes des services de géo-

localisation en matière de précision et de sécurité tout en offrant des garanties robustes de confidentialité. Ces preuves de localisation permettent une validation efficace des données tout en minimisant les risques de suivi et de profilage des utilisateurs.

Modèle de menace et garanties. Le modèle de menace envisagé dans le cadre du prototype repose principalement sur deux adversaires réalistes : un pair honnête-mais-curieux au sein du réseau et un observateur passif capable de surveiller le trafic. L'objectif de notre conception est donc de garantir à la fois l'authenticité et l'intégrité des transactions grâce à la combinaison d'ECDSA et de SHA-256, mais aussi la non-répudiation grâce à l'utilisation systématique d'identités jetables. La rotation régulière de ces identités, couplée au routage par Tor, contribue en outre à briser les corrélations temporelles et à dissocier les sessions d'activité successives. La confidentialité en transit est assurée par l'usage systématique de TLS, tandis que la résistance face à un observateur global est renforcée par deux mécanismes complémentaires : l'anonymisation réseau via Tor et l'agrégation par géoadresses, qui empêche de relier directement une transaction à un individu.

Ainsi, le dispositif proposé offre un socle robuste de garanties en matière de sécurité et de vie privée. Les transactions bénéficient de protections cryptographiques assurant leur validité et leur non-altération, tandis que la dés-corrélation des identités et l'anonymisation réseau réduisent les risques de profilage. Enfin, l'agrégation sur géoadresses contribue à conserver l'utilité analytique des données tout en limitant la granularité accessible à un attaquant. Ce compromis illustre la capacité du système à concilier intégrité des données et préservation de la confidentialité dans un environnement distribué.

4.3.3 Routage via le réseau Tor

Avant d'expliquer son intégration dans notre système, il est essentiel de présenter Tor et son fonctionnement. Tor, ou *The Onion Router* (21), est une technologie permettant d'anonymiser les communications sur Internet. Il fonctionne en acheminant le trafic via un réseau mondial de relais bénévoles, en appliquant plusieurs couches de chiffrement, comme les couches d'un oignon, pour masquer l'origine, le contenu et la destination des communications. Lorsqu'un utilisateur transmet une requête via Tor, celle-ci passe généralement par un circuit de trois nœuds : un nœud d'entrée, un intermédiaire et un de sortie. Chaque nœud ne connaît qu'une partie des informations nécessaires pour traiter la requête, garantissant l'anonymat. Par exemple, le nœud d'entrée connaît l'adresse IP de l'expéditeur, mais pas sa destination finale, tandis que le nœud de sortie connaît la destination, mais pas l'origine.

Dans notre système, Tor est utilisé pour dissocier les identités temporaires de l'adresse IP des utilisateurs. Cette méthode anonymise le trafic Internet associé à la transmission des données de géolocalisation, atténuant ainsi le risque d'exposition des adresses IP lors des interactions en ligne ainsi que le risque de suivi à travers différentes identités jetables. En effet, une telle exposition pourrait permettre de retracer les données jusqu'à un utilisateur spécifique, compromettant son anonymat et sa confidentialité. Bien que l'utilisation de Tor puisse ralentir la transmission et que certains sites bloquent les nœuds de sortie, cette solution reste théoriquement viable dans notre contexte. Notre conception limite volontairement la taille des transactions pour préserver la confidentialité des données. Ce design minimaliste garantit que le réseau Tor n'est pas surchargé, même en conditions réelles. Bien que le prototype ait été testé localement, où les performances de Tor ne reflètent pas les conditions réelles d'un réseau géographiquement distribué, la nature légère des transactions rend cette hypothèse de viabilité raisonnable. Les tests complets sur un réseau global restent une étape future pour valider expérimentalement cette affirmation.

Rôle d'Orbot dans notre système. Orbot (63) est une application mobile open-source qui agit comme un proxy permettant aux appareils Android d'utiliser le réseau Tor pour anonymiser leur trafic Internet. Il fonctionne comme une passerelle entre les applications installées sur l'appareil et le réseau Tor, assurant que toutes les communications sortantes passent par le réseau Tor avant d'atteindre leur destination. Dans notre système, Orbot joue un rôle clé en intégrant Tor à l'application mobile utilisée par les utilisateurs. Grâce à Orbot, l'application peut transmettre les transactions de géolocalisation via le réseau Tor sans nécessiter d'intégration complexe ou de modifications significatives de l'application mobile. Cette solution simplifie l'implémentation tout en garantissant un anonymat robuste pour les utilisateurs. En outre, Orbot permet une gestion plus flexible des connexions au réseau Tor, rendant son intégration viable même dans un environnement mobile. En utilisant Orbot, notre système exploite les avantages de Tor pour protéger les adresses IP des utilisateurs et s'assure que chaque transaction reste dissociée des identités numériques permanentes, tout en minimisant l'impact sur les performances des appareils mobiles.

4.3.4 Géo-pools

Définition (Géo-pool). Une géo-pool est l'unité logique de gouvernance et de traitement des données pour une zone géographique, qui regroupe (1) un canal Fabric local dédié, (2) les politiques d'accès/validation associées, et (3) les acteurs responsables de la zone. Dans le prototype, les géo-pools sont implémentés *via* des canaux Fabric hiérarchisés (locaux, interrégionaux, global). Les géo-pools constituent un composant

important de notre architecture décentralisée, conçu pour optimiser la gestion des données géographiques dans le réseau Hyperledger Fabric. Ce concept a évolué pour tirer parti des fonctionnalités avancées de Fabric, telles que la gestion des canaux, tout en répondant aux défis identifiés lors des premières phases de développement.

Initialement, les géo-pools étaient envisagés comme un moyen d'attribuer dynamiquement des sous-réseaux Tor à des zones géographiques spécifiques, afin de valider les données des utilisateurs par des pairs proches géographiquement. Cette approche s'inspirait des mécanismes de consensus comme la preuve d'enjeu (PdE), où les zones densément peuplées bénéficiaient naturellement d'un plus grand nombre de validateurs. Cependant, avec le choix d'Hyperledger Fabric comme base technologique et l'implémentation des identités jetables, cette conception a été repensée. La structure des canaux offerte par Fabric a permis de traduire l'idée des géo-pools en une méthodologie plus adaptée, où chaque organisation du réseau est responsable d'une zone géographique spécifique. Ce changement a également permis de mieux répondre aux besoins d'évolutivité et de gestion décentralisée.

Dans le prototype final, les géo-pools sont implémentés à travers la structure des canaux de Fabric, selon une hiérarchie bien définie :

- *Canaux locaux*. Chaque organisation gère un canal dédié à sa zone géographique, où sont traitées les transactions locales et les données spécifiques à cette région.
- *Canaux interrégionaux*. Des canaux supplémentaires permettent la communication entre les zones géographiques adjacentes, facilitant la coordination des données partagées entre elles.
- *Canal global*. Un canal centralisé est utilisé pour agréger les données provenant de toutes les zones, permettant la création de visualisations globales telles que des cartes thermiques (*heatmaps*) dans l'application de visualisation.

Les géopools offrent plusieurs avantages significatifs dans la solution actuelle. En attribuant des responsabilités spécifiques à chaque organisation pour la gestion de sa zone géographique, la gouvernance et la gestion des données sont clarifiées, permettant une meilleure répartition des tâches. La séparation des canaux allège la charge sur le réseau global, optimisant ainsi les performances en facilitant un traitement rapide et localisé des données. De plus, la structure actuelle permet une évolutivité naturelle : l'intégration de nouvelles zones géographiques peut être facilement réalisée par la création de nouvelles organisations et des canaux correspondants. Enfin, le canal global assure une vue d'ensemble des données pour des analyses et des visualisations globales tout en maintenant une gestion locale précise des informations. Dans

un déploiement réel, les organisations participantes pourraient aussi bien être des entreprises privées que des entités publiques telles qu'une municipalité, une université, une ONG ou encore un organisme gouvernemental.

Contrats Intelligents : Les contrats intelligents du prototype ont volontairement été conçus de manière minimaliste pour limiter le coût calculatoire. Ils couvrent trois fonctions principales : (1) l'enregistrement et la validation des transactions de géolocalisation, (2) la gestion et la rotation des identités jetables ainsi que (3) la publication d'agrégats statistiques pour la visualisation. Dans tous les cas, seules des empreintes cryptographiques (hashes) sont stockées sur la chaîne, tandis que les clés privées et les données sensibles demeurent sur l'appareil de l'utilisateur, assurant ainsi un haut niveau de confidentialité. Ce système sera décrit plus en détails dans la section 5.4.

4.3.5 Perturbation de la localisation

Même en intégrant une couche de confidentialité via l'utilisation d'identités temporaires dans notre système de géolocalisation décentralisé, le risque de détection de motifs, d'habitudes de déplacement stables et constants demeure. Ainsi, il est possible, par exemple, d'analyser des identités présentant des points de départ et d'arrivée récurrents pour tenter de déduire les lieux de résidence et de travail des utilisateurs (54). Afin de contrecarrer cette vulnérabilité, nous proposons d'introduire une perturbation des positions selon une fréquence d'injection faible, c'est-à-dire appliquée seulement sur une fraction limitée des transactions générées. L'objectif est de maintenir l'utilité globale du système (les trajectoires conservent leur cohérence) tout en réduisant la possibilité de déduire des points sensibles comme le domicile ou le lieu de travail(69).

Plusieurs méthodes peuvent être employées pour intégrer ce type de bruit dans les données (56). Premièrement, il est envisageable de demander aux utilisateurs d'enregistrer leurs adresses personnelles et professionnelles localement. Nous pourrions alors identifier les géoadresses les plus proches de ces points géographiques et sélectionner aléatoirement d'autres adresses dans un périmètre prédéfini, établissant ainsi de faux emplacements voisins. Ces faux emplacements seraient utilisés à chaque fois que l'utilisateur termine son trajet à son domicile ou lieu de travail, brouillant les pistes quant à ses véritables allées et venues.

Une autre approche consiste à «purger» certaines identités en générant un motif de déplacement pseudo-aléatoire dans l'environnement immédiat de l'utilisateur. Ceci pourrait être déclenché à un moment aléa-

toire, où l'identité en cours adopterait un itinéraire imprévu à travers la zone environnante jusqu'à l'épuisement de ses transactions disponibles avec l'identité jetable. Cette action forcerait l'utilisateur à créer une nouvelle identité, instaurant ainsi un schéma de mouvement dissocié de ses déplacements habituels. Cette stratégie pourrait s'avérer particulièrement utile pour les utilisateurs restant inactifs pendant de longues périodes, tels que ceux travaillant majoritairement depuis leur domicile. Néanmoins, il est crucial que la perturbation de la localisation n'affecte pas outre mesure l'utilité des données. Ainsi, les identités purgées pourraient par exemple être marquées pour éviter qu'elles ne faussent l'ensemble des données collectées.

Dans le cadre du développement du prototype, bien que l'injection de bruit soit une fonctionnalité en cours de développement, des tests ont été effectués sur deux méthodes dans les scripts de génération de données de simulation. La première méthode utilise deux variables contenant les géoadresses représentant le domicile et le lieu de travail. Lors de la génération des mouvements simulés, si ces adresses apparaissaient comme destination de la prochaine transaction, un motif pseudo-aléatoire était généré pour brouiller ces emplacements. La seconde méthode, visant à épuiser toutes les transactions restantes pour une identité, est actuellement partiellement opérationnelle. Son implémentation complète dépend de l'intégration d'OpenRouteService, un service ouvert de calcul d'itinéraires basé sur les données d'OpenStreetMap. qui permettra de générer des itinéraires réalistes tout en maintenant un niveau élevé de confidentialité.

Ces deux approches illustrent les efforts en cours pour développer des mécanismes robustes de perturbation de localisation, essentiels pour préserver l'anonymat des utilisateurs face à des analyses malveillantes visant à déduire leurs habitudes de déplacement.

Paramétrage de la perturbation. La génération de bruit sur les traces est contrôlée par trois paramètres complémentaires conçus pour offrir un compromis mesurable entre confidentialité et utilité des données. La *fréquence* p fixe la proportion de transactions affectées par une perturbation (par exemple : $p \in [0{,}005; 0{,}010]$). Un faible p introduit des « points d'ombre » suffisants pour compliquer l'identification de lieux sensibles (domicile, travail) sans altérer significativement les tendances agrégées. L'*intensité spatiale* $r_{\max}$ détermine l'amplitude maximale du déplacement artificiel autour du point sensible (ex. 100–300 m) : elle est choisie pour être suffisamment large pour masquer précisément un point d'intérêt tout en restant cohérente avec l'échelle urbaine et la résolution des géoadresses. Enfin, la *cohérence topologique* impose que toute trajectoire perturbée respecte le graphe routier (itinéraires réalistes) et limite l'allongement temporel induit (dilatation temporelle bornée, p. ex. $< 15\%$) afin de préserver la plausibilité des déplacements et l'utilité

pour les analyses mobilité.

