

Titre: Gestion de la mobilité dans les réseaux denses de cinquième
Title: génération (5G)

Auteur: Éric Mayeul Olusegun Fafolahan
Author:

Date: 2019

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Fafolahan, É. M. O. (2019). Gestion de la mobilité dans les réseaux denses de
Citation: cinquième génération (5G) [Thèse de doctorat, Polytechnique Montréal].
PolyPublie. <https://publications.polymtl.ca/3944/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/3944/>
PolyPublie URL:

**Directeurs de
recherche:** Samuel Pierre
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Gestion de la mobilité dans les réseaux denses de
cinquième génération (5G)**

ÉRIC MAYEUL OLUSEGUN FAFOLAHAN

Département de génie informatique et génie logiciel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie informatique

Mai 2019

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

**Gestion de la mobilité dans les réseaux denses de
cinquième génération (5G)**

présentée par **Éric Mayeul Olusegun FAFOLAHAN**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Alejandro QUINTERO, président

Samuel PIERRE, membre et directeur de recherche

Ronald BEAUBRUN, membre

Ahmed KARMOUCH, membre externe

DÉDICACE

*Au
Comte de Saint-Germain . . .*

REMERCIEMENTS

Je voudrais, en premier lieu, remercier mon directeur de recherche, le Prof. Samuel Pierre pour son encadrement, sa disponibilité, pour le support académique et financier dont j'ai bénéficiés durant toutes mes années d'études. J'ai eu la chance de collaborer avec lui sur plusieurs projets qui m'ont enrichi sur les plans technique, professionnel et humain. Merci M. Pierre. Je voudrais aussi exprimer ma gratitude au Prof. Alejandro Quintero avec qui j'ai travaillé au département. Cette expérience m'a beaucoup enrichi sur le plan technique. Merci au Prof. Ronald Beaubrun pour ses conseils et suggestions qui ont été décisifs à un moment crucial durant mes études au doctorat.

Je voudrais également remercier M. Alejandro Quintero, M. Ronald Beaubrun, et M. Ahmed Karmouch pour avoir accepté d'évaluer mon travail.

J'adresse aussi mes remerciements à tous mes collègues du LARIM. Vos suggestions et vos critiques constructives durant les séminaires m'ont enrichi au fil des années. Collectivement, vous avez été une source de motivation et de dépassement de soi. Merci à Hazem, Almathe, Ronald, Marième, Valérie, Georges, Eunice, Aurel, Moussa, Saida, Lamia, Nasrin, Farnoush, Amir, Claudy, Olson...Individuellement, vous avez contribué, par vos qualités, à faire du Lab un cadre de travail convivial. Nous formions une belle diversité au sein de laquelle, chacun avait sa place. Merci particulièrement à Valérie, Hazem, Aurel et Eunice pour votre écoute, votre disponibilité et vos conseils. Merci Mathieu et Franjieh pour votre collaboration durant les projets auxquels nous avons participé.

Merci à mes amies Andrée Caron, Valérie Roy, Sabrina Boisvert, Alexandra Vincent, Marnie Vodounou, Jessica Ostiguy, et Eugénie Halgrain. Vous êtes exceptionnelles, et votre soutien a toujours été inconditionnel.

Je voudrais également exprimer ma gratitude et ma reconnaissance à mes parents, Édith et Emmanuel, à ma soeur Eunice et à mon beau-frère Cédric. Malgré la distance et tous les sacrifices que les circonstances pouvaient exiger, vous m'avez toujours soutenu dans cette longue aventure, au meilleur de vos capacités. Je vous en suis infiniment reconnaissant. Merci papa pour ton soutien financier et tes encouragements à aller le plus loin possible dans mes études. Je te remercie pour ton assistance spirituelle, Maman. Tu m'as toujours encouragé à

rester fidèle à mes principes.

Enfin, je ne saurai jamais assez remercier, Divine Providence, mon alliée de chaque instant en qui, j'ai la Vie, le Mouvement et l'Être...

RÉSUMÉ

Les réseaux de communications mobiles ont connu de profondes avancées technologiques au cours des deux dernières décennies. La croissance du nombre d'abonnés mobiles ainsi que l'accès à des forfaits de données illimitées, souvent à des tarifs préférentiels, ont engendré une demande de bande passante, de vidéo et de données en forte croissance. Ces progrès significatifs ont favorisé le déploiement de nouveaux services et de nouveaux cas d'utilisation tels que l'Internet-des-objets (IoT), la réalité augmentée et virtuelle, les réseaux de villes intelligentes, les véhicules autonomes et l'automatisation industrielle. Aux technologies existantes, s'ajouteront de nouveaux modes de communication dans le but de répondre à plusieurs cas d'utilisation des réseaux mobiles qui sont encore difficiles à satisfaire à ce jour. Le résultat à long terme de cette nouvelle tournure dans le monde de la réseautique mobile est désigné sous le vocable de réseaux de cinquième génération (5G). Au-delà du déploiement d'applications avancées, les réseaux 5G offriront de nouvelles opportunités de revenus aux fournisseurs de services lorsqu'ils seront combinés aux fonctionnalités avancées telles que l'analyse de données, l'apprentissage automatique et à l'intelligence artificielle.

Dans ce contexte, un large consensus est aujourd'hui établi sur la nécessité d'accroître la capacité du réseau par un déploiement massif de cellules de petite taille (Small Cell, SCs), d'un rayon de couverture réduit et à faible puissance. On parle alors d'une ultra-densification du réseau dont le but essentiel est de favoriser la proximité des points d'accès des utilisateurs finaux. Cependant, la densification du réseau implique des relèves fréquentes des usagers mobiles (MNs) entre les SCs et les zones de service. En effet, le rayon de couverture réduit des SCs rend plus complexe la phase de sélection des relèves en plus d'accroître la fréquence de celles-ci. Ces relèves entraînent des dégradations, des perturbations et des déconnexions qui peuvent entraver l'objectif d'un accès transparent aux services du réseau. En outre, la fréquence des relèves engendre une latence et une charge de signalisation élevées dans le réseau. De plus, l'omniprésence d'applications temps réel exige une latence faible du réseau. Dans ce contexte, la gestion de la mobilité demeure encore un enjeu et il s'avère donc indispensable de concevoir de nouveaux protocoles de gestion de la mobilité capables répondre aux exigences de performances strictes des réseaux 5G.

Aujourd'hui, la gestion de la mobilité s'articule principalement autour de deux approches. L'approche de gestion centralisée de la mobilité (CMM) et l'approche de gestion distribuée de la mobilité (DMM). L'approche CMM repose sur la présence d'un point d'ancrage central qui est responsable de la signalisation de mobilité et du transfert des données du MN. L'avantage

de l'approche CMM réside dans sa simplicité. Cependant, avec l'explosion du nombre de MNs connectés, cette approche a révélé ses limites. La centralisation de l'architecture du réseau entraîne une augmentation significative de la charge de signalisation sur le point d'ancrage du trafic, une faible évolutivité, un routage sous-optimal, l'existence d'un point de rupture dont la défaillance entraînerait une dégradation significative de performance globale dans le réseau, des failles de sécurité, ainsi qu'une utilisation non-optimale des ressources du réseau. Pour faire face à ces limitations, l'approche DMM est l'une des alternatives considérées au sein de l'Internet Engineering Task Force (IETF). Elle répartit le trafic du MN sur les points d'accès qui lui sont les plus proches, contribuant ainsi à la décentralisation des architectures des réseaux mobiles. Cependant, malgré le fait que plusieurs approches de la DMM visent l'atténuation des problèmes liés à la CMM, elles ne répondent pas toujours aux besoins de performance désirés. De récentes études de performance ont montré des limitations considérables de leur performance, surtout lorsque le temps de résidence des MNs au sein des SCs est court, ou lorsque le nombre de sessions actives amorçées au sein des réseaux visités est élevé. De plus, la plupart des approches distribuées ne tiennent compte que d'une seule relè. De ce fait, lorsque la durée des sessions du MN est s'étire dans le temps, elle peuvent s'étaler sur plusieurs relèves. Dans ce cas, l'efficacité de l'approche DMM dépend en grande partie de la durée des sessions par rapport au temps de résidence des MNs au sein des SCs. Dans un réseau densifié de SCs, lorsqu'un MN se déplace fréquemment et qu'il initie de nouvelles sessions au sein des différents réseaux visités, les performances des approches distribuées sont limitées en raison du nombre élevé de tunnels devant être configurés et gérés par les points d'ancrage distribués. Par conséquent, si la proportion de ces longs flux est trop importante, les approches distribuées risquent d'être très peu performantes au sein des réseaux denses 5G. Par ailleurs, les mécanismes de gestion de la mobilité actuels, manquent de support de mobilité adaptable à un profil de mobilité spécifique ou aux besoins spécifiques d'un MN. L'approche distribuée offre le même support de gestion de la mobilité à tous les MNs, quel que soit leur profil ou les exigences des l'application en cours d'exécution. Cependant, contrairement aux génération précédentes des réseaux mobiles, il est reconnu que les périphériques utilisateurs sont appelés à jouer un rôle plus important dans l'écosystème des réseaux 5G. En raison de la diversité des applications, des services et des cas d'utilisation émergents, offrir le même support de mobilité serait contre performant. Il est donc indispensable de mettre en œuvre dans la 5G, des mécanismes de gestion de la mobilité flexibles, adaptables aux exigences des applications des MNs, mais qui impliquent davantage les MNs.

Au regard de tout ce qui précède, cette thèse vise à concevoir une architecture de gestion de la mobilité dans les réseaux denses 5G. Afin d'atteindre cet objectif global, le travail a été réparti sur trois principales phases.

Lors de la première phase, un modèle de densification du réseau a été proposé pour le support de la mobilité. Le modèle proposé partitionne un domaine local fonctionnant sur le protocole LISP en plusieurs zones de localisation de service (LSAs), dans le but d'offrir un support de mobilité locale. Une LSA est une structure logique du réseau qui regroupe en son sein, un ensemble de SCs localement administrées par une entité du réseau dénommée point d'ancrage local (LA). La mise en oeuvre dudit modèle suppose une phase préalable de planification du réseau afin de déterminer la position des LAs au sein du domaine local. Par la suite, un modèle mathématique a été conçu afin de modéliser l'approche de densification proposée. Les deux étapes précédentes nous ont permis de proposer un protocole, dénommé LISP-LSA, basé sur le mécanisme d'ID/Locator Split afin d'offrir un support de la mobilité dans un réseau dense. Selon son mode de fonctionnement, le protocole LISP-LSA engendre deux types de relèves à savoir : la relève intra-LSA ainsi que la relève inter-LSA. Une relève intra-LSA implique deux SCs situées dans le même LSA, tandis qu'une relève inter-LSA implique deux SCs situées dans deux LSA différents. L'évaluation numérique des performances a révélé qu'en fonction de la taille du LSA et du taux de mobilité du MN, le protocole LISP-LSA est capable de réduire les coûts de signalisation des relèves intra-LSA et inter-LSA dans des proportions allant jusqu'à 38% et 21% respectivement, en comparaison à la version améliorée du protocole LISP-MN.

Tenant compte de l'explosion du trafic vidéo et de l'omniprésence des applications sensibles au délai, le protocole LISP-LSA a été étendu lors de la deuxième phase du projet afin d'intégrer un support de la QoS à son fonctionnement. Afin d'évaluer les performances du protocole étendu, plusieurs métriques ont été considérées dans le cadre de notre analyse. Il s'agit du coût de signalisation, du coût de livraison des données, de la latence, du taux de perte de paquets, et de la charge active de la passerelle du domaine (GW). Les deux types de relèves (intra-LSA et inter-LSA) ont été étendus avec de nouveaux modes de relèves proactives. Les relèves proactives exploitent l'information sur l'état des liens de la couche liaison de données (L2) afin d'anticiper la rupture du lien sans fil et débiter de manière anticipée l'exécution de la relève. Les relèves proactives ont l'avantage de réduire le coût de signalisation, la latence de la relève, la perte de paquets ainsi que le coût d'acheminement des paquets. En cas d'échec de la procédure de relève proactive, le MN doit passer en mode de relève réactive qui consiste à s'enregistrer à nouveau sur la SC de destination. Cette opération est nécessaire et est due à l'échec de transfert du contexte de mobilité du MN. Pour chaque métrique considérée, un modèle analytique d'évaluation de performance a été proposé pour chacun des protocoles LISP-MN et LISP-LSA. Lesdits modèles ont été implémentés pour une évaluation numérique. Mieux, les deux protocoles ont aussi été implémentés avec le simulateur de réseau ns-3.26. Les résultats numériques ont révélé que le protocole LISP-LSA réduit d'environ 30% le coût de signalisation du réseau. Mieux, lorsque le taux de mobilité est élevé dans un réseau densifié, le

coût de signalisation du réseau est réduit d'environ 26%. Par ailleurs, plus le MN est éloigné du coeur du réseau, le coût de signalisation est réduit dans une proportion allant jusqu'à 64%. Quant au coût de livraison des données, il a été réduit d'environ 66% tandis que la latence et le taux de perte de paquets ont été réduits de 54%. Les résultats analytiques ci-dessus mentionnés ont été validés par des résultats de simulation. Ces derniers ont révélé une réduction du coût de signalisation d'environ 36% tandis que le coût de livraison des données a été réduit de 62%. La latence et la perte de paquets quant à eux, ont été réduits d'environ 61.5% et 90% respectivement. Par ailleurs, il a été observé que le protocole LISP-LSA réduit également la charge active du GW dans une proportion allant jusqu'à 90%.

La troisième phase du travail a consisté en la proposition d'une architecture de gestion de la mobilité. La conception de ladite architecture a nécessité la définition de ses composantes et leur structure interne, ainsi que la définition des mécanismes de contrôle. Les mécanismes de contrôle définissent les procédures de fonctionnement par lesquelles l'architecture proposée est capable d'assurer la gestion de la mobilité des utilisateurs. Ils impliquent une interaction entre plusieurs composantes de l'architecture. Par la suite, nous avons proposé le modèle analytique du débit et de la probabilité d'interruption des sessions lors des relèves comme métriques complémentaires à l'évaluation de performance entamée à la phase précédente. Les modèles développés ont ensuite été implémentés avec le logiciel MATLAB pour une évaluation numérique. Les composantes de l'architecture ont également été implémentés avec le logiciel de simulation de réseaux ns-3.26 afin de valider les performances du protocole LISP-LSA. Les résultats numériques ont montré qu'en fonction de la distance entre le MN et le GW, le débit du réseau est amélioré d'environ 32%. Mieux, en fonction de la taille du LSA, le débit est amélioré d'environ 22%. Quant à la probabilité d'interruption des sessions, elle est réduite d'environ 40% lorsque le rayon de couverture des SCs tend à se réduire. Mieux, il a été observé que le protocole LISP-LSA réduit le taux d'interruption d'environ 55% lorsque la vitesse du MN est élevée tandis qu'il est éloigné du coeur du réseau. Les résultats analytiques mentionnés ont été par la suite validés par des résultats de simulation. Nous avons observé qu'en fonction de la distance qui sépare le MN du GW, le débit du réseau est accru dans une proportion allant jusqu'à 36%. Mieux, en fonction du taux d'arrivée des paquets le débit mesuré a été amélioré d'environ 28%. Lorsque le réseau se densifie, le débit du réseau est significativement amélioré dans une proportion allant jusqu'à 90%. Quant au taux d'interruption des communications, celui-ci est considérablement réduit dans des proportions allant jusqu'à 70%, en fonction de la vitesse du MN et de la distance avec le GW. Néanmoins, il a été observé que le taux d'interruption des sessions tend généralement à être élevé lorsque le délai sur le réseau de transit augmente.

ABSTRACT

Mobile communications networks have experienced tremendous technological advances in the last two decades. The growth of the number of mobile subscribers and access to unlimited data plans, often at very affordable prices, have led to an increased demand for bandwidth, video and high-growth data. These significant advances have facilitated the deployment of new services and use cases such as Internet-of-things (IoT), augmented and virtual reality, smart city networks, autonomous vehicles, and industrial automation. On top of the existing technologies, new communication modes will arise to respond to several uses cases of mobile systems that are still difficult to meet today. The long-term result of this new trend in the world of mobile networking gives birth to a new paradigm called the fifth generation networks (5G). Beyond deploying advanced applications, 5G networks will offer new revenue opportunities to service providers, when combined with advanced features such as data analytics, machine learning, and artificial intelligence.

In this context, a broad consensus is now established on the need to increase the network capacity through a massive deployment of small cells (Small Cell, SCs), with reduced coverage and low power. This requirement led to the ultra-densification of the network whose primary purpose is to promote the proximity of access points to the end-users. However, the densification of the cellular networks involves many mobile nodes (MNs) going through several handovers between the SCs and the service areas. The shorter SC's radius makes the handover selection phase more complex while increasing its frequency. These handovers lead to service disruptions and disconnections that may hinder the provision of seamless mobility of network services. Moreover, the frequency of the handovers generates a high latency and signaling load in the network. Besides, the ubiquity of real-time applications requires low network latency. In this context, mobility management is still an issue, and it is, therefore, essential to design new mobility management protocols that can meet the stringent performance requirements of 5G networks.

Today, there are two central mobility management approaches namely the Centralized Mobility Management (CMM) approach and the Distributed Mobility Management (DMM) approach. The CMM approach uses an anchoring node that is responsible for the mobility signaling and the transfer of MNs data. The advantage of the CMM approach lies in its simplicity. However, with the explosion of the number of connected MNs, this approach has revealed its limits. The centralization of the network architecture results in a significant increase of the signaling load on the anchor point, low scalability, suboptimal routing, the

existence of a single point of failure which would lead to a considerable network performance degradation, security vulnerabilities, and non-optimal use of network resources. To solve the limitations mentioned above, the DMM approach is one of the considered alternatives within the Internet Engineering Task Force (IETF). It distributes the MN traffic to the closest access points to it, contributing to the decentralization of mobile network architectures. However, even though several DMM approaches aim to mitigate the CMM issues, they do not always respond to the desired performance needs. Recent performance studies have shown considerable limitations in their performance, especially when the MN's cell residence time is shorter, or when the number of active sessions initiated within the visited networks is high. Besides, most distributed approaches take into account only one handover. Therefore, when the MN's sessions last during the time, they can spread over several many handovers. In this case, the effectiveness of the DMM approach largely depends on the relative duration of the sessions versus the MN's SC residence time. In a denser SCs network configuration, as the MN moves frequently and initiates new communications within the different visited networks, the performance of the distributed approaches is limited because of the large number of tunnels to be configured and managed by the distributed anchoring nodes. Therefore, if the proportion of these flows is too large, distributed approaches may perform poorly within dense 5G networks. Besides, current mobility management mechanisms lack adaptable mobility support to specific mobility profiles and MN's needs. The distributed approach provides the same mobility management support to all MNs, regardless of their profile or the requirements of the running application. However, unlike previous generations of mobile networks, it is recognized that user devices are expected to play a more significant role in the 5G network ecosystem. Due to the diversity of applications, services, and emerging use cases, offering the same mobility support would be counter-productive. It is therefore essential to implement flexible and adaptable mobility management mechanisms, with greater MNs involvement, in the 5G networks.

To this end, this thesis aims at designing a new mobility management architecture in dense 5G networks while enabling Quality of Service (QoS) support. In order to achieve this global goal, the work has been divided into three main phases.