Le paramétrage doit être vu comme une politique adaptable selon le contexte (zone urbaine vs périurbaine), le niveau de risque accepté et les objectifs analytiques. En pratique, p et $r_{\max}$ peuvent être ajustés de façon centralisée (par district) ou localement (préférences utilisateur), et évalués via métriques standards : perte d'utilité (ex. divergence entre heatmaps originales et perturbées), taux de réussite de ré-identification estimé, et métriques de cohérence topologique (distance/temps additionnels moyens). Chaque transaction perturbée porte un indicateur `noise_flag` dans le champ *meta* ; ce flag permet d'exclure ou de pondérer ces points lors du calcul des agrégats (p. ex. ne pas compter les transactions marquées dans certaines analyses statistiques), garantissant ainsi que les mécanismes de protection de la vie privée n'introduisent pas de biais imprévus dans les résultats finaux.

CHAPITRE 5
PROTOTYPE

Le prototype présenté dans ce mémoire est actuellement au stade de preuve de concept. Il a permis de démontrer la faisabilité technique des mécanismes de géolocalisation respectueux de la vie privée, intégrant des identités jetables, un réseau Hyperledger, et une application de visualisation. Cependant, plusieurs éléments restent à compléter et à valider dans des environnements réels. Notamment, des tests à grande échelle, l'intégration de scénarios utilisateurs variés, ainsi que l'optimisation des performances pour des charges de données plus importantes sont encore nécessaires. Ce niveau de maturité indique que le système est prêt pour des validations supplémentaires en environnement contrôlé avant un déploiement complet en production. L'architecture globale de ce système repose sur une intégration de plusieurs composants clés, chacun jouant un rôle vital dans la fourniture du service. Cette section décrit la structure générale du prototype, en mettant en lumière l'interaction entre ses divers éléments constitutifs et leur contribution à l'objectif global du projet.

5.1 Architecture générale

Le cœur du prototype est ancré dans la gestion des géoadresses, servant de fondation pour toutes les opérations. Ces données, centrées sur les réseaux routiers et les intersections, forment la base de la couche de données sur laquelle repose l'ensemble du système. L'utilisation de ces points géographiques spécifiques comme ancrages pour les transactions de géolocalisation garantit non seulement une représentation précise et adaptable de l'espace urbain, mais favorise également la confidentialité et le passage à l'échelle.

La collecte et le traitement des données de géolocalisation sont facilités par une application Android dédiée, qui sert de point d'interaction direct avec l'utilisateur. En exploitant les capacités de localisation des dispositifs mobiles, cette application capte les mouvements des utilisateurs, les traduisant en transactions anonymes au sein du réseau. La préservation de la confidentialité est renforcée par l'utilisation du réseau Tor, qui masque les détails de l'utilisateur, assurant ainsi une couche supplémentaire de sécurité.

L'infrastructure chaînes de blocs est batie sur Le réseau Hyperledger Fabric, organisée en une architecture de réseau décentralisée comprenant dix organisations correspondant à différents districts géographiques. Cette structuration privilégie une gestion des données localisée et efficiente, tout en permettant une ex-

pansion aisée du système. La communication entre ces organisations est orchestrée à travers divers canaux, facilitant ainsi l'interaction et l'échange de données tout en maintenant l'intégrité et la confidentialité des informations. Une application de visualisation client-serveur, basée sur Node.js et Angular, offre une interface pour l'interprétation et l'affichage des données de géolocalisation agrégées. Cette application fournit des visualisations intuitives, telles que des cartes thermiques, permettant aux utilisateurs de comprendre les tendances de mouvement sans compromettre la confidentialité individuelle.

Enfin, la validation de l'architecture et des fonctionnalités du prototype est assurée par l'utilisation de données de simulation, qui ont été formés à partir de données de localisation personelles receuillies a travers les premières itérations de l'application Android ainsi que des données produites a partir du dataset utilisé pour les géoadresses. Cette approche permet de tester l'efficacité, le passage à l'échelle et la confidentialité du système dans un environnement contrôlé, permettant ainsi de valider que le prototype est non seulement fonctionnel, mais aussi aligné avec les exigences de confidentialité et de performance. L'architecture globale du prototype est conçue pour être modulaire, flexible et évolutive, permettant une intégration facile de nouveaux composants et une expansion géographique sans perturber l'infrastructure existante.

5.2 Structuration des données de localisation pour les géoadresses

Les géoadresses constituent la couche fondamentale du prototype, conçu pour représenter les informations de géolocalisation dans un cadre décentralisé. Cette section décrit la justification, le processus et les résultats de la structuration des données de localisation pour les géoadresses basées sur le réseau routier, en soulignant son importance pour la fonctionnalité du prototype. L'utilisation des réseaux routiers, et en particulier des intersections, comme fondement pour les transactions de géolocalisation, est dictée par leur omniprésence dans les milieux urbains. Cette stratégie assure que le modèle est évolutif et flexible, adapté à une variété de configurations urbaines. En transformant les déplacements des utilisateurs en interactions avec des points prédéfinis (ici les intersections), le système simplifie la représentation des données tout en renforçant la confidentialité par l'agrégation des mouvements des utilisateurs en nœuds communs.

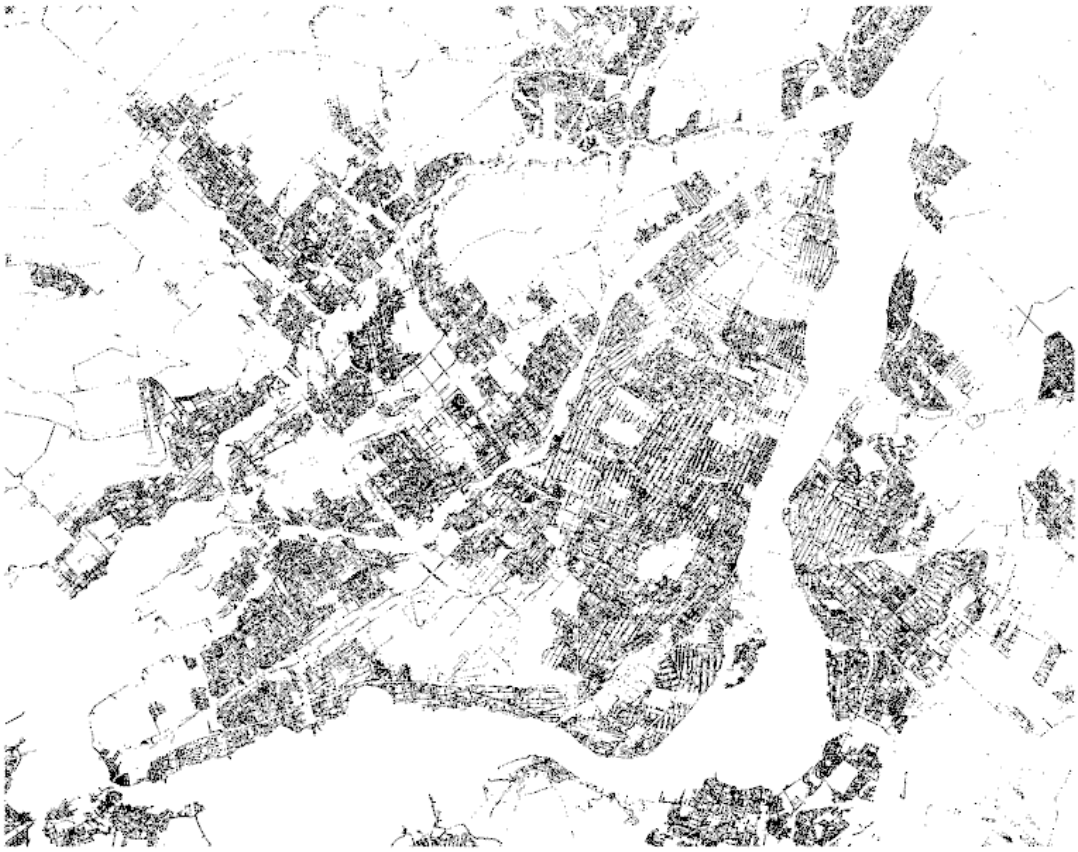

Figure 5.1 – Visualisation des géoadresses sous Gephi.

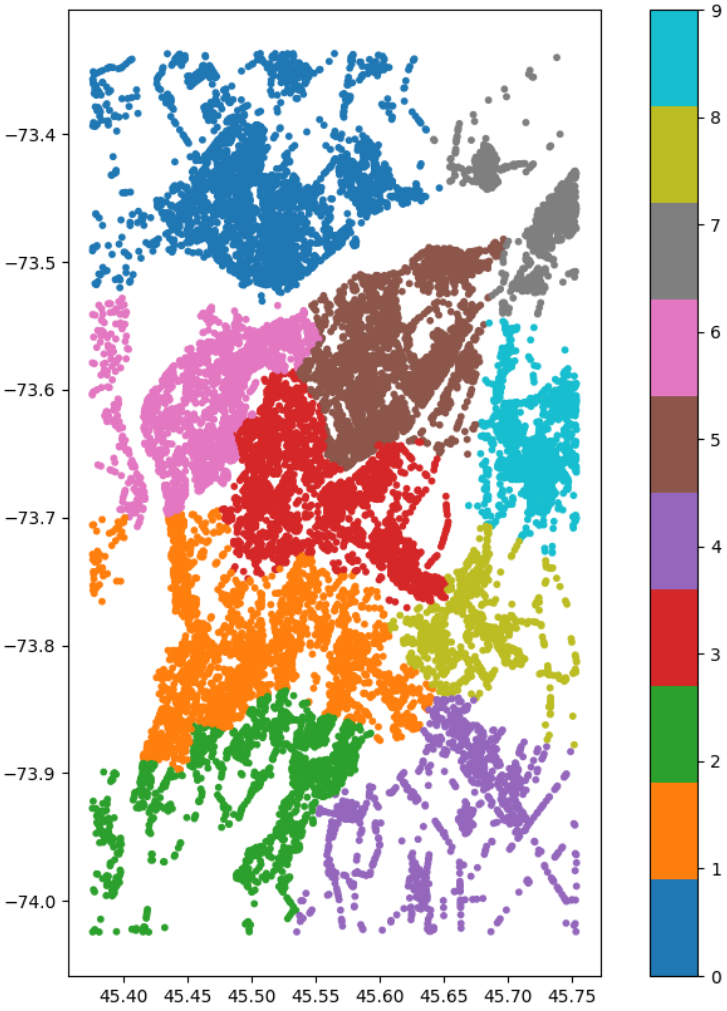

Figure 5.2 – Ensembles de géoadresses apres clustering.

Pour réaliser cela, nous avons d'abord extrait les données d'intersections de rues de la région de Montréal depuis Open Street Map (OSM) via l'API Turbo. Cette extraction visait à réunir un ensemble complet de données, y compris les coordonnées de latitude et de longitude de chaque intersection, couvrant ainsi toute l'île de Montréal et ses environs, comme illustré dans la figure 5.1. Ensuite, chaque point d'intersection, identifié par ses coordonnées, s'est vu assigner une adresse. Nous avons rencontré une difficulté lorsque, dans certaines zones, notamment celles avec de larges boulevards, plusieurs points représentaient en réalité un même emplacement physique. C'était par exemple le cas à l'intersection de Pie-IX et Sherbrooke, à Montréal, où la densité des croisements proches entraînait la génération de plusieurs géoadresses distinctes dans un rayon très réduit. Ce phénomène causait la création de boucles artificielles dans les trajectoires, tant lors des tests sur appareil mobile que dans les simulations.

Pour résoudre ce problème, nous avons appliqué l'algorithme DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) (25), un algorithme de clustering bien adapté aux données géographiques. Cette étape permet de regrouper les points trop rapprochés en une seule géoadresse représentative, réduisant ainsi la fragmentation des données et améliorant la cohérence des trajectoires. DBSCAN fonctionne en regroupant les points proches selon une distance définie (appelée *epsilon*). Il identifie les *core points*, qui ont au moins un nombre minimal de voisins (*minPts*) dans le rayon *epsilon*, et les regroupe en clusters. Les points trop éloignés de tout cluster sont classifiés comme anomalies (*outliers* en anglais). Dans notre cas, nous avons utilisé un rayon de 50 mètres pour regrouper les intersections proches représentant un même emplacement physique, réduisant ainsi la redondance des données et affinant la représentation géographique de Montréal.

Pour rappel, les géoadresses ont été organisées en dix districts, une structuration qui a guidé l'architecture du réseau Hyperledger Fabric. En effet, cette division facilite la gestion des données, s'alignant sur l'approche décentralisée du prototype et favorisant un traitement efficace et une accessibilité des données. La méthodologie mise en place pour les géoadresses assure l'adaptabilité du système à différents contextes urbains. En s'appuyant sur les réseaux routiers comme indicateur universel de la géographie urbaine, le système peut aisément intégrer de nouvelles zones en traitant les données OSM locales, tout en conservant la logique sous-jacente.