During the first phase, a network densification model was proposed for mobility support. The proposed model partitions a local LISP domain into several location service areas (LSAs), in order to provide local mobility support. An LSA is a logical structure of the network that includes within it, a set of SCs locally administered by a network entity called the Local Anchor Point (LA). The implementation of this model assumes a preliminary planning phase of the network to determine the position of the LAs within the local domain. Afterward, a mathematical model was designed to model the proposed densification approach. The two

previous steps allowed us to propose a protocol, called LISP-LSA, based on the ID/Locator Split mechanism in order to enable mobility support in a dense network. The LISP-LSA protocol generates two types of handovers namely: the intra-LSA handover and the inter-LSA handover. The intra-LSA handover involves two SCs located in the same LSA, while the inter-LSA handover involves two SCs located in two different LSAs. The numerical performance evaluation revealed that, depending on the size of the LSA and the MN mobility rate, the LISP-LSA protocol can reduce the signaling costs of the intra-LSA and the inter-LSA handovers up to 38% and 21%, respectively, compared to the improved version of the LISP-MN protocol.

In regards to the massive increase of video traffic and the omnipresence of time-sensitive applications, we extended the LISP-LSA protocol during the second phase of the project in order to integrate the QoS support into its operation. We consider several metrics, namely the signaling cost, the data delivery cost, the handover latency, the packet loss rate, and the processing load on the domain gateway (GW), during the performance analysis. Indeed, both intra-LSA and inter-LSA handovers scenarios have been extended with the new proactive handover mechanisms. Proactive handovers exploit the data link layer (L2) information in order to anticipate the wireless link break and to timely start the handover execution. They have the advantage of reducing the signaling cost, the handover latency, the packet loss, and the data delivery cost. When the proactive handover fails, the MN switches into the reactive handover mode, which requires a new registration on the new SC. This operation is necessary and is due to the failure of the MN's mobility context transfer. Indeed, during our analysis, we considered the LISP-MN and the LISP-LSA protocols and proposed an analytical model for each handover metric that we considered for the performance evaluation. These models have been implemented for numerical evaluation. Besides, we implement both protocols with the ns-3.26 network simulator. The numerical results revealed that the LISP-LSA protocol reduces the signaling cost of the network up to 30%. Moreover, when the mobility rate is high in a densified network, the signaling cost of the network is reduced up to 26%. Moreover, when the MN becomes far from the core of the network, the signaling cost is reduced by up to 64%. The data delivery cost is reduced by 66% while the handover latency and the packet loss rate were reduced by 54%. Simulation results have validated the analytical results mentioned above. They revealed that the signaling cost is reduced by 36% while the data delivery cost is reduced by 62%. The handover latency and the packet loss were reduced by 61.5% and 90% respectively. Furthermore, we observed that the LISP-LSA protocol also reduces the processing load of the GW by up to 90%.

The third phase of the work involved the proposal of a mobility management architecture. The design of this architecture required the definition of its components and their internal

structure, as well as the definition of the control mechanisms. Control mechanisms define the operating procedures by which the proposed architecture is capable of managing user mobility. They involve an interaction between several components of the architecture. Subsequently, we proposed the analytical models of the throughput and the handover blocking probability as complementary metrics to the performance evaluation started during the previous phase. The proposed models were implemented with MATLAB for numerical evaluation. The architecture's components were implemented with the ns-3.26 network simulator to validate the numerical results. The numerical results showed that, depending on the distance between the MN and the GW, the network throughput is increased by about 32%. Moreover, depending on the size of the LSA, the throughput is improved by about 22%. The handover blocking probability is reduced by about 40% with smaller SC's radius. Indeed, we observed that the LISP-LSA protocol reduces the blocking probability by about 55% when the velocity of the MN is high while it is far from the core of the network. Simulation results subsequently validated the analytical results mentioned. We observed that, depending on the distance between the MN and the GW, the network flow is increased by up to 36%. Moreover, depending on the data packets arrival rate, the throughput increases by about 28%. When the network becomes denser, the network throughput is significantly improved by up to 90%. The handover blocking rate is considerably reduced up to 70%, according to the velocity of the MN and the distance with the GW. Nevertheless, we observed that the handover blocking rate generally tends to be high when the delay in the transit network increases.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	vi
ABSTRACT	x
TABLE DES MATIÈRES	xiv
LISTE DES TABLEAUX	xviii
LISTE DES FIGURES	xix
LISTE DES SIGLES ET ABRÉVIATIONS	xxi
LISTE DES ANNEXES	xxiv
CHAPITRE 1 INTRODUCTION	1
1.1 Définitions et concepts de base	2
1.1.1 Que désigne un réseau 5G ?	3
1.1.2 Mobilité	4
1.1.3 Taxonomie de la gestion de mobilité	5
1.1.4 Métriques de la Qualité de Service	6
1.1.5 Densité du réseau	8
1.1.6 Mécanisme de séparation identificateur/localisateur (ID/Locator Split)	9
1.2 Éléments de la problématique	10
1.3 Objectifs de recherche	13
1.3.1 Objectif global	13
1.3.2 Objectifs spécifiques	14
1.4 Principales contributions de la thèse et leur originalité	14
1.5 Plan de la thèse	16
CHAPITRE 2 REVUE DE LITTÉRATURE	18
2.1 Gestion de la mobilité au niveau de la couche TCP/IP	18
2.1.1 Mobilité au niveau de la couche réseau (MIPv6 et ses extensions)	18

2.1.2	Gestion de la mobilité au niveau des autres couches	21
2.1.3	Améliorations basées sur MIPv6 et ses extensions	27
2.2	Mécanismes distribués de la gestion de la mobilité	30
2.3	Gestion de la mobilité par l'approche du Software-defined Networking (SDN)	32
2.4	Protocoles basés sur le mécanisme ID/Locator Split	35
CHAPITRE 3 DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE .		39
3.1	Phase 1 : Conception du protocole de gestion de la mobilité	39
3.1.1	Modèle de densification du réseau	39
3.1.2	Protocole de mobilité LISP-LSA	40
3.1.3	Modèle analytique de représentation du LSA	40
3.1.4	Évaluation de performance	41
3.2	Phase 2 : Gestion de la mobilité avec support de relèves proactives	42
3.2.1	Relèves proactives et réactives	42
3.2.2	Métriques et modèles d'évaluation de performance	42
3.2.3	Implémentation	43
3.2.4	Évaluation de performance	43
3.3	Phase 3 : Conception de l'architecture de gestion de la mobilité	45
3.3.1	Composantes de l'architecture	46
3.3.2	Mécanismes de contrôle	46
3.3.3	Métriques et modèles d'évaluation de performance	47
3.3.4	Implémentation	47
3.3.5	Évaluation de performance	47
CHAPITRE 4 ARTICLE 1 : LOCAL ANCHOR BASED ARCHITECTURE FOR MO-		
BILITY SUPPORT IN IDENTIFICATOR LOCATOR SPLIT NETWORKS . . .		50
4.1	Introduction	51
4.2	Overview of the proposed Local anchor based mobility architecture	52
4.2.1	Local anchor based location management in localized LISP networks	52
4.2.2	Identifiers, Locators and the Mapping System	53
4.2.3	Location update signaling	54
4.3	Analytical model	56
4.3.1	Model Description	57
4.4	Performance Evaluation and Analysis	61
4.4.1	Intra-LSA handover cost	62
4.4.2	Inter-LSA handover cost	62
4.4.3	Cost definition	62

4.5	Numerical Results	63
4.6	Conclusion	66
CHAPITRE 5 ARTICLE 2 : A SEAMLESS MOBILITY MANAGEMENT PROTO-		
	COL IN 5G LOCATOR IDENTIFICATOR SPLIT DENSE SMALL CELLS . . .	67
5.1	Introduction	68
5.2	Background and Motivation	70
5.3	Overview of the proposed Local anchor based network architecture	73
	5.3.1 Intra-LSA Handover Scheme	73
	5.3.2 Inter-LSA Handover Scheme	74
5.4	Analytical Models	76
	5.4.1 Network Model	76
	5.4.2 Mobility and Traffic Models	78
5.5	Performance Evaluation and Analysis	79
	5.5.1 Total Signaling Cost	80
	5.5.2 Data Delivery Cost	82
	5.5.3 Handover Latency and Packet Loss	84
5.6	Analytical and Simulation Results	86
	5.6.1 Numerical Evaluation	86
	5.6.2 Simulation Setup	87
	5.6.3 Numerical and Simulation Results	88
	5.6.4 Discussion	95
5.7	Conclusion	96
CHAPITRE 6 ARTICLE 3 : AN ARCHITECTURE FOR MOBILITY MANAGE-		
	MENT IN 5G IDENTIFICATOR LOCATOR SPLIT DENSE SMALL CELLS . .	97
6.1	Introduction	98
6.2	Background and Motivation	101
	6.2.1 Background	101
	6.2.2 Motivation	102
6.3	Proposed Architecture	103
	6.3.1 Network Components	103
	6.3.2 Packet Routing	106
	6.3.3 Mobility Management	107
6.4	Analytical Models	108
	6.4.1 Assumptions	108
	6.4.2 Network Model	108

6.5	Performance Evaluation and analysis	110
6.5.1	Throughput Model	110
6.5.2	Handover Blocking Probability	113
6.6	Analytical and Simulation Results	116
6.6.1	Numerical Evaluation	116
6.6.2	Simulation Setup	116
6.6.3	Numerical and Simulation Results	117
6.7	Conclusion	124
CHAPITRE 7 DISCUSSION GÉNÉRALE		125
7.1	Synthèse des résultats	125
7.2	Méthodologie	126
7.3	Analyse des résultats	126
CHAPITRE 8 CONCLUSION ET RECOMMANDATIONS		127
8.1	Synthèse des travaux	127
8.2	Limitations de la solution proposée	128
8.3	Améliorations futures	130
RÉFÉRENCES		132
ANNEXES		146

LISTE DES TABLEAUX

Table 4.1	System parameters	63
Table 5.1	Notation	77
Table 5.2	Expression of the signaling cost functions	81
Table 5.3	Expressions of the binding update latencies	85
Table 5.4	Default values	87
Table 6.1	Notation	109
Table 6.2	Expressions of the binding update latencies	114
Table 6.3	Expressions of the packet delivery time	114
Table A.1	Simulation Parameters	148
Table B.1	Default values	151
Table B.2	Simulation Parameters	152

LISTE DES FIGURES

Figure 1.1	Architecture de haut niveau du réseau 5G	3
Figure 4.1	Local anchor based network architecture in localized LISP networks .	52
Figure 4.2	Signaling for the intra-LSA location update scheme	55
Figure 4.3	Signaling for the inter-LSA location update scheme	56
Figure 4.4	2D Grid model	57
Figure 4.5	LSA with bordering states	58
Figure 4.6	Intra-LSA handover signaling cost versus K ($m=0.1$)	64
Figure 4.7	Intra-LSA handover signaling costs versus K ($m=0.05$)	64
Figure 4.8	Inter-LSA handover signaling cost versus K ($m=0.1$)	65
Figure 4.9	Inter-LSA handover signaling cost versus K ($m=0.05$)	65
Figure 5.1	Local anchor based network architecture in localized LISP networks .	72
Figure 5.2	Intra-LSA handover	74
Figure 5.3	Inter-LSA handover	75
Figure 5.4	2D Grid-based network models and the Markov Model	78
Figure 5.5	Timing diagram of LISP-MN Handover	81
Figure 5.6	Timing diagram of LISP-LSA Handover	81
Figure 5.7	Simulated Network in Network Simulator NS3	88
Figure 5.8	Total Signaling Cost versus different parameters	89
Figure 5.9	Data Delivery Cost, Handover Latency and Packet Loss	91
Figure 5.10	Total Signaling Cost and Packet Loss versus different parameters . .	92
Figure 5.11	Handover Latency and Packet Loss versus different parameters	94
Figure 5.12	Total Signaling Cost versus different parameters	94
Figure 5.13	Processing Load on the GW versus different parameters	95
Figure 6.1	Overview of the proposed network architecture and components . . .	104
Figure 6.2	The model used for calculating the handover blocking probability . .	113
Figure 6.3	Simulated Network in Network Simulator NS3	117
Figure 6.4	Throughput versus E_S and $d_{LA,GW}$	118
Figure 6.5	Throughput versus E_S and λ_p	118
Figure 6.6	Throughput versus E_S and K	119
Figure 6.7	Handover Blocking Probability versus v and R	119
Figure 6.8	Handover Blocking Probability versus v and K	120
Figure 6.9	Handover Blocking Probability versus v and $d_{LA,GW}$	121
Figure 6.10	Throughput versus E_S and $d_{LA,GW}$	121

Figure 6.11	Throughput versus E_s and λ_p	122
Figure 6.12	Throughput versus v and λ_d	123
Figure 6.13	Handover Blocking Probability versus v and $d_{LA,GW}$	123
Figure A.1	Registration procedure	146
Figure A.2	Example of link configuration in LSA	147
Figure A.3	Handover latency versus m and $d_{LA,GW}$	149
Figure A.4	Packet Loss versus λ_p and MN's velocity	149
Figure B.1	The movement direction probabilities	151
Figure B.2	Handover Blocking Probability versus v and σ_{core}	153

LISTE DES SIGLES ET ABRÉVIATIONS

3G	Third Generation of Mobile Telecommunication Technology
4G	Fourth Generation of Mobile Telecommunication Technology
5G	Fifth Generation of Mobile Telecommunication Technology
AP	Access Point
BER	Bit Error Rate
BS	Base Station
BU	Binding Update
CES	Core Edge Separation
CMM	Centralized Mobility Management
CoA	Care-of-Address
DAD	Duplicate Address Detection
DAR	Dynamic Address Reconfiguration
DHCP	Dynamic Host Configuration Protocol
DMM	Distributed Mobility Management
DNS	Domain Name System
EID	End-Point Identifier
EID-to-RLOC	End-Point Identifier to Routing Locator
ETR	Egress Tunnel Router
FA	Foreign Agent
FBack	Fast Binding Acknowledgement
FTP	File Transfer Protocol
FMIPv6	Fast Handover for Mobile IPv6
GEID	Global End-Point Identifier
GEID-to-ELOC	Global End-Point Identifier to Egress Locator
GEID-to-LLOC	Global End-Point Identifier to Local Locator
GW	Gateway
HA	Home Agent
HI	Host Identity
HIP	Host Identity Protocol
HIT	Host Identity Tag
HMIPv6	Hierarchical Mobile IPv6
HoA	Home Address
IETF	Internet Engineering Task Force

ID-to-LOC	IDentifier to LOCator
ILS	Identifier/Locator Separation
IMM	Individual Mobility Model
IP	Internet Protocol
ITR	Ingress Tunnel Router
LA	Local Anchor
LCoA	Local Care-of-Address
LDC	Local Data Center
LISP	Locator and Identity Separation Protocol
LISP-DMC	LISP Distributed Mapping Control
LISP-MN	LISP Mobile Node
LISP-LSA	LISP Location Service Area
LLOC	Local Locator
LMA	Local Mobility Anchor
LSA	Location Service Area
LSC	Local Service Center
MAG	Mobile Access Gateway
MAP	Mobility Anchor Point
MIH	Media Independent Handover
MIPv4	Mobile IP version 4
MIPv6	Mobile IP version 6
METIS	Mobile and wireless communications Enables for the Twenty-twenty Information Society
MN	Mobile Node
MS	Mapping System
NAR	New Access Router
NLA	New Local Anchor
NRO	Non-Route Optimized
NxTR	New xTR
PAR	Previous Access Router
PLA	Previous LA
PMIPv6	Proxy Mobile IPv6
PxTR	Previous xTR
QoE	Quality of Experience
QoS	Quality of Service
RCoA	Regional Care-of-Address

RD	Rendezvous Domain
RLOC	Routing Locator
RO	Route Optimization
RSSI	Received Signal Strength Indicator
SC	Small Cell
SCTP	Stream Control Protocol
SDN	Software Defined Networking
SINR	Signal-to-Interference-plus-Noise Ratio
SIP	Session Initiation Protocol
SLA	Service-Level Agreement
SMR	Session-to-Mobility Ratio
TCP	Transmission Control Protocol
ULID	Upper Layer Identifier
UMTS	Universal Mobile Telecommunications Service
xTR	Tunnel Router
Wi-Fi	Wireless Fidelity
WiMax	Worldwide Interoperability for Microwave Access

LISTE DES ANNEXES

Annexe A	Supplemental Material of the paper <i>A Seamless Mobility Management Protocol in 5G Locator Identifier Split Dense Small Cells</i>	146
Annexe B	Supplemental Material of the paper <i>An Architecture for Mobility Management in 5G Identifier Locator Split Dense Small Cells</i>	150

CHAPITRE 1 INTRODUCTION

Au cours des dernières décennies, les réseaux de communications mobiles ont connu de profondes avancées technologiques. Parmi celles-ci, notons entre autres, l'évolution des technologies d'accès, l'intégration des mécanismes de communication basés sur le protocole Internet (IP), le support de services multimédia, la prolifération de périphériques mobiles aux performances et capacités accrues (smartphones, assistants personnels, tablettes mobiles..etc), l'offre de services forfaitaires donnant accès à des téléchargements de données illimitées. Ces progrès significatifs ont favorisé le développement massif de nouvelles applications et balisé le chemin pour de nouvelles possibilités d'utilisation de la connectivité mobile.

Les statistiques récentes révèlent une croissance du nombre d'abonnés mobiles ainsi qu'une demande de bande passante, de vidéo et de données toujours en forte croissance. Il est estimé que le trafic de données mobiles augmentera à un taux annuel de 46% entre 2017 et 2022 [1]. En effet, il est prévu que d'ici 2022, 99% du trafic de données mobiles généré proviendra essentiellement de périphériques intelligents. Cela représente une hausse de 92% par rapport à 2017. D'ici 2022, la vidéo mobile représentera près de 79% du trafic de données mobiles et connaîtra une croissance annuelle anticipée estimée à 35% jusqu'en 2024 [2].

Dans ce contexte, il est peu probable qu'une seule technologie d'accès puisse satisfaire aux prédictions mentionnées. En réalité, les technologies d'accès à haut débit continueront de manière inéluctable leur expansion, afin d'aboutir à des systèmes aux performances et aux capacités plus étendues. Aux technologies existantes s'ajouteront aussi de nouveaux modes de communication dans le but de répondre à plusieurs cas d'utilisations des réseaux mobiles qui sont encore difficiles à satisfaire à ce jour. Le résultat à long terme de cette nouvelle tournure dans le monde de la réseautique mobile est désigné sous le vocable de réseaux de cinquième génération (5G). De manière plus précise, la 5G est une référence à l'intégration transparente des versions évoluées des technologies de communication actuelles à de nouvelles technologies complémentaires, favorisant conjointement l'avènement du paradigme de la société connectée à partir des années 2020 [3].

La 5G définit la prochaine phase des technologies mobiles offrant une bande passante élevée, une couverture plus étendue et une latence ultra-faible [1]. Elle rendra possible le déploiement de nouveaux services, de nouveaux cas d'utilisation tels que l'Internet-des-objets (IoT), la réalité augmentée, la réalité virtuelle, les réseaux de villes intelligentes, les véhicules autonomes, et l'automatisation industrielle [4]. Au-delà du déploiement d'applications avancées, les réseaux 5G offriront de nouvelles opportunités de revenus aux fournisseurs de services lorsqu'ils

seront combinés aux fonctionnalités avancées telles que l’analyse de données, l’apprentissage automatique et à l’intelligence artificielle [1].