Bien que l'implémentation actuelle soit seulement un prototype, des recherches futures pourraient se pencher sur des algorithmes plus sophistiqués pour la génération d'adresses et le regroupement de points.

Perfectionner ces éléments pourrait améliorer les performances du système et la précision des données, élargissant ainsi les possibilités d'application et offrant une couverture urbaine plus complète.

5.3 Application Android

L'application Android joue un rôle central dans la facilitation des interactions des utilisateurs avec le service de géolocalisation décentralisé. Elle repose sur trois fonctions opérationnelles essentielles : le traitement des données de géolocalisation, l'utilisation de Tor pour préserver la confidentialité et la connexion au réseau hyperledger. Chacune de ces fonctions est vitale pour assurer la confidentialité et l'efficacité du service, comme décrit dans la figure 5.3, qui illustre les principales classes et interactions du système.

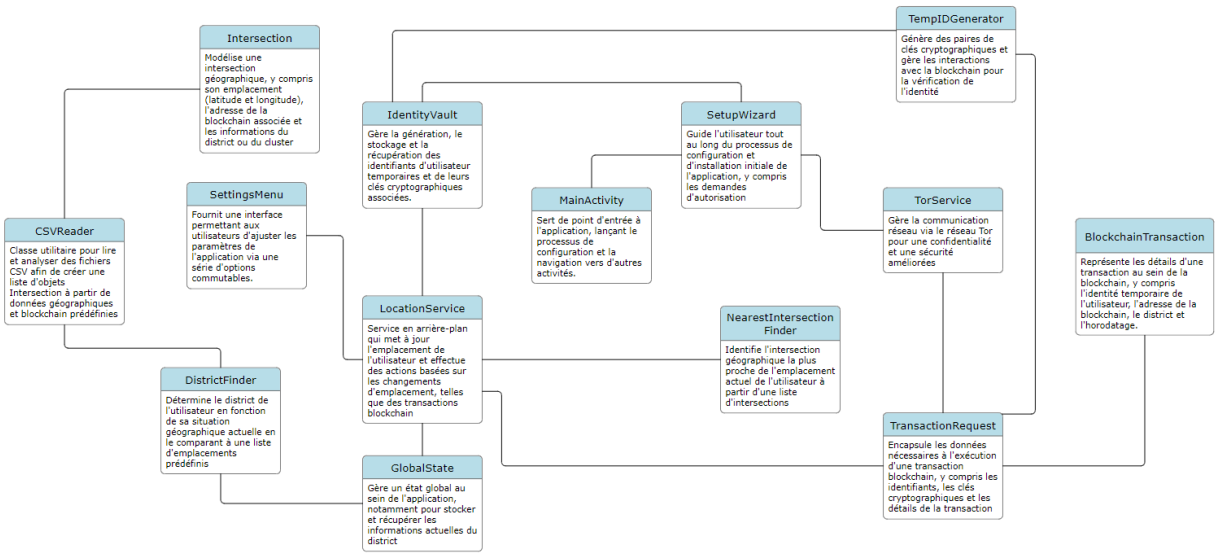

Figure 5.3 – Structure de l'application Android.

La fonction principale de l'application consiste à acquérir et à traiter avec précision les données de géolocalisation de l'utilisateur. Cela inclut la détection de mouvement, où l'application suit en permanence la localisation de l'utilisateur pour identifier ses déplacements, la classe `LocationService`, comme représentée dans la figure 5.3, est responsable de la mise à jour en temps réel de l'emplacement de l'utilisateur. En se référant aux géoadresses, l'application compare ces mouvements à l'adresse la plus proche représentant une intersection de rue. À partir de ces emplacements cartographiés, l'application exécute des transactions via la classe `TransactionRequest`, qui encapsule les données nécessaires à l'interaction avec le réseau Hyperledger, convertissant ainsi les mouvements réels en une série d'interactions avec des points prédéfinis et améliorant la confidentialité des utilisateurs.

Pour rappel, l'application anonymise le trafic Internet via le réseau Tor, dissimulant l'adresse IP de l'utilisateur pour prévenir tout suivi ou profilage basé sur l'activité réseau. Cette fonctionnalité est prise en charge par la classe `TorService`, illustrée dans la figure 5.3, qui gère les communications réseau via le réseau Tor pour garantir une confidentialité accrue. Par exemple, on peut voir dans la figure 5.4 l'IP du nœud d'entrée dans le réseau Tor pour l'utilisateur, ainsi que le nœud de sortie avec une IP différente dans la figure 5.5.

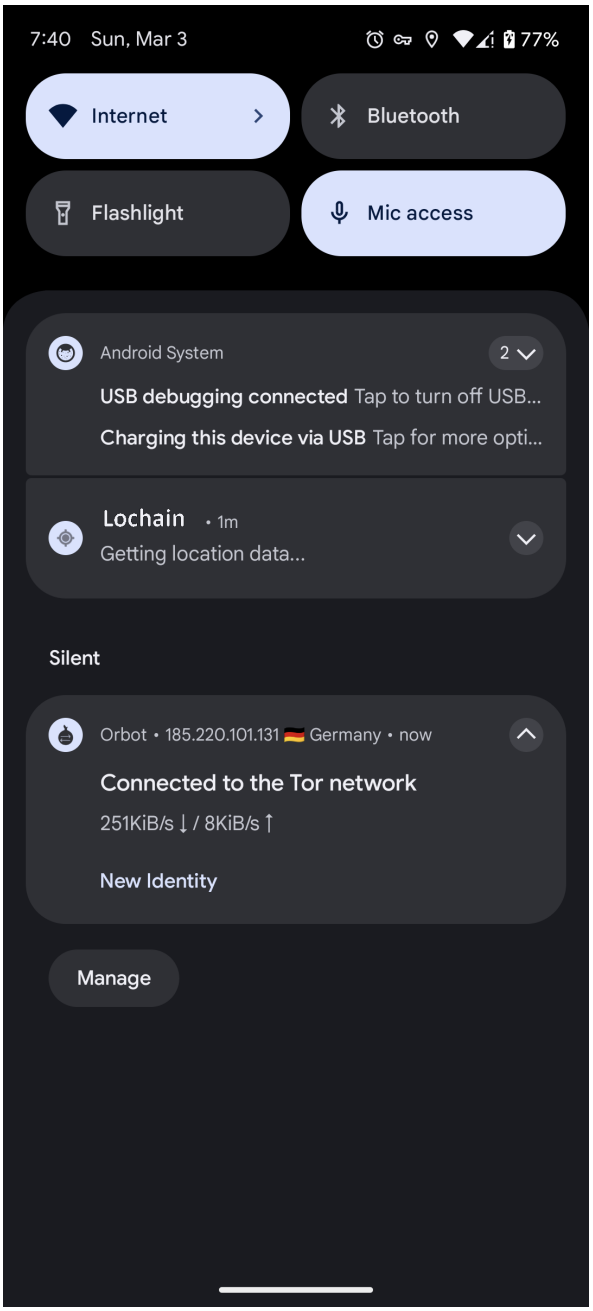

Figure 5.4 – Notifications Application Android.

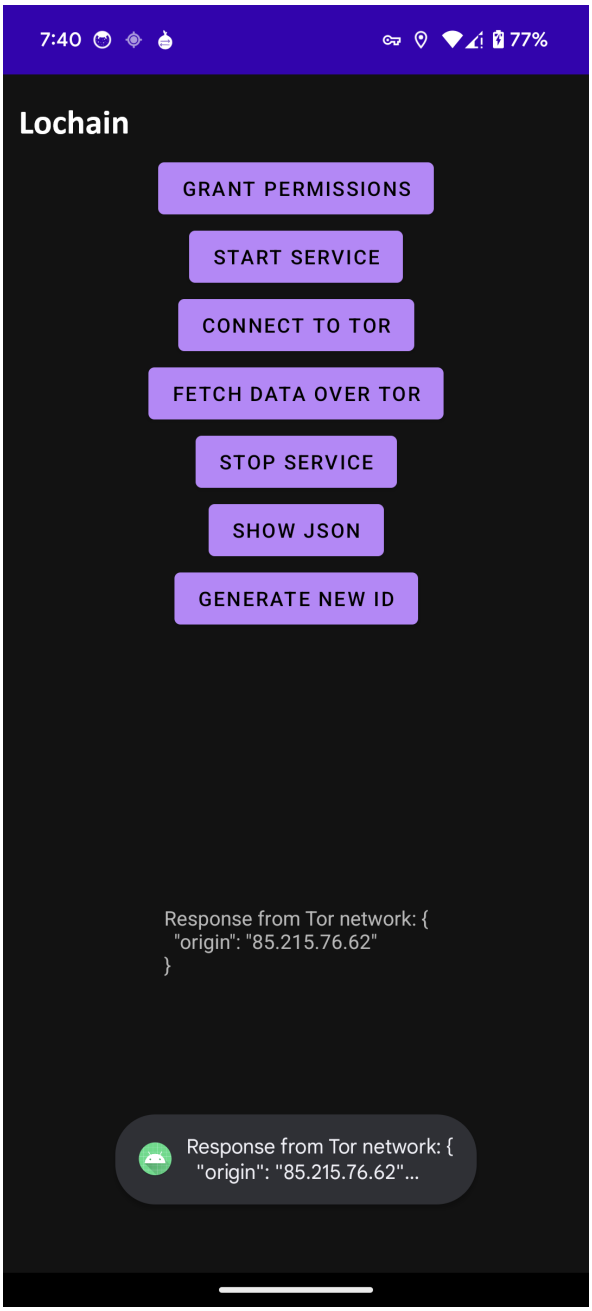

Figure 5.5 – IP Noeud de sortie.

L'autre fonction clé de l'application Android est d'agir comme un lien entre l'utilisateur et le réseau Hyperledger. Ceci est réalisé grâce à l'intégration du SDK Hyperledger Fabric via la classe `MainActivity`, qui coordonne les interactions entre l'utilisateur et les autres composants comme `IdentityVault` et `TempIDGenerator`. Cette intégration facilite la soumission des transactions de géolocalisation et la gestion des identités jetables dans l'architecture de chaîne de blocs. De plus, l'application propose une interface utilisateur intuitive, permettant aux utilisateurs de visualiser leur historique et de gérer leur identité numérique comme illustré dans la figure 5.6 où l'on peut voir les informations concernant une identité temporaire, ses composants principaux étant l'identifiant lui-même, la paire de clés privé/publique, les districts avec lesquels des interactions ont été effectuées pour faciliter l'exécution des contrats intelligents, l'heure de création ainsi que le nombre de transactions avant que l'identité ne soit plus utilisable.

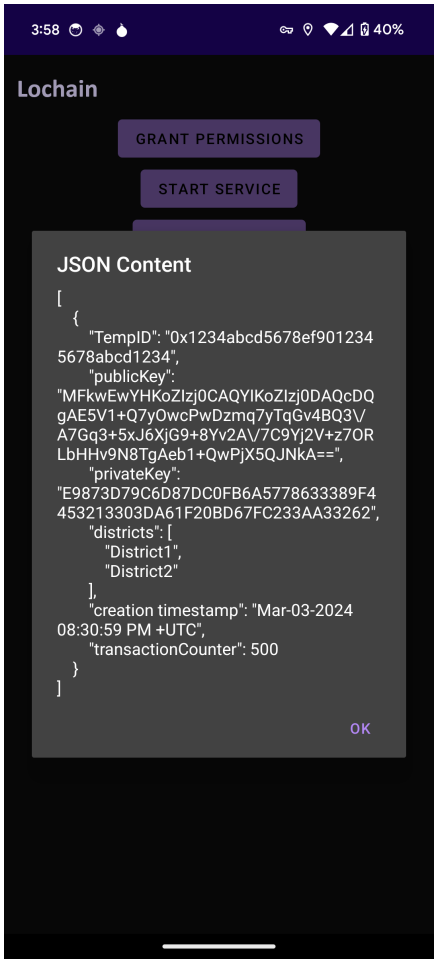

Figure 5.6 – Contenu du coffre d'identité.

En somme, l'application Android est conçue pour être non seulement fonctionnelle, mais aussi pour garantir l'anonymat et la sécurité des données des utilisateurs, des aspects cruciaux dans un contexte dans lequel la protection de la vie privée numérique est de plus en plus primordiale.

5.4 Réseau Hyperledger

Le prototype du service de géolocalisation exploite Hyperledger Fabric pour créer un cadre de réseau décentralisé comprenant dix organisations, chacune correspondant à un district distinct au sein de la zone géographique modélisée. Cette structure prend en charge une expansion évolutive et une gestion coopérative entre différentes juridictions sans contrôle centralisé, ce qui est crucial pour maintenir l'intégrité et la flexibilité du système. Plus précisément, le fondement de la structure organisationnelle du réseau repose sur les géoadresses dérivées des intersections routières, facilitant une approche adaptable et évolutive des services de géolocalisation. Cet alignement garantit qu'à mesure que de nouveaux districts sont ajoutés, les organisations correspondantes peuvent être intégrées de manière fluide et transparente dans le réseau.