Dans le but de répondre convenablement aux exigences liées à l’avènement des réseaux 5G, un large consensus est aujourd’hui établi sur la nécessité d’accroître la capacité du réseau par un déploiement massif de cellules ou d’antennes d’un rayon de couverture réduit et à faible puissance [5, 6]. On parle alors d’une ultra-densification du réseau dont le but essentiel est de favoriser la proximité des points d’accès (AP) des utilisateurs finaux [7]. En plus de l’augmentation de la capacité totale du réseau, l’ultra-densification favorise une planification plus flexible, une efficacité spectrale, un gain énergétique sur les périphériques mobiles, et des économies substantielles aux opérateurs de réseaux [5]. Cependant, la densification du réseau est encore à ce jour, un enjeux majeur dans la gestion de la mobilité des périphériques mobiles et elle nécessite encore des investigations [8]. En effet, les utilisateurs mobiles (MNs) seront sujets à des relèves fréquentes qui pourraient nuire à la Qualité de l’Expérience Utilisateurs (QoE) [9].

Dans cette thèse, nous nous consacrons à la gestion de la mobilité dans les réseaux denses et ultra-denses 5G dans le but d’offrir aux usagers du réseau, une Qualité de Service (QoS) adéquate. Les métriques de QoS que nous considérons dans le cadre de notre étude sont : le coût de signalisation, le coût d’acheminement des données, la latence des relèves, le taux de perte de paquets, le débit du réseau, ainsi que la probabilité d’interruption d’une session durant une relève.

Ce chapitre d’introduction est organisé comme suit. Dans une première étape, nous définissons les termes ainsi que les concepts de base nécessaires à la compréhension du cadre de ce travail. Par la suite, nous présenterons les éléments de la problématique en lien avec la gestion de la mobilité dans les réseaux denses 5G. Les objectifs de la recherche sont ensuite présentés, suivis des principales contributions faisant l’originalité de cette thèse. Le présent chapitre prendra fin sur une présentation sommaire des chapitres subséquents de la thèse.

1.1 Définitions et concepts de base

Cette section présente et définit les concepts qui seront référencés dans le cadre de ce travail. Nous commencerons par la présentation des notions liées à la gestion de la mobilité au sein des réseaux mobiles, suivi de la présentation des métriques de QoS que nous considérons dans le cadre de ce travail. Nous définirons par la suite la notion de densité dans les réseaux cellulaires que nous avons adopté dans le cadre de cette thèse. Pour clore cette section, nous présenterons les notions liées au mécanisme de séparation Identificateur/Localisateur

(ID/Locator) seront ensuite présentés pour clore cette section.

1.1.1 Que désigne un réseau 5G ?

Selon Ericsson [3], la 5G désigne l'intégration transparente des versions évoluées des technologies de communication actuelles à de nouvelles technologies complémentaires, favorisant conjointement l'avènement du paradigme de la société connectée à partir des années 2020. La Figure 1.1 présente une illustration de haut niveau de l'architecture de la 5G.

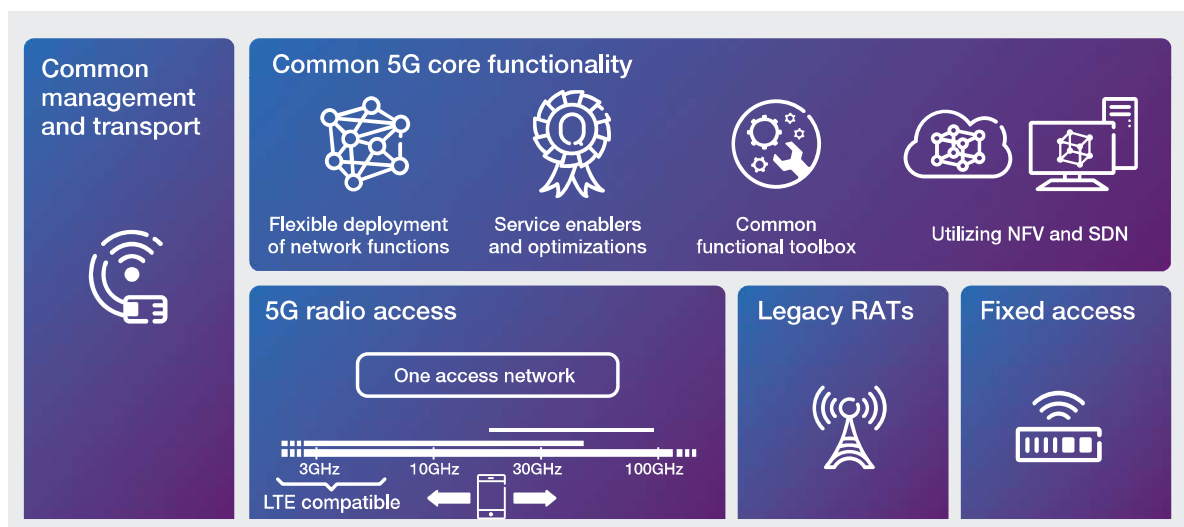


Figure 1.1 Architecture de haut niveau du réseau 5G

D'un point de vue architectural, la 5G définit une infrastructure d'intégration d'une diversité de technologies d'accès, comprenant à la fois les technologies existantes telle que le réseau LTE, de nouvelles technologies de réseaux d'accès fixes ou mobiles, tout en maintenant une compatibilité avec les technologies existantes. En outre, la 5G intégrera dans un environnement complètement virtualisé, flexible et programmable, des paradigmes émergents de la réseautique tels que le réseau défini par le logiciel ou Software-Defined Networking (SDN) et la virtualisation des composantes fonctionnelles du réseau ou Network Function Virtualization (NFV).

D'un point de vue du spectre, la 5G se verra attribuer de nouvelles bandes de fréquences dans le but de faire face à la demande de connexion en croissance des périphériques connectés, des réseaux extrêmement densifiés, et des nombreux cas d'utilisation de la connectivité mobile.

Du point de vue des utilisateurs et des clients, la 5G définira un écosystème d'accès à des services ubiquitaires, à des débits de données très élevés, à une latence très faible, à une QoS

et une QoE grandement améliorés.

1.1.2 Mobilité

Les réseaux mobiles du futur, dont ceux de la 5G, sont destinés à combiner en leur sein des infrastructures de télécommunication ainsi que des mécanismes de gestion avancée basés sur le protocole IP. Une grande diversité de réseaux d'accès ou de technologies de communication sans fils, permettront aux usagers ou noeuds mobiles (MNs) d'accéder aux services offerts à travers plusieurs points d'attaches au réseau ou points d'accès (APs).

Au sein des réseaux de communication, le terme *mobilité* fait référence à la possibilité d'accéder à des services, indépendamment de la localisation et du déplacement de l'utilisateur. On distingue trois types de mobilité : la mobilité terminale, la mobilité de service, et la mobilité personnelle [10]. Nous nous intéressons dans le cadre de ce travail à la mobilité terminale. La gestion de la mobilité quant à elle, regroupe deux composantes : la gestion de la localisation (*location management*) et la gestion de la relève (*handover management*) [11–13].

1.1.2.1 Gestion de la localisation

La gestion de la localisation permet au réseau d'accès d'identifier le AP courant du MN afin de lui acheminer ses données. Elle englobe les opérations d'enregistrement de la localisation (*location registration*), de mise à jour de localisation (*location update*), et de livraison de données (*data delivery*). L'enregistrement de la localisation permet au MN lors de sa première attache au réseau d'accès, de s'identifier en vue d'accéder aux services disponibles. Au cours de ses déplacements, le MN procède aux mises à jour périodiques de localisation afin que le réseau puisse retrouver à tout moment sa position courante. L'acheminement des données fait référence à la capacité du réseau à trouver l'emplacement courant du MN afin de lui transmettre ses données [11, 12].

1.1.2.2 Gestion de la relève

Une relève (*handover*) désigne le transfert automatique du canal de communication d'un PA vers un autre lors des déplacements du MN. La gestion de la relève permet donc au réseau d'assurer et de maintenir une continuité des communications en cours lors des relèves successives. Elle s'exécute en trois étapes [14] :

- **La phase d'initiation** : durant cette phase, le MN ou un noeud du réseau identifie la nécessité de procéder à une relève. Cette phase peut également s'amorcer suite à un changement des conditions courantes du réseau.

- **La phase de préparation ou de sélection** : elle permet d'établir une connectivité au niveau de la couche de liaison de données (L2) et de maintenir la connectivité au niveau de la couche IP [11]. Durant cette phase, dont le rôle clé est de décider du moment de transfert du MN vers le nouveau AP, le réseau doit aussi mobiliser les ressources nécessaires à l'opération [14]. Un algorithme de décision s'exécute en tenant compte de plusieurs paramètres tels que l'indicateur de la force du signal reçu (RSSI), le ratio signal sur interférence plus bruit (SINR), le taux d'erreur par bits (BER)...etc.
- **La phase d'exécution** : durant celle-ci, une signalisation de relève permet de transférer l'ensemble des informations associées au MN (son contexte de mobilité), ainsi que ses données de l'AP de départ vers le nouveau AP du MN. Cette phase permet d'assurer la continuité des sessions du MN [11].

1.1.3 Taxonomie de la gestion de mobilité

La gestion de la mobilité implique plusieurs interactions entre le MN et les APs du réseau dans le but de lui garantir une continuité de sessions. Plusieurs facteurs permettent de caractériser une relève. On peut citer par exemple le type de réseau d'accès, le nombre de liens maintenus par le MN avec les APs durant la relève, le domaine administratif impliqué, le moment de la décision de relève, l'entité d'exécution et d'assistance de la relève, ainsi que la stratégie d'exécution de la relève.

- **Type du réseau d'accès** : on parle de *relève horizontale* lorsque les PAs impliqués durant la relève sont de la même technologie. Dans le cas contraire il s'agit d'une *relève verticale*. Dans le cadre de cette thèse, nous considérons des relèves horizontales.
- **Nombre de points d'attache au réseau** : on parle de relève dure (*hard handover* ou *break-before-make*) lorsque le MN libère sa connexion au AP précédent avant d'établir une nouvelle liaison sur le AP de destination. On parle de relève en douceur (*soft handover* ou *make-before-break*) lorsque le MN établit une nouvelle liaison sur le AP de destination avant de libérer sa connexion avec le AP précédent. Dans le cadre de cette thèse, nous considérons des relèves dures.
- **Domaine administratif** : le domaine administratif identifie un groupe de systèmes et de réseaux administrés par une seule entité ou autorité d'administration. On parle de *relève intra-domaine administratif* lorsque les deux APs impliqués à la relève sont localisés dans le même domaine administratif. Par contre, une relève *inter-domaine administratif* implique deux APs situés dans deux domaines administratifs distincts. Dans le cadre de cette thèse, nous avons considéré une unité d'administration du réseau avec le concept de zone de localisation de service (*Location Service Area, LSA*). Une

LSA désigne un ensemble de APs administrés par un noeud du réseau appelé point d’ancrage local (*Local Anchor, LA*). Ainsi, nous distinguons les relèves intra-LSA, de même que les relèves inter-LSA.

- **Moment de la décision de relève** : l’implication du MN durant la procédure de relève permet de distinguer une *relève proactive* ou une *relève passive*. Lors d’une relève proactive, le MN décide du moment du transfert du lien d’accès vers le nouveau AP. Cependant, lorsque le transfert du lien ne s’effectue pas à tant, une rupture du lien sans fil survient et la relève passe en *mode réactif*. Le mode réactif nécessite généralement une nouvelle authentification au réseau, suite à l’échec du transfert du contexte de mobile du MN. Lors d’une *relève passive*, le MN n’a aucun contrôle sur la décision de transfert. Dans le cadre de notre travail, nous considérons le mode proactif et le mode réactif lors de la phase de décision de la relève.
- **Entité de décision et de contrôle** : selon l’entité qui décide de la phase de contrôle de la relève, une relève peut être contrôlée par le MN et assistée par une entité du réseau et vice-versa. Lorsque la relève est contrôlée par le MN (*Mobile-Controlled Handover*), ce dernier collecte en toute autonomie les informations qui lui sont nécessaires et décide de son exécution. Dans le cas d’une relève contrôlée par le réseau (*Network-Controlled Handover*), une entité du réseau pilote la phase de décision. La relève est dite assistée par le MN (*Mobile-Assisted Handover*) lorsque les informations nécessaires à la prise de décision sont collectées par le MN et transmises ensuite à l’entité du réseau qui prend la décision d’exécution. Dans le cas contraire, on parle de relève assistée par le réseau (*Network-Assisted Handover*). Dans le cadre de cette thèse, nous avons adopté une approche collaborative dans laquelle, les relèves sont contrôlées par le MN et assistées par le réseau.
- **Stratégie d’exécution de la relève** : lorsque l’objectif visé lors d’une relève est de minimiser le temps de déconnexion du MN, on parle de *fast handover*. Par contre, lorsque l’objectif est de réduire le nombre de paquets perdus durant la relève, on parle de *smooth handover*. Lors d’une relève transparente (*seamless handover*), l’objectif visé est le support de la QoS ainsi que la satisfaction des préférences de l’utilisateur. Elle implique donc une relève rapide sans une dégradation de performance perçue par l’utilisateur.

1.1.4 Métriques de la Qualité de Service

Les opérations de gestion de la mobilité engendrent une consommation des ressources du réseau. Il s’avère donc indispensable de procéder à une gestion efficiente des ressources du réseau lors de la mobilité des MNs, afin de garantir une QoS adéquate à ses usagers.

La QoS regroupe un ensemble de propriétés mesurables en lien avec la performance de bout-en-bout d'un service réseau. Elle est garantie à l'avance par un contrat de niveau de service (SLA) entre un utilisateur et un fournisseur de services afin de répondre aux besoins spécifiques des applications client [15]. La QoS a une incidence directe sur la QoE. Contrairement à la QoS, la QoE est une mesure subjective de performance qui dépend de la perception de l'utilisateur. Plusieurs métriques servent de cadre d'évaluation de la performance des mécanismes ou protocoles de gestion de la mobilité. Dans le cadre de cette thèse nous avons adopté les métriques suivantes :

- **Le coût de signalisation** : le coût de signalisation (*signaling cost*) désigne le nombre de messages de signalisation transmis durant la procédure de relèvement. Cette métrique affecte la performance du protocole de mobilité et reflète son efficacité. Il est désirable que le coût de signalisation soit minimisé.
- **Le coût de livraison des données** : le coût de livraison des données (*data delivery cost*) permet de quantifier la consommation des ressources du réseau par le protocole de mobilité, afin d'acheminer les données d'une source vers une destination. Il est désirable que cette métrique soit minimisée.
- **La latence de la relève** : la latence d'une relève (*handover latency*) désigne le temps durant lequel le MN ne peut recevoir ni transmettre des données, suite à la rupture du lien sans fil durant la relève. C'est le temps écoulé entre le début et la fin de la relève. Ce temps est évalué en sommant plusieurs facteurs à savoir : le temps de relève au niveau de la technologie d'accès (*Layer-2 handover time*), le temps de détection du mouvement du MN (*movement detection time*), et le temps de configuration de l'adresse IP (*IP configuration time*).
- **Taux de perte de paquets** : Le taux de perte de paquets (*packet loss*) permet de quantifier le nombre de paquets de données que le protocole de mobilité n'a pas réussi à transmettre durant la relève. Il est désirable de minimiser la perte de paquets lors d'une relève car cette métrique influence les applications temps réel.
- **Le débit** : Le débit (*throughput*) permet de mesurer le taux de paquets délivrés avec succès au MN durant une session de communication.
- **La probabilité d'interruption d'une session** : la probabilité d'interruption d'une session durant une relève (*handover blocking probability*), désigne l'éventualité qu'une session amorcée subisse une interruption abrupte durant une relève. Cette interruption peut être causée par un manque de ressources sur le AP de destination, ou lorsque le temps de résidence du MN sur le AP de destination est inférieur à la durée nécessaire à l'exécution complète de la signalisation de mobilité. Dans ce cas, le MN entame une nouvelle relève avant la fin de celle déjà en cours d'exécution. Dans cette thèse, nous

avons consid   le temps de r  sidence dans une SC, ainsi que le temps de signilation d      la rel  ve comme principaux facteurs d'interruption.

1.1.5 Densit   du r  seau

La tendance de croissance du trafic mesur   au cours des derni  res ann  es force un d  ploiement massif de cellules d'un rayon de couverture r  duit (*Small Cell* ou SC) afin d'accro  tre la capacit   globale du r  seau [7, 16]. Dans un r  seau de troisi  me g  n  ration (3G) on peut d  nombrer environ quatre    cinq stations de base (BS) au kilom  tre carr   (Km²), et environ huit    dix dans les r  seaux de quatri  me g  n  ration (4G) [8]. La densit   d'un r  seau 5G a   t   d  finie selon deux approches dans la litt  rature [7].

1.1.5.1 D  finition

D  signons respectivement par λ_b et λ_u le nombre de SCs et de MNs au Km².

Dans [17], un r  seau est consid  r   dense lorsqu'un grand nombre de SCs est d  ploy   comparativement au nombre de MNs. Autrement dit, $\lambda_b \gg \lambda_u$.

Ind  pendamment de λ_u , la densit   du r  seau est aussi d  finie selon une mesure quantitative. Selon [8], un r  seau est dense lorsque $40 \leq \lambda_b \leq 50$. Selon [18], un r  seau est ultra-dense lorsque $\lambda_b \geq 1000$.

Dans le cadre de cette th  se, nous consid  rons pour la suite du travail qu'un r  seau est dense lorsque $\lambda_b \geq 40$ et ultra-dense lorsque $\lambda_b \geq 1000$. Dans la suite du document, l'expression *densification du r  seau* permettra de d  signer un r  seau dense et un r  seau ultra-dense.

1.1.5.2 Mod  le de densit   et nature des cellules consid  r  es

Nous consid  rons dans le cadre de cette th  se un mod  le de densification horizontal [7] pouvant   tre d  ploy   dans un environnement ext  rieur (*outdoor*), un environnement int  rieur (*indoor*) ou un voisinage de SCs (*neighborhood small cell*) [19]. Un voisinage de SCs d  finit une configuration dite *indoor-to-outdoor* dans laquelle des SCs d  ploy  es dans un environnement *indoor* permettent aux usagers d'un environnement *outdoor* (pi  tons, v  hicules    faible vitesse) d'acc  der au r  seau. Deux types de SCs sont consid  r  s    savoir : les pico-cellules dont le rayon de couverture peut atteindre 100 m  tres et les femto-cellules dont le rayon de couverture est inf  rieur    30 m  tres [7]. Nous supposons que les MNs du r  seau ont une v  locit   maximale de 10 m  tres par secondes.

Dans le cadre de notre travail, nous consid  rons qu'une LA est une pico-cellule, tandis que

les SCs du LSA sont des femto-cellules.