Selon nos données, nous nous retrouvons donc avec 10 organisations comme indiqué dans la figure 5.2, Il s'agit d'un compromis pratique permettant de démontrer la gestion multi-zones et l'utilisation de canaux Fabric distincts, tout en maintenant la complexité du déploiement à un niveau raisonnable pour un prototype expérimental. Un nombre inférieur de districts aurait réduit l'intérêt de la démonstration en ne montrant qu'une décentralisation très limitée, tandis qu'un nombre beaucoup plus élevé aurait entraîné une surcharge technique difficile à gérer dans le cadre des ressources disponibles. Il convient de souligner que ce paramètre est entièrement ajustable : dans un déploiement réel, le nombre de districts pourrait être adapté à la taille du territoire et aux besoins des organisations participantes.

Une fonctionnalité intéressante à exploiter d'Hyperledger Fabric est le système de canaux, qui ont des registres qui leur sont propres, cela nous permet d'organiser le flux d'informations de manière à avoir un traitement localisé des données, chaque organisation représentant une zone utilisant trois catégories de canaux décrites ci-après.

- *Canal d'ingestion de données*. Ce canal est utilisé afin d'enregistrer les transactions qui représentent les mouvements des utilisateurs dans chaque district, sécurisant ainsi la localisation des données, une instance unique de ce canal peut exister pour chaque organisation.
- *Canaux périphériques*. Étant donné que chaque organisation représente une zone géographique, il est possible d'avoir une instance de ces canaux pour chaque organisation avoisinante. Ils facilitent

la communication entre districts, agissant comme des registres pour les interactions inter-districts transparentes et vérifiables.

— *Canal global*. Ce canal centralise l'agrégation et la mise à jour périodique des données de mouvement pour la visualisation, tout en préservant l'anonymat des utilisateurs et en assurant une cohérence des données à travers le système. Toutes les organisations y participent.

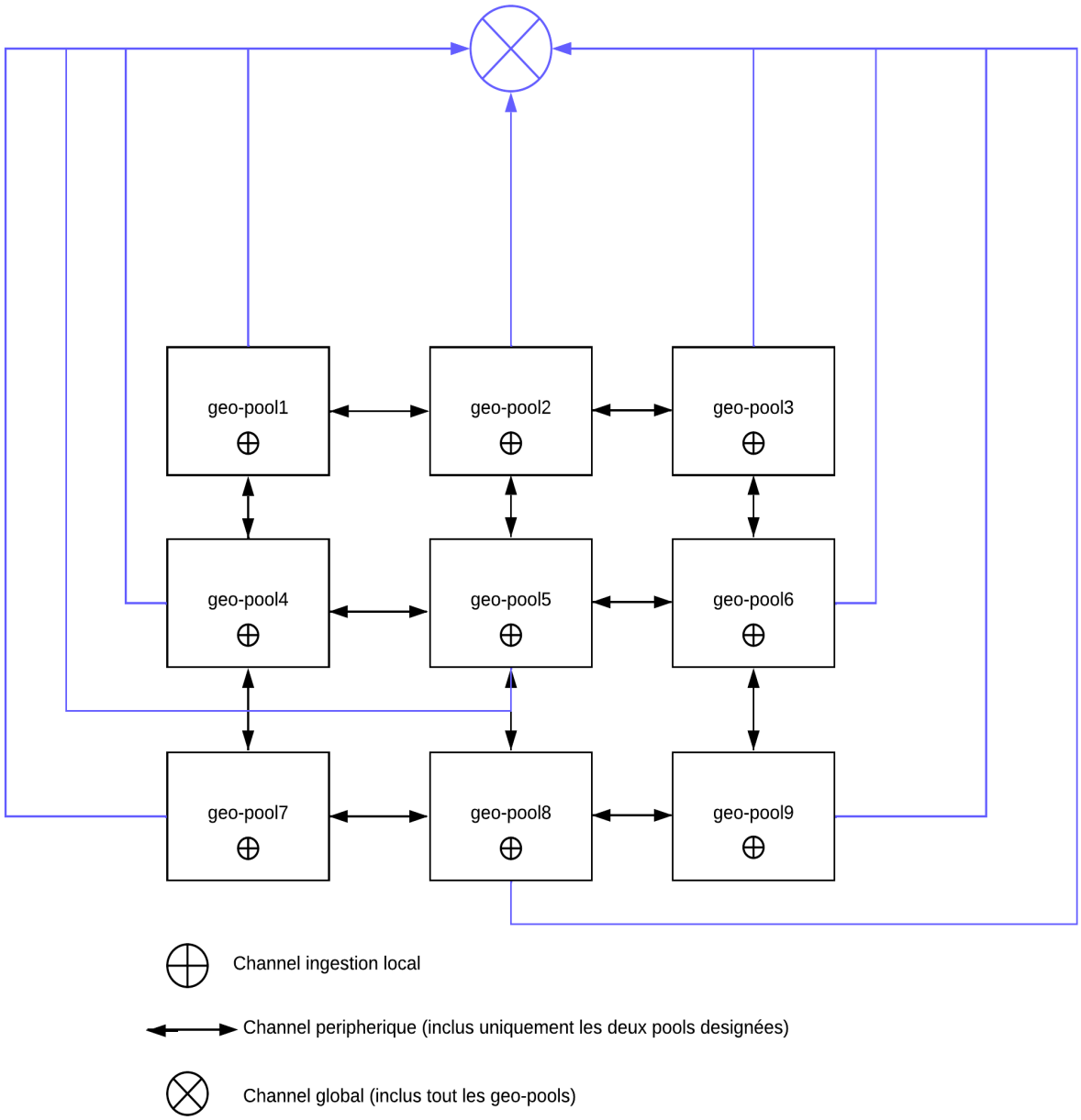

Figure 5.7 – Organisation des géo-pools dans le prototype : chaque géo-pool dispose de son canal local (⊕), est relié à ses voisins par des canaux périphériques (↔) et participe au canal global (⊗) qui agrège toutes les données.

Le réseau emploie des contrats intelligents pour la gestion des identités, l'enregistrement des transactions et la visualisation des données, favorisant une approche minimaliste du côté de la chaîne de blocs pour maintenir l'efficacité et la rapidité du système :

— *Contrats de gestion des identités*. Ces contrats automatisent la création et le renouvellement des pools d'identités jetables et facilitent les revendications des utilisateurs sur ces identités. Ce système favorise un juste équilibre entre l'autonomie utilisateur et la génération contrôlée d'identifiants, essentielle pour la confidentialité.
— *Contrat d'enregistrement des transactions*. Cette étape effectue l'enregistrement des transactions de géolocalisation dans le canal d'ingestion, reflétant l'engagement du système pour la simplicité opérationnelle.
— *Contrats de visualisation de données*. Ils soutiennent l'accès respectueux de la confidentialité aux statistiques de mouvement, permettant une visualisation des données individuelles et agrégées sans compromettre l'identité des utilisateurs.

Cette structure est relativement simple à comprendre dans le cadre de ce prototype où le nombre d'organisations est limité à 10, néanmoins il est important de prendre en compte les capacités de mises à l'échelle et les possibles difficultés qui pourraient être rencontrés à ce niveau (différences importantes de trafic entre différentes zones, goulot d'étranglements, etc.), il existe des techniques visant à répondre à ces problématiques, l'une d'entre elle étant la fragmentation (section 2.9).

Cette solution est particulièrement intéressante à prendre en compte dans notre contexte, grâce à notre structure d'organisation et de canaux, qui ont chacun leurs registres qui leur sont propres, le fait que nos données soient stockées localement, notre solution émule la fragmentation, de plus un des aspects les plus complexes à gérer dans les systèmes utilisant la fragmentation est la communication inter-fragments. Cette difficulté dépend directement du type de données traitées. Contrairement aux systèmes financiers, où une transaction peut nécessiter une interaction entre fragments distants dans le réseau, notre système de géolocalisation se distingue par la nature locale de ses données.

La probabilité qu'une organisation ait besoin de communiquer avec une autre dépend principalement de sa localisation géographique et des zones qu'elle partage avec ses organisations voisines. Par conséquent, les interactions sont généralement limitées aux fragments adjacents. Dans ce contexte, les canaux périphériques de notre architecture remplissent efficacement le rôle de mécanismes de communication inter-fragments. Ces canaux permettent de gérer les échanges entre zones voisines tout en garantissant la cohérence et la confidentialité des données.

5.4.1 Coût calculatoire (ordre de grandeur)

L'évaluation du coût calculatoire repose sur des ordres de grandeur issus de la littérature académique et des rapports de performance concernant les signatures ECDSA et Hyperledger Fabric. Côté client (Android), la signature ECDSA sur la courbe P-256 est réalisée en environ 0.3 à 2 ms sur un téléphone de gamme intermédiaire, tandis que la sérialisation et l'écriture via TLS ajoutent un surcoût négligeable de l'ordre de 0.1 à 0.5 ms (57). Du côté des pairs Fabric, la vérification d'une signature ECDSA prend en moyenne 0.2 à 1.0 ms par transaction sur un processeur moderne, et les opérations de lecture/écriture dans l'état (CouchDB ou LevelDB) se situent généralement entre 1 et 5 ms par transaction selon le backend choisi (35). Enfin, l'ordonnancement via Raft introduit un coût dépendant de la configuration du système : dans les scénarios typiques, l'assembleur produit des blocs de 10 à 100 transactions avec une latence de 50 à 500 ms, en fonction des paramètres `BatchTimeout` et `BatchSize` (8). Ces valeurs doivent être considérées comme des ordres de grandeur représentatifs, servant à situer la charge relative des différents composants, plutôt que comme des mesures précises de performance du prototype.

Les mesures sont issues de la littérature et de rapports de performance (benchmarks Fabric, ECDSA sur terminaux mobiles). Dans le cadre du prototype, ces chiffres doivent être considérés comme des ordres de grandeur académiques. Nous privilégions la publication de distributions (médiane/p95) plutôt que de valeurs ponctuelles. Ainsi, la signature et vérification cryptographique coûtent quelques millisecondes au plus, ce qui est négligeable comparé aux délais réseau. Le vrai facteur limitant reste la latence de communication (Tor, Fabric) plutôt que le coût calculatoire pur.

5.5 Application de visualisation

L'application de visualisation est conçue dans le but de rendre les données de géolocalisation non seulement accessibles, mais également intelligibles pour les utilisateurs. Elle repose sur une architecture client-serveur, utilisant Node.js pour le backend et Angular pour l'interface utilisateur, exploitant ainsi le SDK Hyperledger Fabric pour présenter de manière efficace les données complexes de localisation stockées dans la chaîne de blocs à travers des visualisations intuitives. L'un des avantages de cette architecture est la possibilité d'avoir une instance du backend déployée avec chaque nœud de chaque organisation, cela dans l'optique générale de localisation des données et d'équilibrage de la charge. L'application adopte une conception minimaliste, limitant les interactions utilisateur aux fonctionnalités essentielles telles que la barre de recherche et l'option de basculement entre les modes de visualisation tels qu'illustrés dans la figure 5.8. Cette approche

garantit une expérience utilisateur fluide et accessible, facilitant la navigation et l'interprétation des données présentées sans exiger de connaissances techniques approfondies.

L'application propose deux modes principaux de visualisation des données. Tout d'abord le mode identité, illustré dans la figure 5.8, permet à l'utilisateur d'effectuer une recherche basée sur une identité temporaire. En saisissant une identité dans la barre de recherche, l'utilisateur peut visualiser toutes les transactions associées à cette identité spécifique. Ce mode est particulièrement utile pour examiner les interactions ou mouvements d'une identité temporaire donnée dans le système. Ensuite le mode heatmap, présenté dans la figure 5.9, utilise les données agrégées du canal global pour générer une carte thermique des déplacements dans différentes zones géographiques. Ce type de visualisation met en évidence les zones les plus fréquentées, en offrant une vue statistique sans compromettre la confidentialité individuelle.

Gestion des identités temporaires. Il est important de préciser que le système ne permet pas à un utilisateur de découvrir ou d'explorer les identités temporaires d'autres participants. Chaque utilisateur conserve uniquement, sur son appareil, les informations relatives aux identités temporaires qui lui ont été attribuées. Les requêtes effectuées via l'application de visualisation se limitent donc strictement à ses propres identités. Par ailleurs, les identités temporaires ne sont pas réutilisables. Bien que leur réutilisation partielle pourrait sembler, à première vue, brouiller davantage les pistes, cela irait à l'encontre d'un des objectifs principaux du système : garantir que toutes les traces générées soient de taille équivalente afin de limiter les attaques par inférence fondées sur des différences de longueur des trajectoires. La non-réutilisation des identités constitue ainsi une condition essentielle pour maintenir l'homogénéité des traces et préserver la robustesse du mécanisme de confidentialité.

Pour aborder les enjeux de confidentialité, l'application impose un délai de 24 heures avant de rendre les données de localisation accessibles. Cette mesure, couplée aux stratégies de préservation de la confidentialité du système comme l'usage d'identités temporaires et l'intégration de Tor, assure que le suivi des mouvements individuels est fortement limité, contribuant ainsi à protéger la vie privée des utilisateurs. Ainsi, l'application met en œuvre un processus d'agrégation des données pour créer des visualisations sous forme de cartes thermiques en mettant un accent particulier sur l'anonymisation et la précision. En se basant sur des statistiques agrégées diffusées via le canal global, elle permet de révéler les tendances d'activité sans compromettre la confidentialité individuelle.

Optimisations techniques du backend.

L'architecture modulaire permet le déploiement d'une instance backend Node.js avec chaque nœud chaîne de blocs de district, offrant d'importantes opportunités d'optimisation détaillées ci-dessous :

- **Prétraitement et agrégation localisés.** Lors de la publication des données agrégées sur le canal global, chaque instance backend agrège localement les données brutes de géolocalisation de l'organisation, résumant les statistiques de déplacement avant la soumission sur la chaîne de blocs. Cela réduit considérablement la charge inutile des transactions chaîne de blocs, optimise les performances des nœuds de validation et permet une correction locale des erreurs.
- **Mise en cache décentralisée pour la visualisation.** Les backends Node.js maintiennent des caches locaux des visualisations fréquemment demandées. Cela réduit considérablement la latence des requêtes, fournissant des visualisations quasi-instantanées tout en minimisant les requêtes chaîne de blocs et la charge du réseau.
- **Soumissions en lot asynchrones.** Les transactions sont mises en tampon et soumises de manière asynchrone par lots (ex. : toutes les dix minutes ou après un nombre fixe de transactions). Cette technique de mise en lot réduit considérablement les surcharges transactionnelles, améliorant ainsi l'évolutivité de la chaîne de blocs.
- **Validation distribuée et contrôle de qualité des données.** Grâce à la structure standardisée des données exploitant les géoadresses, les instances backend locales effectuent des vérifications de validation (ex. : détection d'anomalies, cohérence des données) avant la soumission sur la chaîne de blocs. Cela réduit les erreurs dans le registre chaîne de blocs, préservant la qualité des données et minimisant les inefficacités de stockage.
- **Communication pair-à-pair (P2P).** Les backends Node.js communiquent directement avec les nœuds géographiquement adjacents, synchronisant les mises à jour récentes localement afin de réduire les interactions redondantes avec la chaîne de blocs et la congestion des nœuds de validation globaux.

Ensemble, ces optimisations mettent en évidence les avantages pratiques et l'importance stratégique de l'architecture de visualisation Node.js/Angular de LoChain.

Les aspects de sécurité liés aux API et à l'architecture backend, ainsi que les méthodes de gestion des fortes charges de trafic, nécessitent un développement ultérieur. Néanmoins, la structure actuelle du prototype,

avec une instance backend associée à chaque nœud réseau, pose les bases pour une évolution et un traitement efficace des données distribuées.

Figure 5.8 – Mode identité : Visualisation des transactions associées à une identité temporaire.

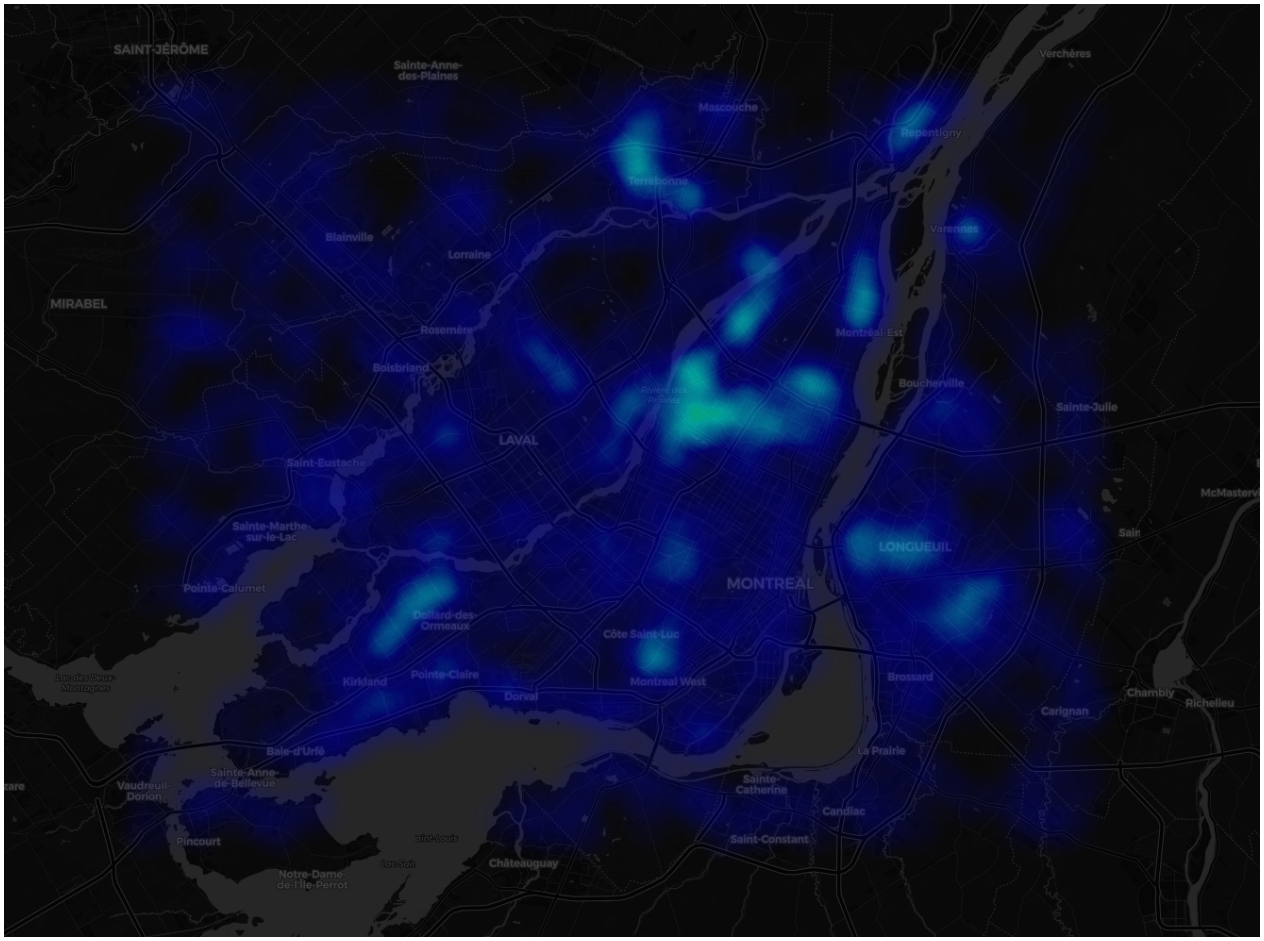

Figure 5.9 – Mode heatmap : visualisation agrégée des statistiques de déplacement.

5.6 Données de simulation

Le développement d'un service de géolocalisation qui préserve la confidentialité est un processus qui nécessite des tests et validations pour s'assurer de l'exactitude, de l'efficacité et de la conformité avec les principes de confidentialité. Cette partie aborde l'approche adoptée pour générer et utiliser des données de simulation, essentielles à la validation du système et à l'évaluation de ses performances.

La première étape de ce processus a consisté à collecter des données de géolocalisation réelles en enregistrant mes propres mouvements à des intervalles réguliers lors du développement de la première version de l'application Android. Ces données ont été stockées dans un fichier de référence. Initialement, l'idée était de proposer à plusieurs personnes d'installer l'application pour générer des jeux de données plus diversifiés. Cependant, en raison des incertitudes quant à la nécessité de soumettre cette démarche à un comité d'éthique, la décision a été prise de limiter la collecte de données à mes propres déplacements pour plus de simplicité et de conformité. Ces données personnelles ont servi de base pour valider la capacité de l'application Android à enregistrer les données, tout en fournissant un ensemble de données authentiques pour la validation du système. En prenant ces données comme base pour comparer les patterns et afin de créer un environnement d'utilisation réaliste, nous avons simulé une base de 10 000 utilisateurs générant chacun 1 000 transactions.

La génération des profils utilisateurs a attribué à chaque utilisateur simulé un ensemble d'identité jetables, conforme aux mécanismes précédemment décrits, et réparti les utilisateurs de manière aléatoire entre les différents districts géographiques modélisés. Leurs trajectoires ont été simulées en se basant sur les réseaux routiers extraits d'OpenStreetMap, où les intersections routières servaient de points de départ et d'arrivée pour les déplacements. À chaque étape, une transaction de géolocalisation était générée, incluant l'identité jetable, la géoadresse correspondant à l'intersection atteinte, et un horodatage précis, assurant un enregistrement cohérent et précis des mouvements simulés.

Pour évaluer la cohérence de ces traces, une vérification a été effectuée à deux niveaux. Dans un premier temps, les trajectoires issues des données personnelles ont été comparées aux transactions correspondantes générées en géoadresses, afin de s'assurer que la traduction était fidèle aux mouvements réels. Dans un second temps, les données simulées ont été soumises à des règles de continuité temporelle et spatiale : une transaction générée à un instant t ne peut être suivie d'une transaction à $t + 1$ située à plusieurs kilomètres de distance, sans qu'un chemin routier logique existe entre les deux. Ces contraintes

ont garanti que les trajectoires simulées restaient plausibles à grande échelle, même lorsqu'il n'était plus possible de vérifier manuellement l'ensemble du million de transactions générées.**

Cette simulation à grande échelle a été conçue pour évaluer la capacité du système à gérer de gros volumes de données, ce qui est crucial pour tester l'évolutivité et les performances de l'application de visualisation. L'agrégation des données de géolocalisation simulées pour créer des cartes thermiques a illustré comment le système peut présenter les tendances de déplacement de la population de façon anonyme et sécurisée. Ce processus d'agrégation, qui comptabilise les occurrences aux emplacements géographiques spécifiques, valide la fonctionnalité de l'application de visualisation et confirmé l'efficacité de la méthode d'anonymisation des données.

Pour valider davantage ces fonctionnalités, plusieurs scénarios d'évaluation ont été définis. L'un des scénarios principaux a consisté à analyser la capacité du système à agréger les transactions simulées de manière cohérente tout en préservant la confidentialité des utilisateurs. La précision des visualisations agrégées a été mesurée en comparant les heatmaps générées à partir des données simulées avec les modèles prévus, afin d'évaluer le taux de concordance entre les données agrégées et les emplacements géographiques réels. De plus, pour vérifier que les identités jetables étaient bien équivalentes en termes de traces, deux stratégies d'indexation ont été testées. La première a indexé les transactions par identités jetables, tandis que la seconde les a classées en fonction du nombre de transactions effectuées à chaque géoadresse. Les résultats obtenus ont permis d'évaluer la cohérence des traces générées et la robustesse du système face à ces critères.

Bien que le processus de simulation ait offert des aperçus importants sur la fonctionnalité du système, les questions relatives au réalisme des mouvements et à l'applicabilité dans des conditions réelles ont mis en lumière la nécessité d'améliorations futures. Le modèle de mobilité utilisé dans la simulation actuelle repose sur les intersections routières extraites d'OpenStreetMap comme points de départ et d'arrivée des déplacements. Cela suggère implicitement que les utilisateurs se déplacent principalement en véhicule ou suivent les itinéraires routiers. Cependant, ce choix peut limiter la représentation d'autres types de mobilité, comme les déplacements à pied, à vélo ou par les transports en commun. Il est cependant pertinent de noter que l'approche par géoadresses n'est pas limitée aux trajets routiers. Dans les réseaux piétonniers ou cyclables, où les intersections sont moins fréquentes, il serait possible de définir des géoadresses à partir d'autres points d'ancrage pertinents, tels que les entrées de parcs, les croisements de pistes cyclables, ou

encore les arrêts de transport collectif. L'ensemble S des géoadresses pourrait ainsi être adapté au type de mobilité visé, tout en conservant les mêmes propriétés d'uniformité et de confidentialité. Cette extension constituerait une piste intéressante pour de futurs travaux, afin de rendre la méthode applicable à une plus large variété de modes de déplacement.

5.7 Modèle de menaces

L'établissement d'un modèle de menaces pour notre service de géolocalisation décentralisée est essentiel pour identifier, comprendre et atténuer les risques potentiels à la sécurité et à la vie privée : Afin de structurer cette analyse, nous nous appuyons sur le cadre théorique de Dolev-Yao (22), couramment utilisé dans la littérature pour modéliser les menaces dans les systèmes distribués. Ce modèle suppose un adversaire puissant capable d'intercepter, modifier et réinjecter tout message circulant sur le réseau, mais incapable de casser les primitives cryptographiques considérées sûres (p. ex. ECDSA, SHA-256). Dans certains cas, nous complétons cette perspective par une catégorisation inspirée du modèle STRIDE de Microsoft (47), qui distingue les menaces en six familles (usurpation d'identité, altération, répudiation, divulgation d'information, déni de service et élévation de privilèges). Ces références fournissent une base théorique solide à notre modèle de menaces, garantissant que l'analyse repose sur des cadres établis plutôt que sur une simple intuition.