1.1.6 Mécanisme de séparation identificateur/localisateur (ID/Locator Split)

Le mécanisme d'ID/Locator Split répartit en deux blocs distincts la couche réseau et les couches supérieures de la couche TCP/IP. Il s'agit du bloc de l'IDentificateur (ID) et du bloc du Localisateur (Localisateur). Au sein de l'architecture actuelle du protocole IP, tous les MNs du réseau sont à la fois identifiés et repérés dans l'espace de routage par une adresse IP. Une architecture basée sur le concept d'ID/Locator Split utilise des valeurs distinctes d'identificateurs et de localisateurs. Cela permet à la couche réseau de changer de localisateur sans devoir modifier les identificateurs de sessions précédemment établis au niveau des couches supérieures [20]. Cette section présente quelques concepts clés liés au mécanisme de séparation d'ID/Locator Split à savoir :

- **L'identificateur d'un MN** : un identificateur désigne le nom attribué à un MN afin de le distinguer de manière unique dans le réseau. L'identificateur est représenté sous la forme d'une série de chiffres, de caractères, de symboles ou représenté sous une autre forme. Les identificateurs sont généralement moins lisibles (binaires) et sont obtenus en combinant un nom et une fonction de hachage. Ils peuvent être partagés au sein d'un certain nombre de réseaux spécifiques, publics, ou privés [21] ;
- **Le localisateur d'un MN** : le localisateur d'un MN désigne sa localisation courante dans la topologie du réseau. En d'autres termes, le localisateur désigne l'identification au niveau de la couche réseau d'une interface ou d'un ensemble d'interfaces [21]. Les localisateurs sont indiqués dans le champ d'adresse IP lors de l'acheminement des données à travers le réseau. Une fonction de mappage permet d'associer un identificateur à un ou plusieurs localisateurs ou vice versa ;
- **La fonction de stockage identificateur/localisateur** : La fonction de stockage dans le mécanisme ID/Locator Split (*ID/Loc Mapping Storage Function*) enregistre l'information d'association entre l'identificateur et le localisateur d'un MN. Cette fonction met également à jour les informations d'association et les fournit aussi à d'autres fonctions ou à d'autres requêtes. Au sein d'une architecture ID/Locator Split, la fonction de stockage peut être implantée au niveau du MN. Elle peut également être déléguée à une composante fonctionnelle du réseau [21] ;
- **La fonction d'association identificateur/localisateur** La fonction d'association ou de mappage identificateur/localisateur (*ID/Locator mapping function*) récupère l'information de correspondance disponible au sein de la fonction de stockage et l'utilise au niveau des entêtes de paquets. Les composantes architecturales destinées à

exécuter les fonctions de mappage communiquent entre elles afin de distribuer, de mettre à jour et de faciliter l'accès à l'information de mappage. Elles interrogent les composantes de stockage à travers une requête de mappage (*mapping request*) afin de retirer l'information d'association correspondant à l'identificateur d'un MN. Si l'information est trouvée dans la base de données, celle-ci est retournée dans une réponse de mappage (*mapping response*) par la composante de stockage. Dans le cas contraire, la requête est transférée à une autre composante de stockage jusqu'à ce que l'information recherchée soit trouvée. Plusieurs composantes architecturales dédiées aux fonctions de mappage peuvent être déployées au sein de l'architecture du réseau afin d'améliorer le routage en terme de sécurité, de fiabilité et de QoS.

1.2 Éléments de la problématique

Autour de 2020, près de cinquante milliards de périphériques seront connectés. Ils nécessiteront un accès ubiquitaire à l'infrastructure du réseau 5G qui devra supporter des applications, des services et des cas d'utilisation révolutionnaires [22]. A cet effet, les réseaux IP existants nécessitent d'importantes mises à jour architecturales dans le but d'évoluer vers la 5G. La densification du réseau, par le déploiement de cellules d'un rayon de couverture plus réduit est un facteur clé qui rendra possible l'atteinte des objectifs de performance visés par la 5G [8, 23]. Elle permettra d'accroître la capacité du réseau [24], d'offrir une meilleure stabilité des liens et un débit de transfert de données élevé. Elle rendra possible la mise en oeuvre d'une meilleure efficacité énergétique et spectrale [24]. Cependant, la gestion de la mobilité au sein des réseaux denses demeure encore un enjeu important [6, 23]. Elle implique un transfert des MNs entre les SCs et les zones de service entraînant des dégradations, des perturbations et des déconnexions qui peuvent entraver l'objectif d'un accès transparent aux services du réseau. En effet, le déploiement massif de SCs dotés d'un rayon de couverture réduit, rend plus complexe la phase de sélection des relèves [8], en plus d'accroître la fréquence de celles-ci [9]. En outre, la fréquence des relèves engendre une latence ainsi qu'une charge de signalisation élevées dans le réseau. D'une part, l'émergence de nouvelles applications telles que l'Internet tactile, la réalité virtuelle, les jeux multijoueurs en ligne, les communications de type machine exigent une latence faible du réseau. Ces applications nécessitent des interactions en temps réel avec l'utilisateur, et une latence élevée du réseau entraînera une dégradation de la QoE [6]. D'autre part, dans un contexte de densification du réseau, le raccordement individuel des SCs au réseau coeur est une solution très coûteuse. Il s'avère donc nécessaire de mettre en oeuvre une approche de raccordement moins onéreuse par le biais de liens à capacités élevées et à faible latence [19]. Considérant l'avènement de l'IoT et l'explosion du nombre de

périphériques mobiles connectés, une forte charge de signalisation due à la mobilité nuirait à l'évolutivité du réseau ainsi qu'à ses performances. Il s'avère donc indispensable de concevoir des mécanismes de gestion de la mobilité pouvant répondre aux exigences de performances strictes des réseaux 5G [22]. Les modifications architecturales nécessaires à la transition vers la 5G, impliquent aussi une évolution des protocoles existants pour un contrôle et une gestion adéquates. En particulier, de nouveaux protocoles de gestion de la mobilité en 5G sont encore nécessaires pour garantir des performances optimales [22]. Par ailleurs, l'architecture des réseaux actuels est encore fortement centralisée et hiérarchique [25]. La charge de signalisation de contrôle est gérée au niveau du réseau coeur et devrait encore augmenter de manière significative à cause du nombre croissant de périphériques mobiles connectés nécessitant un accès ubiquitaire aux services du réseau. À plus long terme, les réseaux 5G qui, inévitablement seront densifiés, entraîneront un coût de signalisation encore plus élevé pour la gestion de la mobilité [26].

Aujourd'hui, la gestion de la mobilité encore centralisée (CMM), repose sur des plans de contrôle et de données convergeant vers la même entité ou passerelle de paquets. Dans cette configuration, ladite passerelle est à la fois responsable de la signalisation de mobilité et du transfert des données du MN en lui offrant un service de retransmission. L'avantage de l'approche centralisée réside dans sa simplicité, car l'existence du point d'ancrage permet de suivre les mouvements du MN par le biais de tunnels créés avec les routeurs d'accès auxquels les usagers mobiles du réseau sont actuellement connectés [27]. Les deux protocoles les plus représentatifs de la CMM sont Mobile IPv6 (MIPv6) [28] et Proxy Mobile IPv6 (PMIPv6) [29]. MIPv6 introduit un agent de mobilité au sein du réseau domestique comme point d'ancrage de la mobilité, tandis que PMIPv6 tente de gérer la mobilité localement dans le réseau visité en introduisant un agent centralisé de mobilité locale (LMA). Cependant, avec l'explosion du nombre de périphériques mobiles connectés, l'approche centralisée est inadéquate et a révélé ses limites [30]. La centralisation de l'architecture du réseau entraîne une augmentation significative de la charge de signalisation sur le point d'ancrage du trafic, une faible évolutivité, un routage sous-optimal, l'existence d'un point de rupture dont la défaillance entraînerait une dégradation significative de performance globale dans le réseau, des failles de sécurité, ainsi qu'une utilisation sous-optimale des ressources du réseau [30, 31]. Dans ce contexte, l'IETF a reconnu ces inconvénients et a spécifié un cadre préliminaire de gestion de la mobilité décentralisée appelé gestion de la mobilité distribuée (DMM) [32]. Aujourd'hui, la DMM est l'une des alternatives considérées à l'approche centralisée. Elle répartit le trafic du MN sur les PAs qui lui sont les plus proches, contribuant ainsi à la décentralisation des architectures des réseaux mobiles [26, 31]. Les protocoles basés sur la DMM étant encore récents, les solutions proposées reposent principalement sur le protocole

réseau PMIPv6 dont les limitations sont connues.

Cependant, malgré le fait que plusieurs approches de la DMM visent l'atténuation des problèmes liés à la gestion centralisée de la mobilité, elle ne répondent pas toujours aux besoins de performance désirés. De récentes études de performance ont montré des limitations considérables de leur performance, lorsque le temps de résidence des MNs au sein des SCs est court, ou lorsque le nombre de sessions actives amorçées au sein des réseaux visités est élevé [26, 33, 34]. De plus, la plupart des approches distribuées ne tiennent compte que d'une seule relève [34]. Lorsque la durée des sessions du MN est longue, elle peuvent s'étirer sur plusieurs relèves. Dans ce cas, l'efficacité de mécanismes distribués dépend en partie de la durée des sessions versus le temps de résidence des MNs au sein des SCs [34]. Dans un réseau densifié de SCs, lorsqu'un MN se déplace fréquemment et qu'il initie de nouvelles sessions au sein des différents réseaux visités, les performances des approches distribuées diminuent en raison du nombre élevé de tunnels devant être configurés et gérés par les points d'ancrage distribués [26]. Par conséquent, si la proportion de ces longs flux est trop importante, les approches distribuées risquent d'être très peu performantes au sein des réseaux denses 5G.

De plus, les mécanismes de gestion de la mobilité actuels, y compris ceux de la DMM, manquent d'un support de mobilité adaptable à un profil de mobilité spécifique ou aux besoins spécifiques d'un MN. L'approche distribuée offre le même support de gestion de la mobilité à tous les MNs, quel que soit leur profil ou les exigences des l'application en cours d'exécution [35]. Cependant, contrairement aux générations précédentes des réseaux mobiles, il est reconnu que les périphériques utilisateurs sont appelés à jouer un rôle plus important dans l'écosystème des réseaux 5G [36]. En raison de la diversité des applications, des services et des cas d'utilisation émergents, offrir le même support de mobilité serait contre performant. Il est donc indispensable de mettre en oeuvre dans la 5G, des mécanismes de gestion de la mobilité flexibles, adaptables aux exigences des applications des MNs, mais qui impliquent davantage les périphériques mobiles [35, 36].

Étant tributaires du protocole MIPv6 et de ses dérivés, les approches CMM et DMM héritent de la surcharge sémantique de l'adressage IP qui limite ses performances pour la gestion de la mobilité. Avec l'augmentation du nombre de périphériques mobiles connectés, l'explosion de la taille des tables de routage devient un enjeu majeur pour la gestion de la mobilité. La dualité sémantique de l'adresse IP provient du fait qu'elle indique à la fois la position topologique d'un hôte et sert aussi à l'identifier au niveau des couches application et transport. Cependant, l'efficacité du routage est mieux réalisée lorsque l'adressage est de préférence découplée du contexte topologique [37]. Ainsi, la dualité fonctionnelle inhérente à l'adressage IP, ne convient plus à son utilisation dans les réseaux hétérogènes 5G pour la gestion de mobilité. Afin de

parer aux limitations inhérentes au protocole IP pour la gestion de la mobilité et faciliter un routage internet évolutif, le mécanisme d’ID/Locator Split a été proposé. Les protocoles basés sur ce mécanisme séparent l’identificateur et le localisateur d’un MN et permettent une gestion plus flexible de la mobilité. Elle supportent de manière native une domiciliation multiple des MNs. Cependant, la performance de ces protocoles repose essentiellement sur le système de résolution sous-jacent. En particulier, leur évolutivité devient critique dans un environnement mobile en présence de plusieurs périphériques pour lesquels, la mise à jour de localisation deviendra fréquente au niveau des composantes de mappage et de stockage du système de résolution.

En effet, à chaque relève effectuée par le MN, son nouveau localisateur doit être mis à jour dans le système de mappage souvent centralisé et localisé dans le réseau coeur. Cette approche, sujette à une signalisation élevée, est inadéquate dans un contexte de SCs densifiées ou les MNs à vitesse élevée subiront des relèves fréquentes. En outre, si les informations de mappage du MN sont stockées au niveau de plusieurs routeurs distants, la fréquence des relèves induira des informations de localisation obsolètes. Par conséquent, les paquets de données destinés à un MN seront transmis à une SC à laquelle le MN ne serait plus attaché [38]. En plus du trafic de signalisation additionnel généré pour la mise à jour de localisation, le taux de perte de paquets est aussi élevé à cause d’une latence de relève élevée. La continuité des sessions actives nécessite la mise en oeuvre de mécanismes qui tirent pleinement parti de l’infrastructure IP pour un support transparent de la mobilité. De plus, il est indispensable de maintenir une QoS adéquate à tous les types de trafic de données. Au regard de ce qui précède, une question fondamentale se pose :

- Comment mettre en oeuvre une gestion de la mobilité efficiente dans les réseaux denses 5G ?

1.3 Objectifs de recherche

Cette section présente l’objectif global ainsi que les objectifs spécifiques de cette thèse.

1.3.1 Objectif global

L’objectif global de cette thèse est de concevoir une architecture de gestion de la mobilité dans les réseaux denses 5G.

1.3.2 Objectifs spécifiques

Afin d'atteindre cet objectif global, nous avons établi plusieurs objectifs spécifiques qui s'énumèrent comme suit :

1. Concevoir un protocole de gestion de la mobilité dans les réseaux denses 5G en utilisant le mécanisme ID/Locator Split ;
2. Étendre le fonctionnement du protocole proposé afin d'intégrer les mécanismes de relèves proactives à la gestion de la mobilité dans les réseaux denses 5G ;
3. Concevoir une architecture de gestion de la mobilité dans les réseaux 5G ;
4. Implémenter une partie des composantes de l'architecture ;
5. Évaluer les performances du protocole et de l'architecture proposée.

1.4 Principales contributions de la thèse et leur originalité

La principale originalité de cette thèse repose sur la mise en oeuvre du mécanisme d'ID/Locator Split pour la gestion de la mobilité dans les réseaux denses 5G avec un support de la QoS.

Les principales contributions portent sur la proposition d'un modèle de densification du réseau, la proposition d'un modèle mathématique de représentation du réseau densifié, la proposition d'un protocole de gestion de la mobilité avec un support de la QoS, la proposition de modèles analytiques d'évaluation de performance du protocole de mobilité, la proposition d'une architecture de gestion de la mobilité. Elle sont détaillées comme suit :

1. **Nouveau modèle de densification du réseau.** Dans un contexte de densification du réseau par le déploiement massif de SCs afin d'en accroître sa capacité, notre première contribution a été la proposition d'un modèle horizontal (2D) de densification du réseau. Nous avons proposé le concept de zone de localisation de service (Location Service Area, LSA) dans un réseau dont le fonctionnement est basé sur le mécanisme d'ID/Locator Split. Dans cette configuration, une LSA est localement gérée par une nouvelle composante du réseau appelée point d'ancrage local (Local Anchor, LA). Ce modèle présente plusieurs avantages. Il est applicable autant dans un contexte indoor (espace résidentiel, bureaux, grandes surfaces...etc) que dans un environnement outdoor (parcs extérieurs, zones de quartiers dont la vitesse de circulation est très réduite...etc) afin d'offrir une connectivité aux usagers du réseau. La proximité du LA de la périphérie du réseau favorise un gestion locale de la mobilité au sein des architectures ID/Locator Split car le temps de résolution identificateur/localisateur est considérablement réduit. En outre, d'un point de vue économique, le raccordement

des SCs au LA réduit le coût de déploiement global du réseau densifié en évitant des liaisons directes des SCs au réseau coeur. Cette approche réduit le trafic de signalisation de la mobilité qui se trouve désormais restreinte à une partie délimitée du réseau. Mieux encore, elle favorise un allègement de la charge totale de la passerelle du réseau (GW) en la répartissant sur l'ensemble des LAs. Par ailleurs, l'utilisation du mécanisme d'ID/Locator Split dans ce modèle de densification favorise une évolutivité du réseau, car l'aggrégation de route permet de réduire la taille des tables de routage. De plus, ce mécanisme favorise la domiciliation multiple des MNs et renforce la résilience du réseau ;

2. **Modèle mathématique de représentation du réseau densifié.** Notre deuxième contribution a été la proposition d'un modèle mathématique du réseau densifié. A cet effet, nous avons représenté un LSA et l'ensemble des SCs qu'il contient par une chaîne de Markov (2D) à temps discret. Ce modèle permet de suivre le déplacement et l'activité d'un MN à travers le réseau. Le modèle intègre plusieurs paramètres tels que la taille du LSA, le taux d'arrivée et la durée des sessions du MN, le temps de résidence du MN au sein d'une SC. Ce modèle a permis d'évaluer la probabilité d'occurrence d'une relève intra-LSA et celle d'une relève inter-LSA ;
3. **Protocole de gestion de la mobilité avec des mécanismes de relève proactives.** Plusieurs protocoles existent dans la littérature pour la gestion de la mobilité basée sur le mécanisme ID/Locator Split. Cependant, au meilleur de nos connaissances, aucun de ces protocoles ne tenait compte d'une mise en oeuvre dans un contexte densifié de réseaux 5G. Notre troisième contribution a été la proposition d'un protocole de gestion de la mobilité appelé LISP-LSA. Ce protocole se base sur le modèle de densification proposé ci-dessus et s'adapte à deux types de relèves : Les relèves intra-LSA et les relèves inter-LSA. Mieux, le protocole utilise les informations des autres couches de la pile TCP/IP afin de supporter des relèves proactives (anticipées) pour un meilleur support de la QoS. Cette décision se base sur une approche collaborative de gestion des relèves contrôlées par le MN et assistées par le réseau afin d'intégrer les préférences utilisateur lors des décisions de relève. De plus, la gestion locale de la mobilité permet de réduire le coût de signalisation, la latence des relèves, le taux de perte de paquets, la probabilité d'interruption des sessions et améliore considérablement le débit du réseau. Le protocole proposé favorise des relèves transparentes et garantit une meilleure QoE ;
4. **Modèles analytiques d'évaluation de performance.** L'évaluation de performance des protocoles ID/Locator Split existant était basée sur des modèles analytiques très limités qui ne pouvaient pas s'appliquer à un réseau densifié. En nous basant sur le

modèle horizontal 2D du LSA, notre quatrième contribution a été la proposition de modèles analytiques détaillés de plusieurs métriques d'évaluation de performance telles que le coût de signalisation, le coût de livraison des paquets, la latence des relèves, la perte de paquets, le débit, la probabilité d'interruption des sessions durant une relève. Mieux, les modèles proposés tiennent compte de l'impact conjoint de plusieurs paramètres du réseau (durée des session, taux d'arrivée des paquets, taille du LSA, temps de résidence au sein des SCs) sur les métriques évaluées ;

5. **Architecture de gestion de la mobilité.** Notre cinquième contribution a été la proposition d'une nouvelle architecture de gestion de la mobilité dans les réseaux denses 5G. Cette architecture réutilise le concept du LSA, ainsi que le protocole LISP-LSA ci-dessus mentionnés. Dans cette architecture, le trafic de signalisation est séparé du trafic de données. La signalisation du réseau est gérée au niveau d'un réseau overlay dont le fonctionnement est basé sur le protocole LISP. Le plan de données quant à lui est composé d'un ensemble de routeurs tunnels dont les fonctionnalités ont été étendues. Les composantes essentielles de l'architecture proposée ont été implémentées dans le simulateur de réseau ns-3.26 afin de valider les résultats numériques de notre étude par des résultats de simulation. Les métriques évaluées ont montré de meilleures performances en comparaison à l'existant dans la littérature.

1.5 Plan de la thèse

Dans ce chapitre, nous avons présenté les définitions et concepts de base de notre étude, énoncé la problématique, présenté les objectifs de notre recherche ainsi que les principales contributions de cette thèse. Le chapitre 2 présentera une revue sélective de la littérature. Nous présenterons les mécanismes de gestion de la mobilité à différents niveaux de la couche TCP/IP, les approches distribuées de gestion de mobilité, les protocoles basés sur le mécanisme ID/Locator Split.

Dans le chapitre 3, nous présenterons la démarche d'ensemble du travail en montrant la corrélation entre les objectifs présentés à la Section 1.3 et les articles réalisés dans le cadre de cette thèse.

Le chapitre 4 présente le texte intégral de l'article intitulé « *Local Anchor-Based Architecture for Mobility Support in Identifier Locator Split Networks* ». Cet article introduit le concept de zone de localisation de service (LSA) dans un domaine local LISP. Une LSA définit une structure de gestion de la mobilité au niveau local. Par la suite, nous avons proposé un modèle mathématique de représentation du LSA en utilisant une chaîne de Markov à deux dimensions à temps discret. Le protocole de gestion de mobilité LISP-LSA a été introduit dans ce travail

afin de définir les différents scénarios de relève des MNs. Une évaluation de performance, faite sur la base du modèle analytique, nous a permis d'obtenir des résultats qui ont fait l'objet d'un article publié dans le cadre de la conférence ICC2017 (*Second International Conference on Internet of things, Data and Cloud Computing. ACM, Mars 2017*).

Le chapitre 5 présente le texte intégral de l'article intitulé « *A Seamless Mobility Management Protocol in 5G Locator Identifier Split Dense Small Cells* ». Dans cet article, le protocole de mobilité LISP-LSA a été étendu afin d'intégrer le support de relèves proactives à la gestion de la mobilité. Pour chaque mode de relève (intra-LSA et inter-LSA) proposé dans l'article précédent, de nouveaux mécanismes de relève proactive et réactive ont été introduits afin de permettre des relèves rapides. De nouveaux modèles analytiques d'évaluation de performance ont été proposés en tenant compte des scénarios de relèves anticipées qui ont été proposés. Le protocole LISP-LSA a été ensuite implémenté avec le simulateur de réseau ns-3.26 afin de valider par une approche de simulation, les résultats numériques obtenus. Cet article est en révision mineure pour publication dans la revue *IEEE Transactions on Mobile Computing*.

Le chapitre 6 présente le texte intégral de l'article intitulé « *An Architecture for Mobility Management in 5G Identifier Locator Split Dense Small Cells* ». Dans cet article, l'architecture complète de la gestion de mobilité est présentée en intégrant le concept de LSA et le protocole LISP-LSA présentés dans les articles précédents. Les différentes composantes de l'architecture sont présentées en détail, ainsi que les mécanismes de résolution et de routage des paquets des données. De plus, de nouveaux modèles analytiques d'évaluation de performance ont été développés. L'architecture a ensuite été implémentée dans le simulateur de réseau ns-3.26 afin de valider par une approche de simulation, les résultats numériques obtenus. Cet article a été soumis dans la revue *IEEE Transactions on Mobile Computing*.

Le chapitre 7 présente une discussion générale en regard des aspects méthodologiques et des résultats obtenus dans le cadre de cette recherche.

Pour conclure cette thèse, le chapitre 8 présente d'une part, une synthèse de tous les travaux réalisés et discute d'autre part, des limitations ainsi que de nouvelles avenues de recherche qui peuvent être explorées.

CHAPITRE 2 REVUE DE LITTÉRATURE

La gestion de mobilité demeure l'une des composantes indispensables pour offrir aux usagers d'un réseau mobile un accès ubiquitaire aux services. Elle permet de gérer à la fois la localisation des MNs et l'acheminement des paquets qui leur sont destinés de manière convenable. Nous présentons dans ce chapitre les mécanismes de gestion de la mobilité à différents niveaux de la couche TCP/IP, les approches distribuées de gestion de mobilité, les protocoles basés sur le mécanisme ID/Locator Split.

2.1 Gestion de la mobilité au niveau de la couche TCP/IP

2.1.1 Mobilité au niveau de la couche réseau (MIPv6 et ses extensions)

Mobile IP [39] est un protocole définit par l'IETF afin de gérer la mobilité d'un MN à travers l'ensemble des réseaux visités. Dans le but d'identifier le MN, gérer ses mouvements et lui acheminer les paquets, le protocole MIPv4 lui assigne deux adresses différentes : le *Home Address (HoA)* qui représente l'adresse du MN dans son réseau domestique et le *Care-of-address (CoA)* l'adresse du nœud mobile dans chacun des réseaux visités au cours de son déplacement [11]. Fondamentalement, l'architecture du protocole MIPv4 introduit en plus du MN, deux entités fonctionnelles que sont le HA (Home Agent) et le FA (Foreign Agent). MIPv4 supporte la mobilité par trois opérations de base que sont la découverte de l'agent de mobilité, l'enregistrement et le transfert de paquets. Lors de sa présence dans son réseau domestique, le MN reçoit les paquets qui lui sont destinés à travers son HA. En cas de mobilité vers un nouveau réseau, le MN amorce la phase de découverte de l'agent de mobilité en envoyant des messages de découverte au FA. A la suite de cette étape commence la phase d'enregistrement ou le MN découvre sa position dans le nouveau réseau et obtient une adresse dans le réseau visité. Par des messages de mises à jour de liens, le MN met à jour les associations de mobilité auprès de son HA afin d'associer sa nouvelle adresse dans le réseau visité à celle de son réseau domestique [11]. A la dernière étape, tous les paquets émis par les ses nœuds correspondants à destination du MN lui sont acheminés par l'intermédiaire de son HA. Lorsque ce dernier reçoit un paquet à destination d'un MN, il encapsule ces paquets et les transmet au MN en se référant à son CoA dans le réseau visité. Les paquets sont transmis à travers un tunnel établi entre le HA et le FA du réseau visité. Ceux-ci sont interceptés par le FA, qui le décapsule et les achemine ensuite au MN.

Malgré le support de la mobilité du protocole MIPv4, des limitations ont été identifiées dans

son fonctionnement. Le problème du routage triangulaire génère des chemins excessivement longs lors de l'acheminement des paquets entre le MN et ses nœuds correspondants ce qui entraîne une surcharge du réseau et des routeurs. De plus, l'absence de mécanismes de communication entre les différents FA traversés entraîne une perte de paquets significative en plus de la signalisation de mobilité très élevée. Cette signalisation est d'autant plus élevée que le MN est géographiquement éloigné de son réseau domestique. Pour remédier aux limitations de MIPv4, le protocole MIPv6 a été proposé pour un meilleur support de la mobilité.

MIPv6 [28] est un protocole bien connu de gestion de mobilité. Contrairement à MIPv4, le fonctionnement de MIPv6 ne fait pas intervenir un FA au niveau du réseau visité. Un MN est identifié par deux adresses : le Home address (HoA) et le Care-of-address (CoA) [40]. L'adresse HoA est une adresse statique permanente servant à identifier les sessions en cours d'exécution au niveau de la couche application. Grâce à elle, le MN est joignable par ses nœuds correspondants qui peuvent lui transmettre des paquets. Le CoA est une adresse temporaire utilisée à des fins de routage lorsque le MN est attaché à des réseaux visités. La première étape dans un cas de mobilité consiste à établir l'adresse (CoA) en utilisant le préfixe du réseau visité. Le MN informe donc son HA et tout nœud correspondant en envoyant des messages de mise à jour de liens (Binding Updates, BU). Ces messages contiennent les adresses HoA et CoA courants du nœud mobile. Les paquets à destination du MN lui sont donc acheminés par le HA vers le CoA à travers un tunnel établi entre les deux parties. Lorsque le MN retourne à son réseau domestique l'association de mise à jour entre le HoA et le CoA est révoquée. MIPv6 introduit de plus un mécanisme d'optimisation de route afin de résoudre le problème du routage triangulaire en vue d'améliorer les performances du réseau [11]. L'idée est de faciliter le routage entre le MN et ses nœuds correspondants par un échange de requêtes/réponses entre les différentes parties pour l'établissement d'un canal de communication sécurisé entre les deux parties. Malgré l'amélioration de performances observées au niveau de MIPv6, son adoption a été cependant lente, à cause des limitations qui lui sont connues : latence élevée lors des relèves, taux de perte de paquets élevés et des charges additionnelles de signalisation [11]. De plus, la mobilité locale d'un MN est gérée de la même manière qu'une mobilité globale, ce qui entraîne une dégradation notable de performance au niveau du trafic temps réel. Toutes ces limitations ont conduit à l'élaboration de nouveaux protocoles de mobilité dans le but d'améliorer les performances de MIPv6.

FMIPv6 (Fast Handover for Mobile IPv6) [41] a été proposé dans le but de réduire la latence observée lors des relèves basées sur MIPv6 et minimiser ainsi les temps d'interruption de service. L'idée sous-jacente au protocole FMIPv6 est d'anticiper au niveau du MN la relève en informant le prochain routeur d'accès (newAR) de la relève imminente par des triggers de la couche liaison de données (L2). Cette anticipation permet d'initialiser la relève au niveau

L3 alors que le nœud mobile est toujours attaché au point d'accès précédent (oldAR). Ce mode de fonctionnement devrait permettre au MN d'anticiper de manière proactive la relève et réduire le temps nécessaire à la relève L3 après l'achèvement de celle de niveau L2. Au cours de la relève, un tunnel temporaire s'établit entre les deux routeurs pour la redirection des paquets jusqu'à la fin de la procédure de relève et qu'un nouveau canal de communication s'établisse entre le MN et le nouveau point d'accès. Cependant, la précision de la procédure de prédiction dépend essentiellement de la détection au niveau L2 et du temps de détection. Des prédictions erronées risquent de nuire à la mobilité transparente souhaitée.

FMIPv6 permet de générer des déclencheurs en vue d'anticiper sur la procédure de relève au niveau de la couche liaison de données mais celui-ci ne tient pas compte des fausses alertes engendrées dans certains cas. FMIPv6, l'un des protocoles de mobilité les plus représentatifs, fonctionne en mode proactif ou en mode réactif, selon la transmission ou non du message FBBack (Fast Binding Acknowledgement) pendant que le nœud mobile se trouve toujours rattaché au routeur précédent (PAR) [42]. En mode proactif, grâce à l'assistance des déclencheurs (triggers) de niveau L2, la relève au niveau L3 peut être anticipée ou réalisée conjointement avec la relève de niveau L2, ce qui réduit de manière significative sa durée totale. L'initialisation anticipée de la relève au niveau L2 est donc cruciale pour de bonnes performances en mode proactif. Cette condition est cependant difficile à obtenir en situation pratique. L'échec de la relève en mode proactif entraîne l'exécution de celle-ci en mode réactif. L'exécution de la relève en mode réactif occasionne cependant un délai supplémentaire dû au temps de découverte du routeur d'accès en plus du mécanisme DAD (Duplicate Address Detection) qui doit s'exécuter [42].

HMIPv6 (Hierarchical Mobile IPv6) [43] est un protocole de gestion de mobilité locale destiné à réduire la signalisation et la charge engendrée par les déplacements fréquents d'un nœud mobile. Son fonctionnement repose sur l'utilisation de points d'ancrages intermédiaires MAP (Mobility Anchor Point) dans le réseau. Ceci permet de localiser le MN, indépendamment de la position de ses correspondants et son HA. Dans son architecture, les paquets à destination du MN sont tunnelisés vers le point d'ancrage local MAP, identifié par une adresse régionale (RCoA). Le MAP réachemine à son tour par un tunnel, les paquets vers le MN, identifié dans la zone contrôlée par une adresse locale (LCoA). Dans cette procédure, la signalisation de mobilité est donc restreinte à la zone contrôlée par le MAP. Comparativement à MIPv6, les latences de relèves sont moins élevées en plus de la diminution de la charge additionnelle de signalisation dues aux mises à jour.

PMIPv6 (Proxy Mobile IPv6) [29] permet de gérer de la mobilité basé sur le réseau et n'impliquant pas la participation du MN dans la signalisation impliquée. Le principe de

fonctionnement de PMIPv6 repose sur MIPv6 dont il fait une extension en réutilisant la signalisation ainsi que de nombreux concepts tels que le HA [44]. Une fois entré dans un domaine PMIPv6, le réseau de service s'assure de maintenir le MN sur son réseau domestique, en s'assurant que celui-ci puisse obtenir son HoA sur n'importe quel réseau d'accès. Autrement dit, le réseau de service affecte à chaque MN, un préfixe unique dont il se sert sur tous les réseaux d'accès du domaine PMIPv6. Du point de vue du MN, le domaine PMIPv6 apparaît comme son domaine domestique. PMIPv6 introduit deux nouvelles entités : le Mobile Access Gateway (MAG) qui agit comme un nœud de relais entre un MN et un agent de mobilité locale (LMA) dont le rôle est similaire à celui HA dans MIPv6 [45]. Le rôle principal de l'entité MAG est de détecter les mouvements du MN et d'initier la signalisation liée à sa mobilité avec le LMA. En outre, un tunnel s'établit entre le MAG et le LMA afin de permettre au MN d'utiliser une adresse de même préfixe que celui de son réseau domestique en émulant ainsi sa présence sur son réseau domestique malgré sa mobilité sur différents réseaux d'accès. Bien que la signalisation entre le nœud mobile et le réseau soient supprimées, PMIPv6 souffre d'inconvénients similaires à MIPv6. Lors de l'exécution du transfert de lien, il y arrive un moment où le MN est incapable d'envoyer ou de recevoir des paquets en raison de la latence et des pertes de paquets engendrées par le fonctionnement de PMIPv6 [45].

2.1.2 Gestion de la mobilité au niveau des autres couches

Au niveau de la couche liaison de données, le standard MIH (Media Independent Handover) utilise un mécanisme de collecte d'informations aussi bien du réseau, que du client mobile afin d'exécuter de manière coopérative une relève en se basant sur les données reçues de la couche liaison de données (L2).

Un contexte est formellement défini comme toute information pertinente à la situation d'une entité (personne, lieu ou d'un objet) [46]. La sélection du réseau d'accès le plus performant et le plus adapté pour satisfaire les requis des applications de l'utilisateur, est devenue un enjeu important dont l'objectif réel est la maximisation de la qualité de service perçue par celui-ci [47]. Le standard MIH (IEEE 802.21) propose des mécanismes de relèves verticale dans les réseaux hétérogènes, indépendamment du média d'accès utilisé par l'utilisateur, dans le but d'améliorer la qualité des relèves qui sont effectuées et favoriser la continuité des services [48].

Dans le contexte des réseaux hétérogènes, le standard MIH a été utilisé en combinaison avec les protocoles de mobilité standards (MIPv4/MIPv6, FMIPv6, PMIPv6, SCTP, SIP...etc.) mais a aussi servi de noyau d'intégration de plusieurs types de réseaux d'accès (UMTS, Wi-MAX, Wi-Fi...). Les différentes contributions de la littérature se basent sur les différents modules du standard pour la gestion des interfaces multiples, décider de l'instant d'initiali-

sation de la relève ainsi que de la sélection du réseau de destination [49].

Dans [50], un algorithme adaptatif de relève verticale est présenté. L'étude considère un environnement hétérogène composé des réseaux d'accès UMTS, WiMAX et Wi-Fi. Dans cette proposition, les requis de qualité de service utilisateur sont pris en compte aussi bien lors de la phase de préparation de la relève que lors de la phase de sélection du réseau de destination. Lors de la phase de préparation, les triggers d'initialisation de relève sont non seulement générés au niveau de la couche liaison de données (MIH Link Going Down Event) mais aussi au niveau de la couche application. Les triggers de la couche application sont générés lorsque les conditions de qualité de service sur le réseau ne satisfont plus aux besoins des applications en cours d'exécution ou lors d'un changement des préférences des paramètres utilisateurs. De plus, un algorithme de sélection MADM (Multiple Attribute Decision Making) permet de choisir lors de la relève verticale, le réseau présentant les meilleures caractéristiques de qualité de service. L'algorithme proposé présente de meilleures performances de QoS comparativement à une sélection uniquement basée sur la qualité du signal perçu RSS pour un trafic FTP et VoIP. On y observe également une meilleure utilisation des ressources. Cependant aucune précision n'est donnée sur la manière dont les informations sont mises à jour de manière dynamique au niveau du serveur d'informations de MIH.

Dans [51], le module EIS (Enhanced Information Server) à MIH afin d'accélérer la prise des mesures nécessaires aux relèves verticales. Le module EIS recueille des informations plus précises quant à la position du nœud, la durée, le niveau de signal du mobile et utilise ces informations de manière coopérative avec le réseau pour la prise de décision de relève. Comparativement aux mécanismes traditionnels de relève verticale, Le module proposé permet d'obtenir de meilleures performances, néanmoins comme toute approche coopérative, l'enjeu principal demeure la collecte d'avance des informations de relève qui peuvent générer une signalisation additionnelle susceptible d'entraîner une dégradation des performances du mécanisme proposé.

Dans [52], l'algorithme de décision proposé est basé sur MIH et permet d'obtenir des métriques de décision relatives aux points d'accès voisins. Une évaluation basée sur la méthode AHP, permet de déterminer les différents facteurs de pondération associés aux différentes métriques considérées (bande passante, taux d'erreurs de paquets, délai de bout en bout) en tenant compte des requis fonctionnels des applications en terme de QoS. Une fonction de coût des différents paramètres considérés permet la sélection du réseau d'accès le plus adapté en fonction des valeurs obtenues pour les différentes valeurs pondérées des métriques et des réseaux d'accès. L'inconvénient à cette méthode est qu'elle ne spécifie pas les différentes primitives MIH utilisées et ne précise pas comment sont collectées et ensuite gérées les

informations dynamiques nécessaires à la prise de décision.

Dans [53], l'infrastructure eMIHF (enhanced Media Independent Handover Framework) est proposée afin d'améliorer la version standard de MIH. Dans son implémentation, eMIH propose des mécanismes de provision et d'activation des ressources de qualité de service dans le réseau de destination, afin d'assurer une mobilité transparente dans le contexte des réseaux d'accès hétérogènes. Mieux, la solution proposée réalise la translation et l'adaptation des paramètres de qualité de service du réseau d'accès courant vers le réseau de destination, permettant ainsi la mise en œuvre des politiques de QoS dans le réseau cible au cours de la phase de préparation de la relève. eMIH assure une interaction entre les différents mécanismes de mobilité proposés au niveau des technologies d'accès en place et permet d'après les auteurs, de garantir un mécanisme fiable de relève de type make-before-break.

En se basant aussi sur un serveur d'information dénommé CIS (Context-aware Information Server) dans [53], une approche sensible au contexte pour l'initialisation ainsi que les procédures de décision des relèves est proposée. Ce serveur collecte, gère et distribue les informations en temps réel dans le réseau. Ces informations dynamiques sont retirées aussi bien du réseau que du terminal (capacité du terminal mobile, les services, préférences de l'utilisateur et du réseau. Etc). L'intégration des technologies Wi-Fi, WiMAX et UMTS est considérée dans l'infrastructure MIH. Le mécanisme proposé permet de réduire le temps d'initialisation et d'exécution des relèves, cependant la distribution des informations entre les différentes entités révèle une limitation dans l'approche proposée.