1. *Identification des actifs.* Notre initiative s'appuie sur des éléments fondamentaux nécessitant une protection rigoureuse, notamment les données de localisation des utilisateurs, leurs identités temporaires et les mécanismes de sécurité associés, l'infrastructure Hyperledger ainsi que l'application de visualisation. Ces actifs sont vitaux pour assurer l'intégrité et la confidentialité de notre service.
2. *Définition et catégorisation des menaces.* Les menaces peuvent être catégorisées en fonction des acteurs malveillants potentiels et de leurs objectifs :
   — *Acteurs externes (hackers et criminels)* : Ces individus ou groupes extérieurs au système cherchent à accéder illégalement aux données pour des activités malveillantes, comme le suivi ou le profilage des utilisateurs, la revente de données sensibles ou l'extorsion. Ils exploitent souvent des vulnérabilités techniques ou des failles humaines (ingénierie sociale) pour atteindre leurs objectifs.
   — *Entités utilisatrices (entreprises ou organisations)* : Bien que ces entités aient un accès légitime au système, elles peuvent tenter d'exploiter les données au-delà de leurs intentions initiales, par

exemple pour un usage commercial non autorisé, comme la publicité ciblée ou le développement de services concurrents. Leur objectif est généralement financier ou stratégique

- *Utilisateurs mal intentionnés* : Ces individus, bien qu'étant des utilisateurs du système, agissent de manière abusive. Par exemple, ils peuvent tenter de manipuler les données de géolocalisation pour des gains personnels, comme modifier leurs traces pour des avantages frauduleux, ou perturber le fonctionnement du système en générant des données incorrectes.
- *Failles internes (erreurs et vulnérabilités du système)* : Ces menaces ne proviennent pas d'un acteur spécifique, mais résultent de problèmes liés à la conception, au déploiement ou à la maintenance du système. Elles incluent les erreurs de configuration, les failles dans le code ou des politiques de sécurité inadéquates, qui peuvent être exploitées par d'autres acteurs malveillants.

3. *Analyse des vecteurs d'attaque.* Il est possible de supposer plusieurs vecteurs d'attaque clés, notamment l'interception des transactions, la compromission des identités temporaires, les attaques ciblant l'infrastructure de la chaîne de blocs, l'exploitation des données agrégées et les failles de sécurité dans l'application de visualisation.
   - Interception et modification des transactions : Les communications entre l'application utilisateur et la chaîne de blocs peuvent être interceptées ou modifiées par des acteurs malveillants, compromettant l'intégrité des données.
   - Déchiffrement des identités temporaires : Les identités temporaires et leurs clés privées pourraient être compromises si l'appareil utilisateur est perdu ou volé. Cela mettrait en danger l'anonymat des utilisateurs.
   - Attaques contre l'infrastructure de la chaîne de blocs : Cela inclut les attaques de déni de service (DDoS), la manipulation d'identités, et d'autres formes d'attaques visant à perturber le réseau, compromettre l'intégrité des données ou réduire la disponibilité du service.
   - Exploitation des données agrégées : Les données anonymisées publiées, telles que les heatmaps, pourraient être utilisées pour déduire des informations sensibles ou profiler les mouvements des utilisateurs.
   - Failles de sécurité dans l'application de visualisation : Des vulnérabilités dans l'application pourraient permettre un accès non autorisé ou une manipulation des données stockées ou affichées.
4. *Stratégies de mitigation.* Les stratégies suivantes sont mises en œuvre pour répondre aux attaques identifiées. Chaque stratégie est alignée sur des menaces spécifiques :

5.7.0.1 Stratégies de mitigation et limites

Pour contrer les vecteurs d'attaque identifiés, plusieurs stratégies de mitigation ont été intégrées dans le prototype. Tout d'abord, le chiffrement systématique des données en transit via TLS garantit que les communications entre l'application utilisateur et la chaîne de blocs ne peuvent être ni interceptées ni altérées. Ensuite, la rotation fréquente des identités temporaires et des clés cryptographiques réduit la fenêtre d'exposition en cas de compromission et empêche la réutilisation abusive d'identifiants compromis.

La résilience du réseau repose également sur un renforcement de l'infrastructure : seuls les nœuds explicitement autorisés peuvent participer au réseau, et les protocoles de consensus retenus (Raft, TFBP) offrent une meilleure résistance aux attaques connues. À un niveau plus global, l'anonymisation avancée des données agrégées permet de limiter les risques liés à l'exploitation de ces informations. Des techniques comme la confidentialité différentielle assurent que les statistiques publiées (par exemple les heatmaps) ne révèlent aucune donnée sensible permettant de profiler les utilisateurs.

Enfin, la sécurité applicative est renforcée par des audits réguliers et des tests de pénétration. Ces évaluations visent à identifier en continu d'éventuelles vulnérabilités dans l'application de visualisation ainsi que dans l'infrastructure, afin d'y apporter des correctifs proactifs.

Il convient toutefois de souligner certaines limites et hypothèses. Certaines attaques restent en dehors du périmètre traité par le prototype, notamment la collusion entre nœuds, les attaques physiques sur les appareils des utilisateurs ou encore les attaques globales sur le réseau Tor (par corrélation de trafic). Par ailleurs, le système repose sur deux hypothèses implicites : d'une part, que les participants respectent les protocoles définis et, d'autre part, que les nœuds validateurs soient correctement configurés et protégés contre les intrusions. Ces limites ouvrent la voie à des travaux futurs visant à renforcer encore la robustesse du système face à des scénarios adverses plus sophistiqués.

5. *Évaluation et révision continue.* Ce modèle vise à couvrir les menaces les plus pertinentes dans le contexte du prototype tout en reconnaissant les limites imposées par les ressources et le cadre défini pour ce mémoire. Les attaques non traitées représentent des opportunités pour des travaux futurs.

Ce modèle de menaces n'est également pas statique, il doit être régulièrement mis à jour pour refléter l'évolution du paysage des menaces, l'introduction de nouvelles fonctionnalités et les retours des utilisateurs. Une veille sécuritaire proactive et une collaboration avec la communauté de la cybersécurité sont essentielles pour anticiper et contrer les nouvelles menaces.

CHAPITRE 6
TRAVAUX FUTURS ET CONCLUSION

6.1 Test/déploiement

Le prototype développé dans le cadre de cette thèse représente une avancée significative vers l'élaboration d'un service de géolocalisation qui protège la confidentialité des utilisateurs. La prochaine étape cruciale de ce projet consiste à implémenter les applications mobiles et les systèmes backend dans un environnement réel. Cette démarche permettra de recueillir et d'analyser des données utilisateur authentiques, facilitant une évaluation concrète de la performance du système, de l'engagement des utilisateurs et des aspects nécessitant des améliorations. De plus afin de faciliter le développement du prototype les canaux d'ingestion de données utilise le mode de consensus solo sur Hyerpledger, dont la complexité est basse et l'utilité principale est dans la phase de développement, tandis que le canal global, contenant toutes les organisations utilise raft ; le déploiement en environnement réel serait une occasion de déterminer, appliquer et tester les mécanismes de consensus adéquat sur le terrain.

6.1.0.1 Objectifs principaux

Les objectifs principaux de cette expérimentation sont triples. D'une part, il s'agit de valider en pratique le fonctionnement du prototype en le testant dans une variété d'environnements réels, afin de vérifier sa fiabilité et sa facilité d'usage. D'autre part, l'analyse des données collectées dans ces conditions réelles permettra de les comparer avec celles issues des simulations, afin de mesurer l'exactitude du système et d'identifier d'éventuels écarts. Enfin, l'expérimentation vise à optimiser le système à partir des retours utilisateurs et des performances observées, avec pour priorité l'amélioration de l'expérience utilisateur, de l'efficacité globale et du niveau de protection de la vie privée.

6.2 Amélioration des données de simulation/injection de bruit

Bien que le modèle de simulation actuel ait offert une base solide pour la compréhension des schémas de mouvement des utilisateurs, une approche plus sophistiquée pour la simulation de mouvement est requise pour un réalisme accru. Un problème rencontré avec les données de simulation est qu'elles sautent parfois des intersections. Cela est dû à la méthode de génération des données, qui repose sur une *hashmap* attribuant à chaque couple de latitude/longitude une géoadresse. Les coordonnées de latitude et de longitude

sont ensuite utilisées pour évaluer la distance avec les points avoisinants afin de trouver le plus proche. Ce processus heuristique, bien qu'efficace pour simplifier le calcul, peut entraîner des incohérences, telles que des sauts irréguliers entre intersections. Par conséquent, les mouvements simulés manquent occasionnellement de réalisme, car ils ne respectent pas les itinéraires logiques que suivraient des utilisateurs réels. La figure 6.1 illustre cette problématique en montrant comment les déplacements simulés sautent des intersections, créant des patterns incohérents qui ne correspondent pas à des comportements réalistes.

La solution envisagée pour résoudre cette limitation est l'intégration d'OpenRouteService (27), un service de routage avancé basé sur les données d'OpenStreetMap. Ce service permettrait de générer des itinéraires réalistes entre intersections, réduisant les déplacements irréalistes et améliorant ainsi la qualité des données simulées. En adoptant une structure ressemblant à un graphe, où les déplacements se font de manière logique d'une intersection à une autre sans saut. Le second avantage de cette amélioration est qu'elle constituerait une avancée majeur pour l'implémentation de la fonctionnalité d'injection de bruit, car cette dernière fonctionnerait sur la même logique lorsqu'un utilisateur déciderait de se débarrasser de son identité actuelle, tentant de produire un itinéraire réaliste en prenant en compte les routes autours de l'utilisateur, le nombre de transactions de mouvements restantes sur leur identité, etc. sans générer des anomalies trop facilement détectable dans les données.

Figure 6.1 – Problématique d’espacement des données de simulation.

6.3 Modèle économique et incitations

Un défi fondamental pour les technologies décentralisées et respectueuses de la vie privée, comme LoChain, réside dans l'établissement d'un modèle économique durable favorisant une adoption généralisée et un soutien continu. Les systèmes traditionnels axés sur la confidentialité, comme le réseau Tor, dépendent fortement de bénévoles, ce qui limite souvent leur croissance, leurs performances et leur viabilité. Pour surmonter ces limites, LoChain propose un modèle économique pragmatique, spécifiquement aligné sur ses forces architecturales, les normes de confidentialité actuelles et les besoins commerciaux réalistes.

Actuellement, les services de géolocalisation suivent généralement un modèle centralisé, où les données de localisation des utilisateurs sont continuellement collectées et stockées directement par les entreprises. Cette approche entraîne d'importants risques pour la confidentialité ainsi que des responsabilités substantielles en matière de gestion des données et de conformité aux réglementations telles que le RGPD ou la Loi 25 (Québec). Les entreprises doivent investir massivement dans le stockage sécurisé des données, les mesures de protection et les cadres de conformité, ce qui entraîne des coûts opérationnels considérables et des risques de responsabilité juridique.

Dans notre contexte, le terme « transaction » désigne l'opération par laquelle une donnée de localisation, déjà anonymisée via une géoadresse et associée à une identité temporaire, est enregistrée et validée sur la chaîne de blocs. Chaque enregistrement peut ainsi être considéré comme une transaction de localisation, soumise à validation par le réseau. Le modèle incitatif envisagé repose sur le fait que ces transactions, lorsqu'elles sont validées, acquièrent une valeur économique pour certains acteurs. Les entreprises qui reposent sur la fiabilité des données de mobilité — par exemple les services de covoiturage ou de VTC comme Uber, les gestionnaires de flotte, ou encore les assureurs — bénéficient directement de l'accès à des données cohérentes et vérifiées. Dans ce cadre, une partie des frais liés à la validation des transactions peut être assumée par ces acteurs économiques, qui y trouvent un intérêt concret : disposer de données géolocalisées plus fiables, sécurisées et résistantes à la falsification.

Plutôt que d'opter pour un stockage centralisé traditionnel et une gestion directe des données de localisation des utilisateurs, LoChain introduit un modèle innovant exploitant les identités numériques temporaires et la vérification de localisation basée sur la chaîne de blocs. Ce modèle permet aux entreprises de vérifier en toute sécurité la position en temps réel d'un utilisateur sans avoir besoin de stocker, gérer ou contrôler directement des données de localisation sensibles.