Dans [54], les mécanismes d'amélioration de performances de MIH proposés visent à inclure des informations contextuelles des couches supérieures afin d'améliorer la pertinence des décisions de relève. Ces informations des couches supérieures sont relatives aux applications en cours d'exécution et au contexte utilisateur. Un mécanisme de collecte d'informations nécessaires à la relève permet d'interagir avec l'utilisateur mobile et le réseau lors des prises de décision. Néanmoins une limitation de ce travail résulte du fait que le contexte utilisateur considéré n'est pas dynamique. De plus la solution proposée ne tient pas compte de l'existence de plusieurs types de trafic et de classes de QoS dans le réseau.

Dans [55], Lin et al proposent une approche prédictive transversale de relève verticale est présentée en utilisant le protocole MIH pour le support de la QoS et pour une meilleure connectivité dans les réseaux hétérogènes. Cet algorithme fait un mapping des paramètres de QoS entre le réseau UMTS et le réseau WiMAX. Les paramètres de décision de relève considérés sont la vitesse, la distance, le coût, la puissance, le taux de perte de paquets. L'approche proposée permet de réduire le niveau de perte de paquets grâce à une prédiction basée sur un contrôle constant du niveau de signal. Le débit global du mécanisme proposé

est aussi consistant en faveur des applications sensibles à la qualité de service.

Dans [56], les auteurs proposent un algorithme de relève verticale en faisant un équilibrage de charge entre les différents points d'accès dans l'objectif de minimiser la consommation énergétique sur l'ensemble des nœuds mobiles. L'algorithme proposé tend à optimiser par une fonction de coût, la consommation énergétique ainsi que la charge entre les différents points d'accès. Selon les auteurs, la solution proposée offrirait aux opérateurs la flexibilité de maximiser la durée d'utilisation des équipements mobiles ainsi qu'une répartition adéquate du trafic sur plusieurs points d'accès dans le réseau. Néanmoins, cette solution se limite uniquement à un seul domaine administratif [57]. Le standard MIH a été utilisé pour améliorer les performances des algorithmes et des mécanismes de relève afin d'améliorer la qualité de service et faciliter la continuité de services. Les différents services présents au niveau du standard facilitent aussi la gestion des interfaces multiples ainsi que l'implémentation des relèves de type L2 et la mise en œuvre de mécanismes adaptatifs pour satisfaire les besoins des utilisateurs [49].

En dehors des mécanismes de mobilité basés sur la couche réseau, une approche importante qui a suscité beaucoup d'intérêt au cours des dernières années est la gestion de la mobilité au niveau de la couche de transport [58]. Les mécanismes de gestion de mobilité basés sur la couche transport présentent des avantages importants tels que : le support inhérent de l'optimisation de route (évite le problème de routage triangulaire), le support de la domiciliation multiple, l'indépendance du concept de réseau domestique, la possibilité de relèves transparente des nœuds multi-interfaces. De plus, contrairement aux approches basées sur Mobile IP, ces protocoles se basent sur une mobilité de bout-en-bout ce qui permet de réduire les modifications au niveau de l'infrastructure du réseau, diminuant ainsi sa charge en confinant la gestion de mobilité au niveau des équipements terminaux. Dans cette optique, le protocole SCTP [59] est un protocole de la couche transport qui définit la notion d'association pour représenter un contexte d'échange de flux unidirectionnels entre deux nœuds. Par cette approche, ce protocole offre un support natif de la domiciliation multiple et plusieurs adresses IP peuvent être associées à un contexte d'association donné au niveau de chaque nœud lors de la communication. De ce fait, les nœuds à plusieurs interfaces peuvent alors communiquer à travers plusieurs chemins de bout-en-bout. Ceci augmente la résilience du réseau en cas de pannes grâce à l'existence de plusieurs chemins. Le mécanisme de reconfiguration dynamique d'adresse (DAR, Dynamic Address Reconfiguration) associé au protocole SCTP a donné naissance au protocole mSCTP [59]. Ce dernier est une extension de SCTP conçu dans le but de résoudre les défis inhérents aux relèves verticales des réseaux hétérogènes en activant de manière dynamique les différentes interfaces réseaux pour permettre le support de la mobilité au niveau de la couche transport. Le mécanisme DAR permet d'ajouter, de modifier et de

supprimer de manière dynamique des adresses associées au contexte de communication du protocole SCTP.

Le support de la mobilité au niveau de la couche application a également reçu beaucoup d'attention. Les travaux relatifs à ces mécanismes se basent sur l'idée qu'une solution de mobilité indépendante des technologies d'accès, et du réseaux sous-jacent est envisageable pour gérer la mobilité [60]. Dans ce contexte, le protocole SIP (Session Initiation Protocol) [61], initialement conçu pour gérer les sessions multimédia au sein des réseaux IP [11], a été utilisé dans plusieurs mécanismes de mobilité. Dans son fonctionnement, le protocole SIP ne nécessite pas l'établissement d'un tunnel à travers le HA du MN, ce qui s'apparente au mécanisme d'optimisation de route dans le fonctionnement du protocole MIPv6 en vue de réduire les délais d'acheminement de bout-en-bout des paquets. Dans son architecture, le protocole SIP intègre des entités de gestion de mobilité telle que le UA (User Agent), PS (Proxy Server), et les RS (Redirect Server). Ainsi lors du déplacement d'un MN durant une session active, vers un nouveau réseau, ce dernier reçoit une nouvelle adresse réseau et adresse un message d'invitation à ses nœuds correspondants afin d'établir une nouvelle session. Les paquets subséquents sont donc acheminés au MN en utilisant cette nouvelle adresse. Le protocole SIP ne garantit cependant pas le maintien des sessions TCP et les associations de port UDP lors du déplacement des MN [61].

En réutilisant les principes du protocole HMIPv6, P_HIMIPv6 [62] utilise un partenaire d'assistance dans un domaine MAP afin d'effectuer une relève au sein des réseaux véhiculaires. Cette sélection est basée sur la proximité qui est mesurée en utilisant le RSSI pour décider si le partenaire est idéal à la relève. L'approche de gestion de la relève dans P_HIMIPv6 est similaire à celle de FMIPv6, dans la mesure où la mise à jour de l'information de routage est effectuée avant que ne s'opère la rupture du lien physique. L'avantage du P_HIMIPv6 tient aussi du fait qu'il ne nécessite pas de mémoire tampon (buffer). Considérant cependant que l'usage du RSSI uniquement pour la sélection de partenaire de relève n'est pas adéquate aux réseaux véhiculaires, une métrique d'expiration du lien entre véhicules VLET est proposée dans [63] afin de sélectionner le meilleur partenaire de relève dans le réseau véhiculaire. Le protocole proposé dans [63] utilise une topologie de route statique, considère la trajectoire du MN ainsi que les probabilités de choix de routes dans la sélection du véhicule partenaire, afin de maintenir le lien le plus longtemps possible. Le protocole proposé a révélé de significatifs gains de performance.

Dans le but de faciliter des relèves transparentes, le protocole e-SMIPv6 est proposé dans [64]. Ce protocole se base sur l'établissement de tunnels bidirectionnels entre les routeurs d'accès avant l'exécution effective des relèves. Par conséquent, les MNs utilisent l'adresse

précédemment configurée sur le AP précédent la relève. Dans le but de réduire le délai de relève associé à la détection d'adresse, chaque routeur d'accès conserve un pool d'adresses ne pouvant être assignés qu'une seule fois. Afin de minimiser les pertes de paquets, le routeur d'accès précédent procède à une double transmission (bi-casting) autant avec le MN qu'avec le routeur d'accès de destination. Le protocole e-SMIPv6 minimise autant que possible la signalisation de la mobilité pendant la relève, ce qui représente une solution idéale aux MNs dont la vitesse est élevée.

Dans [65], un protocole de multidiffusion transparent est proposé en utilisant MIPv6. Ce protocole vise à réduire la perte de paquets due aux interruptions de service lorsque le MN rejoint un groupe de multidiffusion. L'approche proposée introduit un nouveau nœud dénommé MBA dont le rôle est de mettre en mémoire tampon les paquets lors de l'exécution des relèves.

Une architecture d'intégration est proposée dans [66]. Cette architecture basée sur le projet 3GPP2, permet l'intégration et l'interfonctionnement des systèmes sans fil 4G et examine les problèmes de gestion de la mobilité. Un protocole de gestion de relève, basée sur une approche locale pour la gestion de la mobilité, la découverte des réseaux d'accès et de relève rapide appelé HPIN (Handoff Protocol for Integrated Networks) [67] est proposé. Ce protocole réduit les interruptions de service lors des relèves au sein des environnements sans fil hétérogènes basés sur IPv6.

Le protocole HTM [68] (Hierarchical Transport layer Mobility protocol) est proposé pour la gestion de la mobilité au niveau de la couche transport. Il vise à réduire les délais de significatifs de signalisation du protocole SCTP et gère la mobilité locale et globale dans le but d'assurer des relèves plus efficaces. HTM exploite les fonctionnalités de reconfiguration dynamiques d'adresse du protocole SCTP et introduit un point d'ancrage de mobilité AMU (Anchor Mobility Unit) afin de mener à bien les procédures de relève plus efficaces.

Le protocole SMIPv6 [69] (Seamless Mobile IPv6) vise à réduire le coût de signalisation ainsi que les latences de relèves au sein des réseaux basés sur MIPv6. Il exploite les profils des utilisateurs en vue de prédire la localisation de ceux-ci lors des relèves. En se basant sur l'information de prédiction, le protocole met à jour tous ses CNs avec sa nouvelle adresse avant de quitter son réseau actuel. En outre, en se basant sur les informations de la couche L2, SMIPv6 est capable de prédire avec précision le moment où s'effectue le transfert de lien sans fil.

Dans le but de faciliter l'interopérabilité entre les systèmes hétérogènes de prochaine génération, un composant spécial du réseau dénommé WING (Wireless INterworking Gateway) [70] est proposé en vue de faciliter les relèves globales. Ce composant est déployé entre pairs de

réseaux sans fil voisins afin de réduire la charge de signalisation engendrée au niveau des bases de données des registres d'utilisateurs.

2.1.3 Améliorations basées sur MIPv6 et ses extensions

Plusieurs études ont révélé que de manière générale, les protocoles FMIPv6 et PMIPv6 améliorent de façon significative les performances du réseau, comparativement au protocole MIPv6. Cependant, les protocoles dont le fonctionnement met en oeuvre des relèves anticipées (fast handover) connaissent des performances quelque peu amoindries lorsqu'ils sont utilisés dans un contexte à fort taux de mobilité avec utilisation des applications temps réel [22]. En se basant sur les récentes avancées technologiques liées à la virtualisation, le protocole FMIPv6 et les améliorations qui lui ont été apportées peuvent potentiellement résoudre plusieurs problèmes de gestion de la mobilité dans les réseaux 5G [22]. Ces améliorations visent plusieurs aspects soit en améliorant le fonctionnement du protocole, en ajoutant de nouvelles fonctionnalités ou en procédant à des modifications en vue de renforcer son niveau de sécurité.

Les améliorations dites fondamentales, se rapportent à la version native du protocole FMIPv6 et visent à améliorer les sous-étapes ou phases de son processus général de fonctionnement. Dans [71], la phase d'anticipation de la relève a été améliorée par l'introduction d'un module d'apprentissage permettant d'enregistrer l'historique des positions occupées (trajectoire du mouvement) par le MN. Ce module permet de prédire la localisation du MN lors des relèves subséquentes. De façon similaire, dans [72], un protocole basé sur le système GPS (Global Positioning System) permet à chaque MN de transmettre périodiquement sa position à une unité de contrôle de mobilité (mobility controller, MC). Selon ce protocole, l'unité de contrôle permet d'informer le MN de sa prochaine position lors de ses déplacements. Le protocole proposé dans [73] utilise la radio-fréquence (RFID) comme mécanisme de détection du mouvement du MN. Cette approche favorise une utilisation judicieuse de la bande passante du réseau, en plus de réduire les délais additionnels de signalisation engendrés par les messages de sollicitation nécessaires au fonctionnement de FMIPv6.

Le mécanisme d'anticipation de relève permet de réduire considérablement le temps d'exécution total. Dans le but d'anticiper au niveau L2 l'exécution des déclencheurs, un protocole de prédiction du niveau du signal (RSS) est proposé dans [74]. Dans [75], un mécanisme adaptatif du seuil de déclenchement des relèves en fonction du niveau du signal (RSS) est proposé. L'ajustement se fait en considérant la vitesse du MN en vue du déclenchement de la relève. Une approche similaire est proposée dans [76]. Cependant, le gain de l'antenne de transmission ainsi que le rayon de couverture de la cellule sont considérées en plus de la vitesse du MN. Cette proposition évalue de manière anticipée, le niveau de signal

attendu à la périphérie de la cellule afin d’amorcer la procédure de relève. Suivant cette même philosophie, [77] propose que les déclencheurs de relève au niveau L2 soient remplacés par des algorithmes de prédiction basés sur des algorithmes de forage de données, en utilisant l’historique des mouvements du MN. En exploitant cette information, l’algorithme prédit la prochaine localisation (position) du MN en vue d’anticiper la relève.

Considérant que l’établissement et le maintien des tunnels de communication entre le PAR et le NAR lors des relèves constitue une source potentielle de la dégradation de performance (perte de paquets, surcharge de signalisation), la solution proposée dans [78] introduit un mécanisme anticipé d’établissement du tunnel de communication entre les routeurs d’accès du réseau. Ultimement, cette approche permet de négocier d’avance les caractéristiques des tunnels de redirection de trafic tels que le niveau de sécurité, les méthodes de cryptage des données afin d’améliorer les performances globales du protocole. Une optimisation de route proactive est proposée dans [79]. Cette approche suggère qu’en plus de produire le nCoA, le MN produise également générer deux jetons t1 et t2. Ces deux jetons sont transmis au CN à travers deux chemins distincts en transitant (pour t1) par le HA et par NAR (pour t2). Le CN vérifie pour les deux jetons reçus la nouvelle adresse configurée par le MN, le nCoA. Des lors, le CN accuse réception de la mise à jour de localisation, en transmettant un accusé de réception directement au NAR.

Dans la version native de FMIPv6, le PAR cesse de transmettre des paquets au MN aussitôt qu’il reçoit le paquet de type FBU (Fast Binding Update). Tout le trafic destiné au MN est dès lors, placé dans un tampon. Cependant, dans la pratique, le MN pourrait être capable de maintenir un lien avec le PAR pendant une période significative même après avoir envoyé ledit message FBU [22]. Cela peut donc engendrer une latence de relève significative. Afin de remédier à ce problème, un mécanisme de bi-diffusion (bicasting) est introduit dans [80]. Suivant ce mécanisme, le PAR conserve dans sa mémoire tampon non seulement le trafic entrant destiné au MN, mais envoie également une copie de chaque paquet à ce dernier sur son lien sans fil, même après réception du message FBU [22]. Lorsque le tunnel de communication entre les routeurs PAR et NAR est déjà établi, les paquets peuvent également être retransmis vers NAR.

Certaines modifications ont été apportées au protocole FMIPv6 afin de lui rajouter des nouvelles fonctionnalités. Le protocole FMIPv6 peut faciliter le transfert du contexte de QoS du MN vers le NAR avant que le MN ne s’y attache. Ce contexte peut être transmis par le biais des messages HI/HACK transmis par le PAR. Par conséquent, le NAR ayant reçu les informations du contexte de QoS du MN en avance sera donc capable d’allouer les ressources nécessaires suite aux relèves [22]. À cet effet, un mécanisme d’allocation dynamique

de ressources est introduit dans [81] et [82] en vue de procéder à une gestion adaptative des ressources du réseau pour améliorer son utilisation. Un mécanisme similaire [83] repose sur une entité centrale du réseau appelé Domain Resource Manager (DRM), chargée de la gestion des ressources dans un domaine administratif particulier. Le MN au moment de la relève doit coordonner l'allocation de ressource avec le DRM. Contrairement à l'approche de réservation de ressource au niveau du NAR, le support de la qualité de service de bout-en-bout dans le réseau a également aussi reçu de l'attention. Le protocole CXTP (Context Transfer Protocol) de l'IETF [84] permet de transférer le contexte de mobilité du MN du PAR vers le NAR durant une relève. Afin de supporter la QoS de bout-en-bout, le protocole CXTP a été optimisé dans [85]. Ainsi, le message CTD (context transfert data) a été étendu par l'ajout d'une en-tête de transfert de données afin de procéder à une réservation de ressource tout le long du chemin du PAR vers le NAR, en passant par les nœuds intermédiaires.

Outre les modifications apportées au protocole FMIPv6, plusieurs contributions se sont consacrées à l'amélioration des performances du protocole PMIPv6. Dans [86], une solution de relève localisée est proposée dans le but de résoudre, dans l'architecture de PMIPv6, le problème de goulot d'étranglement qui se pose au niveau du LMA. La solution proposée tient également compte d'un mécanisme d'optimisation de route en faisant usage à la fois de mécanismes de relève rapide ainsi que d'une architecture hiérarchique. L'architecture proposée gère ainsi les relèves sans l'intervention du LMA. Cependant, ce mode de fonctionnement se fait au détriment des MAGs qui se retrouvent ainsi surchargés par la gestion des communications ainsi que le trafic de signalisation. La solution proposée dans [87] est basée sur les grappes (clusters). Le domaine local est divisé en sous-domaines qui constituent les grappes. Chaque grappe comprend un certain nombre de MAGs, avec un MAG élu en tant que tête de grappe. Ce partitionnement réduit la charge sur le LMA, en introduisant des relèves intra-cluster au niveau local qui permettent d'optimiser le chemin de communication en plus de réduire le taux de perte de paquets. Dans le but de réduire la durée des relèves ainsi que le taux de perte de paquets, un protocole combinant des relèves rapides et locales dans le contexte de PMIPv6 appelé FLPMIPv6 (Fast and Local PMIPv6) est proposé dans [88]. Cette solution nécessite l'intervention du MN en faisant usage des protocoles FMIPv6 et MIH. Lors d'une relève, le MN utilise MIH afin d'informer son MAG précédent des différents points d'attaches candidats. Ce dernier procède ensuite à la mise en œuvre de la signalisation de relève avec le MAG de destination. Néanmoins, cette approche nécessite de procéder à des modifications dans la pile de protocoles du MN. En réutilisant aussi le protocole MIH, la solution proposée dans [89] procède à une bi-diffusion des paquets entre les MAGs impliqués dans la relève. Selon cette approche, la relève est initiée par le MN qui signale au MAG précédent son intention de relève. Ce dernier initialise la signalisation de relève avec le MAG de destination, qui

à son tour procède à l'enregistrement du MN auprès du LMA. Le LMA après cette opération procède à la bi-diffusion des paquets aux deux MAGs. Cette approche a le mérite de réduire le taux de perte de paquets ainsi que les délais de relèvement, cependant, il induit une forte charge de signalisation due à la bi-diffusion des paquets. En combinant les avantages des deux protocoles FHMIPv6 (Fast Hierarchical MIPv6) et FPMIPv6 (Fast Proxy MIPv6), le protocole PFHMIPv6 (Proxy-based Fast Handover for Hierarchical Mobile Ipv6) [90] réduit à la fois le coût de la signalisation ainsi que le taux de perte de paquets lors des relèves. Cependant, PFHMIPv6 a hérité du problème des prédictions de relèves erronées du protocole FPMIPv6. Les travaux présentés dans [22] et [91] décrivent de manière détaillées, plusieurs modifications apportées au protocole MIPv6 et ses extensions.

2.2 Mécanismes distribués de la gestion de la mobilité

La gestion distribuée de la mobilité (DMM) offre une approche alternative aux mécanismes de gestion centralisée de la mobilité (CMM) au sein des réseaux IP actuels [31]. L'explosion du volume de trafic de données mobiles, le nombre d'utilisateurs mobiles en forte croissance, ainsi que les exigences de performance de plus en plus strictes, forcent l'adoption de nouvelles architectures de moins en moins centralisées. Il est nécessaire de définir de nouveaux mécanismes de gestion de la mobilité, distribués et activés de manière dynamique [92, 93]. À terme, la DMM envisage de répondre aux limitations des approches centralisées actuelles en vue de faire face aux exigences des réseaux de prochaine génération.

D'une part, l'approche de la DMM repose sur l'existence de points d'ancrage de mobilité distribués au niveau du réseau d'accès. Plus précisément, ces points d'ancrage du trafic sont associés aux premiers routeurs du réseau d'accès, de manière à maintenir une proximité avec les MNs. Le trafic associé au MN est naturellement distribué et ensuite acheminé de manière optimale. En outre, une défaillance d'un point d'ancrage de mobilité est restreinte aux MNs desservis par ce dernier [94]. D'autre part, la DMM s'appuie aussi sur la mise à jour dynamique des points d'ancrage selon l'emplacement du MN. L'acheminement des données est assuré via le maintien d'un tunnel entre l'ancre de mobilité et le MN. Les fonctionnalités de mobilité sont également distribuées au niveau du réseau d'accès et activées de manière dynamique [94]. Dans [31], une analyse de plusieurs approches de la gestion de mobilité distribuée est présentée.

Lorsque la mobilité distribuée est contrôlée par le MN, ce dernier est impliqué dans la signalisation de mobilité. La gestion distribuée de la mobilité sous le contrôle du MN est basée sur le protocole MIPv6 et tente d'améliorer les performances du support de la mobilité. Dans [94], une nouvelle fonction, appelée Access Mobility Anchor (AMA), est introduite en faisant une

extension des fonctionnalités du HA [94]. Ces AMA sont distribuées à la périphérie du réseau d'accès et permettent au MN de configurer son adresse en fonction du préfixe de réseau fourni par l'AMA. Lorsqu'un MN se déplace vers un réseau d'accès voisin, desservi par une autre entité AMA, une nouvelle adresse est configurée dans le réseau visité. L'adresse précédente configurée sur le AMA initial est conservée [33]. En conséquence, une signalisation s'établit entre l'AMA en service et l'AMA d'origine, tandis qu'un tunnel bidirectionnel est créé entre ces deux entités du réseau. Cette solution crée plusieurs tunnels entre les entités AMAs du réseau et lorsque la fréquence de mobilité est élevée, la performance du système peut être compromise par les mises à jour fréquentes et la forte signalisation nécessaire à maintenance de plusieurs tunnels [33]. Une approche DMM contrôlée par le MN est décrite dans [95], avec un mode de déploiement selon lequel les fonctionnalités de HA et de routeur d'accès, introduits dans MIPv6, sont colocalisés et distribués. Le MN hérite des fonctionnalités d'un client MIPv6, mais il doit en outre être en mesure de gérer plusieurs adresses IPs pour une utilisation simultanée, et initialiser plusieurs procédures de communication avec le réseau.

Lorsque la gestion de la mobilité distribuée est contrôlée par le réseau, une entité du réseau jouant à la fois le rôle de routeur d'accès et de point d'ancrage de mobilité, doit détecter la présence d'un MN puis exécuter la procédure de mise à jour de localisation [31,96]. Les points d'ancrage de mobilité distribuée effectuent la signalisation de mobilité nom du MN, comme c'est le cas par exemple dans PMIPv6 [26]. Les fonctionnalités de gestion de la mobilité sont aussi implantées au niveau des routeurs d'accès afin d'ancrer le trafic plus près du MN. Chaque AR doit posséder à la fois des fonctionnalités d'ancrage et de localisation de la mobilité et est également appelé Mobility Capable Access Router (MAR). Lorsque survient une relève, le trafic de données est tunnelisé entre le MAR courant et le MAR d'ancrage de la session. Dans cette configuration, l'architecture du réseau est partiellement distribuée et repose sur une base de données centralisée qui stocke les données de sessions en cours des MNs. Ainsi, suite à la relève, le nouveau MAR récupère les adresses IP du ou des MARs précédents à partir de ladite base de données. Par la suite, le nouveau MAR transmet un message de mise à jour de localisation du MN à chaque MAR d'ancrage. Dans [97], une approche basée sur le protocole PMIPv6 est proposée. Dans cette proposition, le router périphérique appelé Mobility Anchor and Access Router (MAAR), joue le rôle de gestionnaire de mobilité aux préfixes IPv6 à qui il sert à la fois de point d'ancrage et offre les fonctionnalités du protocole PMIPv6 (MAG et LMA). Les MAARs détectent les mouvements du MN, émulent son réseau domestique en masquant sa mobilité en maintenant le même préfixe qui lui a été attribué sur son MAAR initial. Une évaluation analytique de cette approche est présentée dans [98]. L'article présenté dans [99] discute de l'adoption de la mobilité distribuée pour la gestion de la mobilité au sein des réseaux 5G. Un protocole basé sur le paradigme du SDN (Software-

Defined Networking) est proposé en adoptant une approche partiellement distribuée de la mobilité. Le SDN est une architecture des réseaux dans laquelle, le plan de contrôle est séparé du plan de transport des données. Dans cette architecture, un plan central (contrôleur) définit les règles de routage des paquets entre une source et une destination. Ces règles sont ensuite transférées aux commutateurs du plan de données pour assurer la livraison des données. Le protocole proposé introduit un contrôleur réseau (Network Controller, NC) ainsi que plusieurs commutateurs dans le plan de données (DMM-GWs). Les DMM-GWs jouent le rôle de points d'ancrage du trafic des MNs, et reçoivent les règles de transfert transmises par le NC. Lors de son attache, le MN reçoit un préfixe unique qui lui est assigné tant qu'il demeure dans le réseau. Le NC détient une base de données locale permettant de localiser le DMM-GW courant du MN. La gestion de la mobilité est assurée par des règles de réécriture d'adresse et de redirection du trafic entre le DMM-GW servant de point d'ancrage à la session courante, et le DMM-GW actuel auquel le MN est attaché. Cette solution, contrairement aux approches de gestion de la mobilité basés sur PMIPv6, n'introduit pas de tunnels de communication additionnels. Une approche similaire à la précédente est proposée dans [100] dans le contexte d'une architecture de réseau LTE. Le contrôleur SDN est colocalisé avec l'entité MME (Mobility Management Entity) du réseau LTE. Au sein de cette architecture, les autres composantes fonctionnelles du réseau LTE à savoir le HSS (Home Subscriber System), le S-GW (Serving Gateway), et le P-GW (Packet Data Network Gateway) sont implémentées comme des applications déployées au niveau du contrôleur SDN. Le plan de données est assuré dans cette architecture par les routeurs d'accès du réseau DMM traditionnels. La mobilité des MNs est gérée selon deux approches. Avec la première approche, un tunnel de communication est établi entre le commutateur servant de point d'ancrage au trafic du MN et le commutateur courant auquel le MN est attaché. Les tables de flux des deux commutateurs étant mises à jour par le contrôleur central, cette approche est similaire à celle présentée dans [99], à la différence qu'un tunnel est maintenu dans le plan des commutateurs. Selon la deuxième approche, lorsqu'une relève survient, le contrôleur procède directement à une optimisation de route sur le commutateur distant du CN afin de rediriger le trafic du MN vers sa nouvelle position. Cette seconde approche ne nécessite pas l'établissement d'un tunnel.

2.3 Gestion de la mobilité par l'approche du Software-defined Networking (SDN)

Le paradigme du SDN (Software-Defined Networking) est basé sur une séparation de l'architecture du réseau au sein de laquelle, les fonctions d'acheminement des paquets de données sont découplées des fonctions de contrôle du réseau (signalisation) [101]. L'intelligence du réseau est alors transférée vers un contrôleur du réseau, logiquement centralisé, qui en dé-

tient une vue globale. Ce contrôleur interagit avec de simples dispositifs de commutation et fournit une interface de programmation aux applications de gestion de réseau définies par ses utilisateurs. Cette approche permet de découpler le système qui décide de la destination du trafic (le plan de contrôle), du système sous-jacent qui en assure l'acheminement vers la destination sélectionnée (le plan de données) [102].

L'architecture SDN est une nouvelle approche de conception du réseau visant à réduire la nature complexe et statique des architectures traditionnelles du réseau qui sont distribuées [101]. En se basant sur le SDN, les administrateurs du réseau peuvent donc programmer le comportement du trafic et du réseau de manière centralisée, sans nécessiter un accès et une configuration indépendantes de chacun des périphériques matériels du réseau [102]. Le plan de contrôle étant séparé de l'infrastructure physique sous-jacente, il abrite toutes les fonctions destinées au contrôle du réseau. La migration des fonctions de contrôle vers un plan de contrôle programmable, permet aux administrateurs de modifier son comportement réseau en temps réel, et de l'adapter à l'évolution rapide des conditions du réseau en exécutant les applications appropriées du contrôleur [101].

L'architecture SDN a été intégré à plusieurs protocoles, architectures et mécanismes de gestion de la mobilité des réseaux 5G. Le travail présenté dans [103] propose une architecture de gestion des relèves au sein des réseaux ultra-dense 5G en utilisant le SDN. L'architecture proposée vise essentiellement à réduire le délai des relèves. Suivant cette approche, le contrôleur du réseau intègre deux modules : le module de gestion de la mobilité ainsi que le module de contrôle d'admission. Considérant une configuration hexagonale des cellules, le module associé à la gestion de la mobilité permet d'estimer, pour chaque eNB, les probabilités de transitions au sein du réseau en se basant sur une chaîne de Markov. Ce module intègre aussi le moteur de sélection des cellules lors d'une relève. Le contrôleur du réseau intègre aussi des procédures de contrôle d'admission lors des relèves. L'architecture proposée permet de réduire la durée des relèves ainsi que la probabilité d'échec de celles-ci en comparaison à l'architecture traditionnelle du réseau LTE.

Dans [104], l'architecture proposée considère un réseau LTE partiellement virtualisé basé à la fois sur les concepts du SDN et du NFV (Network Function Virtualisation) et propose un mécanisme de gestion de relève utilisant l'interface de relève X2 du réseau LTE. Cette architecture se base sur une configuration hiérarchique à trois niveaux tous intégrés au Cloud. Il s'agit du niveau d'accès, du niveau régional distribué ainsi que du niveau national centralisé. Cette proposition visait essentiellement à évaluer l'impact du réseau SDN sur les mécanismes de gestion de relèves au sein des réseaux 5G, notamment, le temps d'exécution des relèves au sein de ladite configuration. Les résultats obtenus, selon les auteurs, ont montré que

l'architecture proposée satisfait aux critères de performance exigés des réseaux 5G.

Considérant les limites de l'architecture du réseau LTE actuel dues à la centralisation de ses fonctions (monitoring, contrôle d'accès, filtrage et transfert de données..etc), il est proposé dans [105] une architecture de gestion de la mobilité réutilisant les concepts du SDN, du NFV et du Cloud Computing. Cette architecture vise à répartir les composantes P-DWs du réseau LTE traditionnel en les rapprochant des usagers, donc de sa périphérie. Elle vise aussi à séparer les plans de données et de signalisation du réseau, et à offrir un plan de signalisation virtuel et centralisé. Les déplacements des MNs sont gérés au sein de cette architecture en adoptant les concepts de la mobilité partiellement distribuée DMM. A cet effet, les P-GWs du réseau LTE désormais distribués et proches de la périphérie du réseau (P-EGW, PDN Edge Gateway) visent à aplatir le réseau et répartissant ses points d'ancrage de mobilité à sa périphérie. L'ensemble des P-EGWs du réseau forment le plan de données constitués de commutateurs SDN, tandis que le plan de contrôle piloté par un contrôleur SDN gère les procédures de signalisation et d'allocation d'adresses IP. Les fonctions classiques du réseau LTE (MME, HSS, PCRF) sont centralisées et implémentées au niveau du plan contrôle sous forme de fonctionnalités virtualisées du réseau grâce au Cloud Computing. L'évaluation des performances de l'architecture proposée montre qu'elle apporte plus de flexibilité, accroît son évolutivité et réduit les coûts du réseau en comparativement à l'architecture LTE du réseau 4G.

Dans le but d'améliorer la gestion du réseau et de la mobilité au sein des réseaux sans fil SDN, l'architecture proposée dans [106] est basée sur le protocole LISP et vise à supporter plusieurs services tels que le routage évolutif, des relèves transparentes, des mécanismes de gestion du trafic ainsi qu'une gestion centralisée du réseau. L'architecture proposée intègre le protocole OpenFlow au protocole LISP dans le but de réduire le coût de signalisation. De plus, elle offre la possibilité de procéder à un offloading des flux de trafic, en utilisant les identificateurs associés aux MNs, dans le but de faire une utilisation rationnelle la bande passante du réseau. L'évaluation de performance de cette architecture révèle une réduction du coût de signalisation associé à la gestion de localisation.

Combinant la philosophie du SDN et celle de la gestion de mobilité distribuée, la solution proposée dans [107] présente une architecture au sein de laquelle, la mobilité est un service du réseau déployé en tant qu'application logicielle au niveau du contrôleur SDN. Cette approche vise à tirer pleinement partie des bénéfices du SDN. La solution proposée présente plusieurs caractéristiques à savoir : un plan de données distribué, un plan de contrôle centralisé, un support de la mobilité basée sur les flux de trafic de données, le découplage des fonctionnalités du réseau des technologies d'accès sous-jacentes au niveau de l'infrastructure.

Les résultats expérimentaux de l'architecture ont révélé des performances similaires aux architectures distribuées en terme de la latence des relèves, du délai de bout-en-bout, en plus d'apporter une grande flexibilité dans l'administration du réseau. L'idée de l'intégration du SDN à la mobilité distribuée a aussi été explorée dans l'architecture proposée dans [108]. De manière similaire au travail présenté dans [107], les fonctions de gestion de localisation et de relèves sont centralisées au niveau du contrôleur SDN, tandis que l'acheminement des paquets est entièrement distribuée au niveau des routeurs d'accès. Au sein des deux architectures, le contrôleur procède à l'optimisation des routes entre le CN et le MN lors d'une relève. Ce mode de fonctionnement est susceptible d'entraîner des coûts de signalisation additionnels nécessaires à la redirection du trafic de données.

Dans le but d'offrir des services fiables personnalisés tout en limitant les coûts d'installation et d'exploitation au sein des réseaux 5G, la technique de subdivision logique du réseau (network slicing) a été récemment proposée au sein des réseaux sans fils en tant que principal catalyseur de la convergence des services du réseau et des services personnalisés à la demande. En partitionnant un réseau physique en plusieurs partitions logiques (slices) dans le but d'offrir des services personnalisés, les ressources de l'infrastructure du réseau peuvent être allouées de façon dynamique et efficace aux différentes partitions du réseau logique en fonction des requis en matière de qualité de service. Dans la littérature existante, la gestion de la mobilité et l'allocation virtualisée des ressources n'ont pas été suffisamment étudiées pour les réseaux 5G basés sur le partitionnement virtuel du réseau. A cet effet, une architecture de partitionnement logique du réseau 5G est proposé dans [109] dans le but d'offrir un support de la gestion de la mobilité ainsi qu'un mécanisme virtuel d'allocation des ressources (intra-slice et inter-slice) au sein de l'architecture proposée. Les résultats de simulation ont montré que les mécanismes proposés facilitent l'allocation des ressources au sein du réseau d'une manière flexible, pavant ainsi le chemin à un partage de ressources optimale au sein des réseaux 5G.

2.4 Protocoles basés sur le mécanisme ID/Locator Split

Fondamentalement, les architectures basées sur ce mécanisme reposent sur le principe selon lequel, l'identificateur (indépendant de la structure topologique du réseau) est utilisé au niveau des couches application et transport pour identifier les sessions en cours sur les nœuds, tandis que le localisateur sert à repérer un nœud dans la configuration topologique du réseau. Dans cet échange, les processus en cours d'exécution au niveau des couches supérieures sont insensibles au changement de localisateurs dont elles ignorent l'information [110]. Cette section présente quelques protocoles dont le fonctionnement se base sur le mécanisme d'ID/Locator

Split. Selon l'approche utilisée, le découplage identificateur/localisateur s'effectue soit au niveau de l'hôte, soit au niveau du réseau. Certains protocoles se basent aussi sur une approche hybride. Le mécanisme de résolution sous-jacent au fonctionnement demeure néanmoins une composante essentielle dans le fonctionnement des différents mécanismes répertoriés.

Le protocole HIP [111] est l'un des importants protocoles qui mettent en œuvre le découplage de l'adresse IP en identificateur et localisateur au niveau de l'hôte. Le protocole introduit le concept d'identité, une abstraction permettant d'identifier un MN [112]. Ce nouvel espace de nom, dénommé Host Identity (HI) est représenté dans la pile du protocole par la couche Host Identification Layer. Au niveau de cette couche, un identificateur appelé Host Identifier est assigné à chaque MN du réseau. L'identificateur d'un MN est sa clé publique cryptographique. A partir d'une fonction de hachage, le protocole HIP permet d'associer à chaque nœud un HIT (appelé Host Identity Tag) qui est une représentation de l'identificateur d'hôte sur 128 bits. On parle de Local Scope Identifier (LSI) pour une représentation sur 32 bits.

L'idée principale du protocole HIP est d'assigner un nom globalement unique à chaque hôte doté de la pile IP. Le fonctionnement du protocole HIP se base sur un mécanisme dénommé « Base Exchange » au cours duquel, une association de sécurité s'établit entre deux hôtes [113]. La couche transport établit une connexion de bout-en-bout en se basant sur les HIT de la source et de la destination. Au niveau de la couche Host Identification, l'indirection ou mappage local se fait entre l'identificateur de la source et son localisateur (adresse IP). Pour le MN de destination, la fonction de mappage est indispensable. Les localisateurs de la source et de la destination sont donc remplacés dans le paquet à transmettre et envoyés sur le réseau Internet établissant ainsi la communication entre les deux parties [112].

Lors du déplacement d'un MN, celui-ci doit informer ses nœuds correspondants et mettre à jour le système de résolution. HIP gère la mobilité des nœuds grâce à un serveur statique dit, serveur de "rendez-vous". Le nœud une fois admis sur le réseau doit enregistrer l'adresse IP de son serveur de rendezvous dans l'entrée DNS. Dès lors, lorsqu'une source requiert l'adresse IP d'un MN mobile, il obtient l'adresse IP du serveur de son serveur rendezvous. En se basant sur l'adresse reçue, l'hôte source entre en communication avec le serveur de rendezvous, engage ensuite la procédure d'authentification et d'échange de clés. Le premier paquet de données reçu par le serveur de rendezvous est ensuite transféré au MN avec qui la communication s'établit directement désormais sans son intervention [112]. Le nœud mobile au cours de ses déplacements, informe par de fréquentes mises à jour, son serveur de rendezvous de sa position courante.