6.3.1 Exemple illustratif : application de VTC

Pour illustrer, considérons un scénario impliquant une application de VTC comme Uber fonctionnant sur le réseau LoChain :

1. **Génération d'un pointeur par l'utilisateur.** Un utilisateur souhaitant réserver un trajet génère en toute sécurité un *pointeur* chiffré faisant référence à sa dernière transaction de géolocalisation stockée sur la chaîne de blocs. Ce pointeur est signé cryptographiquement à l'aide de la clé privée de son identité temporaire, garantissant son authenticité et sa sécurité.
2. **Demande de vérification par l'entreprise.** L'entreprise reçoit ce pointeur chiffré et le soumet au réseau chaîne de blocs via une requête de vérification signée avec sa propre clé privée.
3. **Vérification par la chaîne de blocs.** : Le réseau chaîne de blocs authentifie la requête de l'entreprise en utilisant les clés publiques et vérifie la transaction de géolocalisation référencée. Le réseau confirme ensuite en toute sécurité la position actuelle de l'utilisateur directement à l'entreprise en temps réel, sans révéler les données brutes au-delà de la vérification demandée.

Cette approche offre plusieurs avantages convaincants pour toutes les parties prenantes :

— **Confidentialité renforcée des utilisateurs** : Les données de localisation des utilisateurs restent chiffrées et privées, avec une visibilité publique retardée d'au moins 24 heures. Cela empêche la surveillance en temps réel et réduit considérablement les risques liés à l'utilisation abusive des données ou au profilage des utilisateurs.
— **Réduction de la responsabilité des entreprises concernant les données** : Les entreprises ne sont plus responsables du stockage ou de la gestion de grandes quantités de données de géolocalisation. Cela simplifie grandement la conformité aux lois sur la confidentialité comme le RGPD et réduit les coûts opérationnels.
— **Données agrégées disponibles publiquement** : Après un délai de 24 heures, les données de géolocalisation agrégées et anonymisées deviennent accessibles au public. Ces données servent des intérêts plus larges, soutenant la planification urbaine, l'optimisation des transports, le développement des infrastructures et la recherche académique.
— **Modèle de revenus durable** : Chaque demande de vérification de localisation effectuée par une entreprise entraîne des frais de transaction minimes payables au réseau chaîne de blocs. Ces frais

créent une incitation économique durable pour les parties prenantes exploitant les nœuds et les validateurs, garantissant ainsi une viabilité financière à long terme sans dépendre uniquement d'un soutien bénévole ou d'entités centralisées.

6.3.2 Équilibre des intérêts et garantie de la conformité

L'adoption réussie de ce modèle économique repose sur un équilibre crucial entre les intérêts en matière de confidentialité des utilisateurs, les besoins opérationnels des entreprises et les bénéfices publics des données de localisation agrégées. Des mesures doivent être mises en place pour prévenir les abus potentiels, tels que la surutilisation des requêtes de vérification de localisation par les entreprises ou leur exploitation indirecte pour suivre les utilisateurs. La conformité aux cadres réglementaires, y compris une gestion claire et explicite du consentement des utilisateurs, reste primordiale pour maintenir la confiance et garantir la légitimité opérationnelle à long terme. En combinant une forte préservation de la confidentialité, une réduction des responsabilités des entreprises et une structure d'incitation économique durable, le modèle économique de LoChain constitue une approche pragmatique et attrayante pour le déploiement à grande échelle des services de géolocalisation décentralisés.

6.4 Gouvernance décentralisée par DAO

Le marché actuel des données de géolocalisation est marqué par un manque de transparence et une absence significative de sensibilisation du public. Les utilisateurs ignorent souvent comment leurs données de localisation personnelles sont collectées, vendues et utilisées. Les réglementations existantes en matière de confidentialité (telles que le RGPD) sont fréquemment contournées ou insuffisamment appliquées, conduisant à des abus systémiques et à une méfiance généralisée.

De plus, le marché des données de géolocalisation est dominé par des entités centralisées dont les incitations économiques entrent souvent en conflit direct avec la confidentialité des utilisateurs. Ce contrôle centralisé des données sensibles entraîne des risques considérables en matière d'abus, de violations de sécurité et de gestion irresponsable des données.

L'architecture de LoChain traite naturellement de nombreuses préoccupations en matière de confidentialité à la source en anonymisant et en décentralisant la gestion des données de géolocalisation. Cependant, au-delà de la simple décentralisation technique, la structure unique de LoChain offre également l'opportu-

nité d'établir une nouvelle norme industrielle transparente, remodelant fondamentalement les incitations économiques et la gouvernance.

Plutôt que de reposer sur une seule entité ou un groupe contrôlant le réseau et les standards de données associés, LoChain peut être exploité comme un réseau basé sur un consortium structuré sous la forme d'une Organisation Autonome Décentralisée (DAO). Une telle structure permet une gouvernance démocratique et coopérative impliquant plusieurs parties prenantes, notamment les entreprises, les autorités locales, les ONG, les institutions de recherche et même les utilisateurs eux-mêmes.

6.4.1 Structure DAO proposée pour la gouvernance des données de géolocalisation

Dans le cadre du modèle DAO, le réseau LoChain fonctionnerait comme suit :

- **Contrôle et propriété décentralisés.** Plusieurs parties prenantes deviennent membres de la DAO, chacune contrôlant et exploitant des nœuds au sein du réseau LoChain. Le pouvoir décisionnel, les incitations économiques et la gouvernance du réseau sont distribués, évitant ainsi tout point de contrôle unique.
- **Incitations économiques transparentes.** Les entreprises accédant aux services de vérification de géolocalisation paient des frais de transaction (comme décrit précédemment). Ces frais financent l'exploitation, l'entretien et l'expansion du réseau contrôlé par la DAO. Les membres de la DAO décident démocratiquement des structures tarifaires, des priorités d'investissement et des améliorations du réseau.
- **Prise de décision équitable et démocratique.** Les décisions clés en matière de gouvernance, y compris les standards de gestion des données, les protocoles de confidentialité et les structures d'incitation, sont prises collectivement par les membres de la DAO via un mécanisme de vote transparent. Les parties prenantes participent selon des rôles et responsabilités clairement définis, garantissant responsabilité et équité.
- **Participation et protection des utilisateurs.** Les utilisateurs peuvent devenir des participants directs ou des parties prenantes de la DAO, garantissant que leurs intérêts (en particulier en matière de confidentialité et de transparence) sont directement représentés dans les décisions de gouvernance. Les droits et protections des utilisateurs sont codifiés dans les documents de gouvernance de la DAO, renforçant ainsi la confiance et la transparence.

6.4.2 Avantages de la gouvernance économique basée sur une DAO

La mise en place d'une DAO pour encadrer la gestion des données de mobilité présente plusieurs bénéfices stratégiques. Tout d'abord, elle permet la standardisation à l'échelle de l'industrie en définissant collectivement des standards transparents et respectueux de la vie privée, la DAO offre une référence universelle pour les entreprises. Cette standardisation réduit l'ambiguïté réglementaire, facilite la conformité et limite la responsabilité associée à la gestion des données personnelles.

Elle constitue également un levier de confiance et de transparence. Grâce à une gouvernance démocratique et ouverte, les décisions sont prises de manière claire et responsable, permettant aux utilisateurs comme aux entreprises de comprendre comment et pourquoi les données sont accessibles, partagées ou monétisées. Cette transparence renforce la confiance du public et favorise l'adoption du système.

Sur le plan économique, la DAO assure la durabilité à long terme en mettant en place un modèle d'incitations clair. Les membres reçoivent des récompenses proportionnelles à leur participation, que ce soit via l'exploitation de nœuds, l'apport de ressources ou l'amélioration de la sécurité et de l'utilité du réseau. Ce mécanisme aligne les intérêts individuels et collectifs autour du maintien d'un réseau robuste et sécurisé.

Enfin, le modèle contribue à réduire les risques juridiques et de conformité. En anonymisant les données dès leur création et en établissant un standard collectif pour leur gestion, la DAO atténue les risques liés aux violations des réglementations telles que le RGPD ou la Loi 25. Les entreprises en retirent un double avantage : une diminution des coûts opérationnels liés à la conformité et une meilleure protection contre les sanctions potentielles.

6.4.3 Défis potentiels et considérations

Bien que prometteuse, l'approche DAO doit aborder certains défis pour rester équitable et viable. Tout d'abord, déterminer une distribution initiale équitable du pouvoir de vote ou des jetons parmi les parties prenantes est essentiel. Les incitations économiques doivent être structurées de manière à éviter toute centralisation du contrôle ou toute influence disproportionnée des entités les plus puissantes. Ensuite, des règles transparentes doivent être définies pour la prise de décision, le vote, la résolution des conflits et la participation des parties prenantes, garantissant ainsi une gouvernance efficace et efficiente. Enfin, les DAO et les modèles de gouvernance basés sur des jetons peuvent soulever des complexités réglementaires. Une

structuration juridique soigneuse et des lignes directrices claires en matière de conformité seront essentielles pour favoriser une adoption généralisée.

L'intégration de la gouvernance DAO dans le modèle économique de LoChain représente une opportunité révolutionnaire de redéfinir la gestion, la monétisation et la réglementation des données de géolocalisation. En démocratisant le contrôle, en établissant des incitations économiques transparentes et en créant des standards industriels coopératifs, LoChain peut impulser une économie des données plus fiable, durable et équitable.

6.5 Conclusion

La réalisation de ce travail de recherche sur le service de géolocalisation décentralisée s'inscrit dans une démarche d'innovation et d'interrogation sur les paradigmes actuels de gestion des données de localisation. En particulier, ce mémoire contribue à une réflexion critique sur les possibilités et les défis que représentent les technologies décentralisées pour la protection de la vie privée dans l'espace numérique. Le système proposé, basé sur une architecture Hyperledger et articulé autour des concepts d'identités temporaires, d'un modèle économique équilibré et d'un cadre de sécurité robuste, offre une voie potentielle vers la réconciliation de la nécessité d'une géolocalisation précise avec l'impératif de la préservation de la vie privée des utilisateurs. Ce projet tente d'inviter à repenser les modèles de gouvernance des données et les interactions entre les individus et le tissu urbain à l'ère numérique.

Cependant, la concrétisation de ce projet nécessitera un développement continu afin de mettre en place les fonctionnalités restantes, ainsi qu'une vigilance constante face aux évolutions technologiques, aux ajustements législatifs et aux exigences sociétales. Il est important de noter que nous ne prétendons pas résoudre complètement le problème de l'anonymat de localisation avec cette solution, car entre le point d'accès sans fil, le réseau cellulaire et les potentielles applications de détection de foule installées par les utilisateurs, cela serait un objectif déraisonnable.

L'implémentation et l'expérimentation dans des contextes réels fourniront des données précieuses pour évaluer la robustesse, l'efficacité et l'acceptabilité sociale du système proposé. Ces étapes futures sont essentielles pour affiner le modèle, assurer son intégration harmonieuse dans le quotidien des utilisateurs et déterminer sa capacité à répondre de manière équilibrée aux enjeux de sécurité et de confidentialité. Ce mémoire ouvre aussi des pistes pour des recherches futures, notamment sur les aspects économiques des

systèmes décentralisés, les implications sociales de la gestion autonome des données de localisation, et les cadres réglementaires adaptés à ces nouvelles technologies.

]

[glo] About GLONASS — glonass-iac.ru. https://glonass-iac.ru/en/about_glonass/. [Accessed 17-05-2024].

[gps] GPS.gov : GPS Overview — gps.gov. https://www.gps.gov/systems/gps/. [Accessed 17-05-2024].

[bei] System — en.beidou.gov.cn. http://en.beidou.gov.cn/SYSTEMS/System/. [Accessed 17-05-2024].

[vic] The World's Oldest Blockchain Has Been Hiding in the New York Times Since 1995 — vice.com. https://www.vice.com/en/article/j5nzx4/what-was-the-first-blockchain. [Accessed 17-05-2024].

[esa] What is Galileo ? — esa.int. https://www.esa.int/Applications/Satellite_navigation/Galileo/What_is_Galileo. [Accessed 17-05-2024].

[6] (2008). *Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Standard*. Rapport technique, U.S. Department of Defense. https://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf.

[7] Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., Caro, A. D., Enyeart, D., Ferris, C., Laventman, G., Manevich, Y. *et al.* (2018a). *Hyperledger fabric : A distributed operating system for permissioned blockchains*.

[8] Androulaki, E., Barger, A., Bortnikov, V. et et al. (2018b). Hyperledger fabric : A distributed operating system for permissioned blockchains. Dans *Proceedings of the Thirteenth EuroSys Conference (EuroSys '18)*. ACM. http://dx.doi.org/10.1145/3190508.3190538

[9] Antonopoulos, A. M. (2017). Mastering bitcoin : Unlocking digital cryptocurrencies. 2nd Edition, Published by O'Reilly Media. Récupéré de https://github.com/bitcoinbook/bitcoinbook

[10] Bano, S., Sonnino, A., Al-Bassam, M., Azouvi, S., McCorry, P., Meiklejohn, S. et Danezis, G. (2019). Sok : Consensus in the age of blockchains. Dans *Proceedings of the 1st ACM Conference on Advances in Financial Technologies*, 183–198.

[11] Bayer, D., Haber, S. et Stornetta, W. S. (1992). Improving the efficiency and reliability of digital time-stamping. Dans *Sequences*, 329–334. Springer.

[12] BBC News (2019). Apple facetime bug allows iphone users to eavesdrop before call is answered. Récupéré le 2025-01-06 de https://www.bbc.com/news/technology-47063973

[13] Brown, R. G. (2016). Corda : An introduction. *R3 Reports*.

[14] Bush, A. (2019). Cartography in the age of digital mapping. *Journal of Digital Geography*, *27*(2), 98–114.

[15] Buterin, V. (2014a). Ethereum : A next-generation smart contract and decentralized application platform. *Ethereum White Paper*, *3*(37), 2–1. Récupéré de https://ethereum.org/en/whitepaper/

[16] Buterin, V. (2014b). Proof of stake : How i learned to love weak subjectivity. *Ethereum Blog*.

[17] Buterin, V. (2016). Sharding faq. https://github.com/ethereum/wiki/wiki/Sharding-FAQ. Accessed : 2024-05-17.

[18] Castro, M. et Liskov, B. (1999). Practical byzantine fault tolerance. Dans *Proceedings of the Third Symposium on Operating Systems Design and Implementation (OSDI)*, 173–186. USENIX Association.

[19] de l'image ou site web (si disponible), A. (2025). Structure d'un arbre de merkle. https://external-content.duckduckgo.com/iu/?u=https%3A%2F%2Form-chimera-prod.s3.amazonaws.com%2F1234000001802%2Fimages%2Fmsbt_0705.png&f=1&nofb=1&ipt=9708b48af46df931fb963578188b74e3b14e72635176ba91cfe9499c84afe1d3&ipo=images. Consulté le 6 janvier 2025.

[20] De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. et Blondel, V. D. (2013). Unique in the crowd : The privacy bounds of human mobility. *Scientific Reports*, *3*, 1376. http://dx.doi.org/10.1038/srep01376

[21] Dingledine, R., Mathewson, N. et Syverson, P. (2004). Tor : The second-generation onion router. Dans *Proceedings of the 13th USENIX Security Symposium*, 303–320., San Diego, CA, USA. USENIX Association. Récupéré de https://www.torproject.org/

[22] Dolev, D. et Yao, A. C. (1983). On the security of public key protocols. Dans *Proceedings of the 22nd Annual Symposium on Foundations of Computer Science (SFCS)*, 350–357. IEEE.

[23] Dolev, S. et Welch, J. L. (2004). Self-stabilizing clock synchronization in the presence of byzantine faults. *J. ACM*, *51*(5), 780–799. http://dx.doi.org/10.1145/1017460.1017463. Récupéré de https://doi.org/10.1145/1017460.1017463

[24] Domo, Inc. (2017). Data never sleeps 5.0. Récupéré le 2025-01-06 de https://www.domo.com/learn/infographic/data-never-sleeps-5

[25] Ester, M., Kriegel, H.-P., Sander, J. et Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. 226–231.

[26] FOAM (2018). Foam whitepaper. Récupéré le 2018-05-02 de https://foam.space/publicAssets/FOAM_Whitepaper.pdf

[27] for Geoinformation Technology (HeiGIT), H. I. (2025). Openrouteservice : A web-based route planning application. Accessed January 2, 2025. Récupéré de https://openrouteservice.org

[28] Ganti, R. K., Ye, F. et Lei, H. (2011). Mobile crowdsensing : current state and future challenges. *IEEE Communications Magazine*, *49*(11), 32–39.

[29] Gavish, B. et Sheng, O. R. (2006). Ip address location finding : An empirical study. *Computers & Security*, *25*(4), 297–307.

[30] Gupta, M., Kim, M. et Saxena, N. (2018). Location-proof systems : Advances and challenges. *IEEE Communications Surveys & Tutorials*, *20*(3), 1988–2008. http://dx.doi.org/10.1109/COMST.2018.2827283. Récupéré de https://doi.org/10.1109/COMST.2018.2827283

[31] Haber, S. et Stornetta, W. S. (1991). How to time-stamp a digital document. *Journal of Cryptology*, *3*(2), 99–111.

[32] Haklay, M. et Weber, P. (2008). Openstreetmap : User-generated street maps. *IEEE Pervasive Computing*, *7*(4), 12–18. http://dx.doi.org/10.1109/MPRV.2008.80

[33] Humphreys, T. E., Bhatti, J. A. et Ledvina, B. M. (2010). The gps assimilator : a method for upgrading existing gps user equipment to improve accuracy, robustness, and resistance to spoofing. Dans *Proceedings of the 23rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2010)*, 1942–1952.

[34] Hyperledger Foundation (2025). Hyperledger fabric. Récupéré le 2025-01-06 de https://www.hyperledger.org/projects/fabric

[35] Hyperledger Performance and Scale Working Group (2020). *Hyperledger Fabric Performance and Scale Study*. Rapport technique, Linux Foundation

[36] Inc., P. L. (2021). Planetscope : Global daily satellite imagery. Accessed : December 26, 2024. Récupéré de https://www.planet.com/products/planet-imagery/

[37] ISO/IEC (2019). Iso/iec 27701 :2019 - security techniques — extension to iso/iec 27001 and iso/iec 27002 for privacy information management — requirements and guidelines. *International Organization for Standardization*. Récupéré de https://www.iso.org/standard/71670.html

[38] Johnson, D., Menezes, A. et Vanstone, S. (2001). *The Elliptic Curve Digital Signature Algorithm (ECDSA)*. Berlin, Heidelberg : Springer. http://dx.doi.org/10.1007/3-540-45664-3_11. Récupéré de https://link.springer.com/chapter/10.1007/3-540-45664-3_11

[39] Lamport, L., Shostak, R. et Pease, M. (1982). The byzantine generals problem. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, *4*(3), 382–401.

[40] Lane, N. D., Eisenman, S. B., Musolesi, M., Miluzzo, E. et Campbell, A. T. (2008). Urban sensing systems : opportunistic or participatory ? Dans *Proceedings of the 9th workshop on Mobile computing systems and applications*, 11–16. ACM.

[41] Leppänen, V. (2010). Trilateration algorithms for mobile device localization based on signal strength measurements. *Proceedings of the 11th International Conference on Telecommunications*, 101–105.

[42] Lohr, S. (2018). Google+ data breach exposed data of up to 500,000 users. https://www.nytimes.com/2018/10/08/technology/google-plus-security-disclosure.html. [Accessed 17-05-2024].

[43] Manjoo (2012). How Companies Learn Your Secrets (Published 2012) — nytimes.com. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html. [Accessed 17-04-2024].

[44] McKinnell, J. (2021). Google faces massive fines after world-first data breach ruling. Récupéré de https://thenewdaily.com.au/life/tech/2021/04/16/google-location-sharing/

[45] Merkle, R. C. (1988). A digital signature based on a conventional encryption function. *Advances in Cryptology—CRYPTO'87*, 369–378.

[46] Mettler, M. (2016). Blockchain technology in healthcare : The revolution starts here. *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 1–3.

[47] Microsoft (2005). *The STRIDE Threat Model*. Rapport technique, Microsoft Corporation. Disponible en ligne : https : //learn.microsoft.com/en-us/previous-versions/commerce-server/ee823878(v=cs.20).

[48] Miraz, M. H. et Ali, M. (2018). Applications of blockchain technology beyond cryptocurrency. *arXiv preprint arXiv :1801.03528*.

[49] Nakamoto, S. (2008). Bitcoin : A peer-to-peer electronic cash system. *Decentralized Business Review*. Récupéré de https://bitcoin.org/bitcoin.pdf

[50] Niemeyer, G. (2008). Geohash : Hierarchical spatial data structure for geolocation encoding. Accessed : December 26, 2024. Récupéré de http://geohash.org

[51] Ohm, P. (2010). Broken promises of privacy : Responding to the surprising failure of anonymization. *UCLA Law Review*, *57*, 1701–1777. Récupéré de https://www.uclalawreview.org/pdf/57-6-3.pdf

[52] Ongaro, D. et Ousterhout, J. (2014). In search of an understandable consensus algorithm. Dans *2014 USENIX Annual Technical Conference (USENIX ATC 14)*, 305–319. USENIX Association. Récupéré de https://www.usenix.org/conference/atc14/technical-sessions/presentation/ongaro

[53] O'Flaherty, K. (2020). Apple iphone location controversy : What you need to know. https://www.forbes.com/sites/kateoflahertyuk/2020/01/24/apple-iphone-location-controversy-what-you-need-to-know/. [Accessed 17-05-2024].

[54] Pyrgelis, A., Troncoso, C. et De Cristofaro, E. (2017). What does the crowd say about you ? evaluating aggregation-based location privacy. *arXiv preprint arXiv :1703.00366*.

[55] Pyrgelis, A., Troncoso, C. et Huguenin, K. (2018). Knock knock, who's there ? membership inference on aggregate location data. Dans *Proceedings of the Network and Distributed System Security Symposium (NDSS)*. Internet Society. http://dx.doi.org/10.14722/ndss.2018.23155

[56] Quercia, D., Leontiadis, I., McNamara, L., Mascolo, C. et Crowcroft, J. (2011). Spotme if you can : Randomized responses for location obfuscation on mobile phones. Dans *2011 31st International Conference on Distributed Computing Systems*, 363–372. http://dx.doi.org/10.1109/ICDCS.2011.79

[57] Sakurai, K. et Watanabe, Y. (2010). Performance evaluation of elliptic curve cryptography on mobile devices. Dans *Proceedings of the International Conference on Information Security*, 142–157. Springer. http://dx.doi.org/10.1007/978-3-642-16161-2_12

[58] Smith, J. et Doe, J. (2020). Google maps and apple maps : A comparative analysis. *International Journal of Mobile Computing*, *15*(3), 123–135.

[59] Smith, J. et Doe, J. (2022). Privacy concerns in community-based navigation applications : A case study of waze. *Journal of Privacy and Data Protection*, *10*(2), 150–162.

[60] Statista Research Department (2021). Usage of google maps and apple maps. Récupéré de https://www.statista.com/statistics/865413/most-popular-smartphone-map-apps-usa/

[61] Team, M. (2024a). Maps.me : Offline maps and navigation. https://maps.me. [Accessed 17-05-2024].

[62] Team, W. (2024b). Waze : Gps navigation, maps, and traffic alerts. https://www.waze.com. [Accessed 17-05-2024].

[63] The Tor Project (2025a). Orbot : Tor for android. Récupéré le 2025-01-06 de https://www.torproject.org/orbot/

[64] The Tor Project (2025b). The tor project : Privacy & freedom online. Récupéré le 2025-01-06 de https://www.torproject.org

[65] Thompson, S. A. et Warzel, C. (2019). Twelve million phones, one dataset, zero privacy. The New York Times. Récupéré le 2025-01-06 de https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html

[66] Union, E. (2016). Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). *Official Journal of the European Union*. Récupéré de https://eur-lex.europa.eu/eli/reg/2016/679/oj

[67] Wikipedia contributors (2018). Facebook–cambridge analytica data scandal. Wikipedia, the free encyclopedia. Récupéré le 2025-01-06 de https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

[68] Wood, G. (2014). Ethereum : A secure decentralised generalised transaction ledger. https://ethereum.github.io/yellowpaper/paper.pdf. Accessed : 2024-05-17.

[69] Xiong, P., Zhu, T., Pan, L., Niu, W. et Li, G. (2014a). Privacy preserving in location data release : A differential privacy approach. Dans D.-N. Pham et S.-B. Park (dir.). *PRICAI 2014 : Trends in Artificial Intelligence*, 183–195., Cham. Springer International Publishing.

[70] Xiong, P., Zhu, T., Pan, L., Niu, W. et Li, G. (2014b). Privacy preserving in location data release : A differential privacy approach. In D.-N. Pham et S.-B. Park (dir.), *PRICAI 2014 : Trends in Artificial Intelligence* 183–195. Cham : Springer International Publishing

[71] Yi, S., Li, Q. et Shi, W. (2020). Privacy-preserving location proofs leveraging zero-knowledge proofs in decentralized systems. *IEEE Transactions on Mobile Computing*, *19*(8), 1783–1795. http://dx.doi.org/10.1109/TMC.2020.2965600. Récupéré de https://ieeexplore.ieee.org/document/8964298

[72] Zandbergen, P. A. (2009). Accuracy of iphone locations : A comparison of assisted gps, wifi and cellular positioning. *Transactions in GIS*, *13*(s1), 5–25.

[73] Zou, S., Xi, J., Wang, H. et Xu, G. (2019). Crowdblps : A blockchain-based location-privacy-preserving mobile crowdsensing system. *IEEE Transactions on Industrial Informatics*, *16*(6), 4206–4218.