Le protocole Shim6 [114, 115] a été développé à l'origine dans le but d'apporter un support de domiciliation multiple (multihoming) à IPv6. Il est aussi considéré comme un protocole

ID/Locator Split. Contrairement à HIP, Shim6 n'introduit pas un nouvel espace de noms mais utilise une adresse IPv6 comme identificateur d'un MN [116]. L'identificateur d'un nœud appelé ULID (Upper Layer IDentificator) est mappé de manière dynamique à son localisateur au niveau de la couche shim6 sublayer déployée comme une sous-couche de la couche réseau. Cette couche intermédiaire assure une indépendance des couches supérieures (Application et transport) et la couche réseau [114].

Les communications basées sur Shim6 sont subséquentes à l'établissement d'un contexte d'association d'identificateurs de couches supérieures (ULID-pair context). Ce contexte d'association correspond à une paire d'identificateur (ULID) que la couche intermédiaire (shim6 sublayer) crée et maintient au niveau de chaque nœud pour une communication. Plusieurs applications de couches supérieures (TCP ou UDP) peuvent communiquer à travers un même contexte de communication. Cependant, pour une communication entre deux hôtes, le contexte établi n'est pas toujours le même [115] au niveau des deux hôtes car le protocole Shim6 supporte des communications par plusieurs chemins.

A chaque contexte de communication, chaque hôte génère et insère un paramètre de 47 bits appelé context tag dans l'entête des paquets transmis. Chaque communication est donc identifiée par deux paramètres context tag. Au cours d'une communication ou survient des erreurs de transmission de paquets ou d'inaccessibilité, les deux hôtes doivent spécifier de nouveaux localisateurs afin de rétablir la communication. Lors de la retransmission des paquets, le context tag du destinataire est inséré dans l'entête des paquets afin de lui permettre de retrouver le contexte d'association précédemment en cours afin de transmettre aux protocoles des couches supérieures les en-tête de message correspondants. Ce mécanisme permet de rendre transparent le changement de localisateurs aux protocoles des couches supérieures. Suite à un changement de localisateur dû à la mobilité, les identificateurs ULID demeurent inchangés tandis qu'un nouveau localisateur est choisi en fonction des adresses disponibles pour le nœud multi-domicilié. Le rétablissement de la communication nécessite cependant la mise en place d'un nouveau contexte d'association [112].

LISP (Locator ID Separation Protocol) [117] est l'un des principaux mécanismes d'ID/Locator Split proposé. Contrairement au protocole HIP qui opère son découplage au niveau de l'hôte, LISP découple l'identificateur d'un hôte de son localisateur au niveau des routeurs périphériques [118]. Les deux principaux concepts liés au protocole LISP sont : les identificateurs d'hôtes EIDs (Endpoint Identifiers) et les localisateurs de routage RLOCs (Routing Locators) [116]. Le protocole LISP est basé sur un mappage EIDs/RLOCs entre les routeurs périphériques et le réseau coeur ainsi que sur un tunneling à base du protocole UDP pour l'acheminement des paquets. Les adresses RLOCs sont utilisées comme localisateurs afin de

router les paquets à travers le réseau IP tandis que les adresses IP originales sont maintenues comme identificateurs des hôtes à des fins de routage local. Pour un routage inter-domaine, les adresses EIDs sont mappées en RLOCs par les routeurs périphériques. Les paquets sont ensuite acheminés vers les routeurs périphériques du réseau de destination. A cette étape, les adresses RLOCs sont supprimées des paquets et ces derniers sont ensuite routés vers l'hôte de destination sur la base de son identificateur EID contenu dans le paquet [119]. Avec le protocole LISP, les identificateurs d'hôtes sont aussi des adresses IPv6 et sont supposées être routables. Les deux principales caractéristiques du protocole LISP sont le mappage et l'encapsulation ainsi que la réécriture d'adresse [116].

CHAPITRE 3 DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE

Cette thèse vise à concevoir une architecture de gestion de la mobilité dans les réseaux denses 5G avec un support de la QoS. Afin d'atteindre cet objectif global, plusieurs objectifs spécifiques ont été définis à la Section 1.3. Ces objectifs ont été répartis en trois principales étapes ayant chacune fait l'objet d'un article de publication scientifique. Dans ce chapitre, nous présentons la démarche de l'ensemble du travail de recherche en mettant en évidence le lien entre les objectifs définis et les trois prochains chapitres de cette thèse.

3.1 Phase 1 : Conception du protocole de gestion de la mobilité

Le déploiement massif de SCs est l'une des avenues envisagées dans le but d'accroître la capacité totale des réseaux 5G. La densification du réseau pose cependant un défi majeur à la gestion de la mobilité car elle provoque des relèves fréquentes dont le coût de signalisation limite les performances du réseau. De plus, le raccordement individuel des SCs au réseau coeur n'est pas une solution économiquement viable aux opérateurs de réseaux mobiles.

Notre premier objectif a donc été de concevoir un protocole basé sur le mécanisme ID/Locator Split pour la gestion de la mobilité dans le réseau dense. Pour atteindre cet objectif, nous avons proposé un modèle de densification du réseau pour le support de la mobilité. Par la suite, nous avons conçu un modèle mathématique afin de modéliser l'approche de densification proposée. A la suite des deux étapes précédentes, le protocole LISP-LSA a été proposé pour la gestion de la mobilité. Nous avons par la suite procédé à une évaluation numérique des performances du protocole proposé. Ce premier objectif atteint a fait l'objet de l'article intitulé « *Local Anchor-Based Architecture for Mobility Support in Identifier Locator Split Networks* » et présenté au Chapitre 4 de la thèse.

3.1.1 Modèle de densification du réseau

Dans le modèle proposé, un domaine local LISP est partitionné en plusieurs LSAs dans le but d'offrir un support de mobilité locale. Une LSA est une structure logique du réseau qui regroupe en son sein, un ensemble de SCs localement administrées par une entité du réseau dénommé point d'ancrage local (LA). La configuration du LSA est une grille horizontale 2D. Ce modèle de représentation a été adapté du scénario du déploiement indoor à faible isolation des réseaux ultra-dense 5G [120]. Nous estimons que le modèle proposé peut également être

déployé dans un environnement outdoor dans lequel la vitesse des MNs est très réduite ou modérée. Au centre de la grille, est positionné le LA de contrôle du LSA. Nous avons fait l'hypothèse que la position des LAs au sein du domaine local est déterminée lors d'une phase préalable de planification par l'opérateur du réseau mobile. Au sein du LSA, les SCs sont alors déployées sur plusieurs couches concentriques autour du LA. Le nombre de couches de SCs déployées représente la taille K du LSA. De plus, chaque bloc de la grille contient une SC et correspond au rayon de couverture de la SC considérée.

3.1.2 Protocole de mobilité LISP-LSA

Le protocole de mobilité LISP-LSA a été introduit dans ce travail afin de gérer la mobilité des MNs conformément à la configuration de densification adoptée. Les déplacements des MNs entraînent deux types de relèves que nous avons considérées à savoir : la relève intra-LSA ainsi que la relève inter-LSA. Une relève intra-LSA implique deux SCs situées dans le même LSA, tandis qu'une relève inter-LSA implique deux SCs situées dans deux LSA différents. Lors de la phase d'exécution des relèves, nous avons adopté une approche contrôlée par le MN et assistée par le réseau. Cette approche a l'avantage de tenir compte des préférences des utilisateurs et de les combiner à la connaissance de l'état du réseau dont dispose le LA. De ce fait, la collaboration mène à une meilleure expérience utilisateur car les paramètres du MN ainsi que l'état du réseau sont pris en compte pour une meilleure décision. Lors d'une relève intra-LSA, la signalisation s'effectue entre le MN, le LA ainsi que les deux SCs impliqués dans la relève. Lors d'une relève inter-LSA, la signalisation de mobilité implique en plus des noeuds précédemment mentionnés, le LA du nouveau LSA de destination du MN ainsi que le GW du domaine local.

3.1.3 Modèle analytique de représentation du LSA

Un modèle analytique de représentation du LSA a été proposé en tenant compte à la fois du positionnement des SCs au sein de la grille de densification, du modèle de mobilité du MN, ainsi que des sessions actives en cours d'exécution lors des relèves. Notre démarche s'est basée sur les recommandations du projet METIS [120] et ceux du projet 3GPP [121] pour la simulation des réseaux de prochaine génération.

Tel que mentionné à la Section 3.1.1, chaque couche de SCs déployée autour du LA porte un indice de position i . Au sein de chaque couche, un second indice de position j a ensuite été attribué à chaque bloc de SC. Dans la grille de déploiement, un nombre maximal de K couches de SCs sont déployés autour du LA. Le modèle de mobilité utilisé au niveau du MN est la marche aléatoire (*Random Walk mobility*) en 2D avec une équiprobabilité de mouvements

dans toutes les directions considérées [120]. Il a été montré que ce modèle est convenable aux déplacements du MN lorsque ceux-ci sont restreints à un espace géographique limité [122]. De plus, la marche aléatoire pouvant être modélisée à l'aide d'une chaîne de Markov [123], l'évolution de l'activité du MN a été modélisée par une chaîne de Markov à temps discret.

Les variables d'état A_i^j du modèle de Markov dénotent la présence du MN avec une session active sur une SC localisée à la position j au sein de la couche i de la grille 2D de déploiement. Connaissant les paramètres de trafic (taux d'arrivée ainsi que la durée des sessions) nous avons donc établi les probabilités de transition entre les différents états de la chaîne de Markov. Ainsi, il a été possible d'évaluer la probabilité d'occurrence d'une relève intra-LSA ainsi que celle d'une relève inter-LSA.

3.1.4 Évaluation de performance

L'évaluation de performance a été réalisée selon l'approche numérique. A des fins de comparaison, nous avons retenu de la littérature, une version améliorée du protocole LISP-MN qui, à quelques variantes près, est représentatif de plusieurs protocoles de la littérature basés sur le mécanisme ID/Locator Split. Les paramètres de simulation utilisés ont été adoptés de la littérature. Deux métriques ont été évaluées à savoir les coût de signalisation d'une relève intra-LSA et le coût de signalisation d'une relève inter-LSA. Pour chacune des métriques considérées, le ratio d'amélioration des performances du protocole LISP-LSA a été évalué en comparaison avec le protocole LISP-MN amélioré.

- **Coût de signalisation de la relève intra-LSA** : Le coût de signalisation de la relève intra-LSA a été évalué en fonction de la taille K du LSA et du taux de mobilité m . L'expérience a montré une augmentation du coût de signalisation des relèves intra-LSA avec l'augmentation de K . Mieux, le gain de performance observé est d'autant plus élevé lorsque le réseau est densifié et m élevé. Le gain de performance observé est de 38% pour $K = 3$.
- **Coût de signalisation de la relève inter-LSA** : Le coût de signalisation de la relève inter-LSA a aussi été évalué en fonction de K et m . Les résultats obtenus ont révélé une augmentation du coût de signalisation des relèves inter-LSA avec la croissance de K . Cependant, le gain de performance observé est moins élevé dans le cas de la relève intra-LSA car le coût de signalisation est plus élevée lors d'une relève inter-LSA. Le gain de performance observé était de 21% pour $K = 3$.

Les performances du protocole LISP-LSA ont révélé l'utilité du modèle de densification proposé. Cependant, la diminution du gain de performance observé avec l'augmentation de la taille K du LSA suggère un compromis à trouver entre K et le taux de mobilité m .

3.2 Phase 2 : Gestion de la mobilité avec support de relèves proactives

La gestion de la mobilité demeure un enjeu important au sein des réseaux mobiles. Face à l'explosion du trafic vidéo et à l'omniprésence des applications sensibles au délai, il s'avère indispensable d'intégrer des mécanismes de relèves anticipées afin de maintenir un bon niveau de qualité de service dans la gestion de la mobilité au sein des réseaux denses 5G. A cet effet, le protocole LISP-LSA proposé à la phase précédente, a été étendu afin d'intégrer des relèves anticipées dans son fonctionnement. Pour atteindre cet objectif, nous avons considéré plusieurs métriques d'évaluation de performance dont nous avons développé les modèles analytiques. Ces modèles ont ensuite été implémentés afin de procéder à l'évaluation de performance du protocole étendu. De ce fait, le deuxième objectif spécifique a été réalisé. Les résultats obtenus ont fait l'objet du Chapitre 5, lequel présente l'article intitulé « *A Seamless Mobility Management Protocol in 5G Locator Identifier Split Dense Small Cells* ».

3.2.1 Relèves proactives et réactives

Les deux types de relèves (intra-LSA et inter-LSA) ont été étendus avec de nouveaux modes de relèves proactives. Les relèves proactives exploitent l'information sur l'état des liens de la couche L2 afin d'anticiper, selon une probabilité notée P_s , la rupture du lien sans fil et amorcer de manière anticipée l'exécution de la relève. Les relèves proactives ont l'avantage de réduire le coût de signalisation, la latence de la relève, la perte de paquets ainsi que le coût de livraison des données. En cas d'échec de la procédure de relève proactive, selon la probabilité $(1-P_s)$, le MN doit passer en mode de relève réactive qui consiste à s'enregistrer à nouveau sur la SC d'arrivée. Cette opération est nécessaire et est due à l'échec de transfert du contexte de mobilité du MN.

3.2.2 Métriques et modèles d'évaluation de performance

Afin d'évaluer les performances du protocole proposé, plusieurs métriques ont été considérées dans le cadre de notre analyse. Il s'agit du coût de signalisation, du coût de livraison des données, de la latence, du taux de perte de paquets, et de la charge active de la passerelle du domaine. Pour chaque métrique considérée, un modèle analytique d'évaluation de performance a été proposé.

Le coût total de signalisation a été évalué en sommant les coûts totaux des relèves intra-LSA et inter-LSA. Chacun de ces coûts a été évalué en faisant une somme pondérée du coût de la relève proactive et du coût de la relève réactive. Les coefficients de pondération ont été évalués en faisant la somme des probabilités stationnaires des états de la chaîne de Markov

établie à l'Étape 1 pour chaque type de relève.

Le coût total de livraison de données a été évalué, en sommant les coûts de livraison sur chaque segment du chemin de transmission des paquets, entre la source et la destination. Pour chaque type de relève, le coût de livraison des données considère le mode proactif et réactif. Il est à noter que le protocole LISP-LSA intègre un mécanisme d'optimisation de route aussitôt après la relève comme indiqué à la Section 5.5.2 du Chapitre 5. La probabilité d'acheminement des paquets de données sur les routes optimisées et non-optimisées ont été considérées lors de l'évaluation.

La latence de la relève a été évaluée en sommant le temps de relève au niveau L2, ainsi que le temps nécessaire aux mises à jour de localisation sur les divers noeuds du réseau. Ici également, les différents cas de relève intra-LSA et inter-LSA ont été considérés en tenant compte des modes proactifs et réactifs du déroulement des relèves.

Le taux de perte de paquets étant proportionnel à la latence de la relève, il a été évalué à partir de la latence de la relève en variant le taux d'arrivée des paquets.

La charge active du GW a été évaluée par rapport au volume de trafic de signalisation généré, dû au nombre de relèves traitées au sein du domaine local.

3.2.3 Implémentation

Les modèles analytiques développés ont été implémentés avec le logiciel MATLAB pour une évaluation numérique. Ces modèles ont également été implémentés avec le logiciel de simulation de réseaux ns-3.26 afin de valider les performances du protocole LISP-LSA. Les résultats de simulation ont été présentés avec un intervalle de confiance à 95%. Les paramètres de simulation utilisés ont été adoptés de la littérature. Nous avons considéré la version améliorée du protocole LISP-MN comme référence en vue de l'évaluation de performance. Pour chacune des métriques évaluées, le ratio d'amélioration des performances du protocole LISP-LSA a été évalué en comparaison à celles du protocole LISP-MN amélioré.

3.2.4 Évaluation de performance

Pour chacune des métriques considérées, nous avons mesuré les performances de chacun des protocoles afin d'évaluer le gain de performance.

3.2.4.1 Résultats numériques

Le coût total de signalisation a été évalué en fonction de K et m . Il a été observé que ce coût de signalisation croît en fonction de K . Le coût de signalisation est d'autant plus élevée lorsque m croît aussi. Ce coût de signalisation est réduit d'environ 30% lorsque $K = 2$. Ce résultat montre que la proximité du LA de la périphérie du réseau permet de réduire la charge de signalisation dans le réseau dont la configuration est densifiée. Le coût de signalisation a également été évalué d'une part en fonction du SMR et K , et aussi en fonction de $d_{LA,GW}$ et K . Le paramètre SMR ou Session-to-Mobility-Ratio, est le ratio du taux d'arrivée des sessions sur le taux de mobilité, tandis que $d_{LA,GW}$ exprime la distance entre le LA et le GW. Dans le premier cas, nous avons observé que le gain de performance est plus élevé, soit de 26% ($K = 2$), lorsque le SMR prend de petites valeurs. Autrement dit, lorsque le taux de mobilité est élevé dans un environnement densifié, le protocole LISP-LSA présente de meilleures performances en comparaison au protocole LISP-MN. Dans le second cas, plus la distance $d_{LA,GW}$ est élevée, plus élevé est le gain de performance observé. Le coût de signalisation a été réduit de 64% pour $K = 2$. Il y a donc un réel avantage à déployer le LA à la périphérie du réseau, lorsque le réseau est dense.

Le coût de livraison des données a été évalué en fonction du taux d'arrivée des paquets λ_p . Il a été observé que ce coût de livraison croît avec λ_p . Le protocole LISP-LSA implémente un mécanisme de transfert de paquet entre les deux SCs impliquées dans une relève, ce qui permet de réduire la perte de paquets engendrée lors des relèves. Les mises à jour de localisation étant plus longues dans le cas du protocole LISP-MN, elle sont également susceptibles d'entraîner un taux de perte plus élevé. Le gain de performance observé est de 66%.

La latence a été évaluée en fonction de K et de m . Nous avons observé que les mécanismes proactifs de relève du protocole LISP-LSA ont significativement contribué à réduire la latence des celles-ci. Le gain de performance réalisé est évalué à 54% ($K = 2$). Ce gain de performance est également dû au temps de mise à jour de localisation réduit du fait de la présence du LA de la périphérie du réseau. Le taux de perte de paquets qui est proportionnel à la latence des relèves a également été réduit de 54%.

3.2.4.2 Résultats de simulation

Le coût de signalisation a été évalué en fonction de m , du SMR et du nombre de MNs dans le réseau. Nous avons observé une augmentation du coût de signalisation avec les valeurs de m . Le gain de performance observé était de 36%. En variant le SMR , le coût de signalisation de la relève a été réduit de 28.5%. En variant le nombre de MNs dans le réseau, le coût de

signalisation a été réduit de 31% pour des vélocité réduites à modérées. Lorsque la vélocité est élevée, le gain de performance est cependant réduit à 17%.

Le coût de livraison des données a été évalué en fonction de λ_p et $d_{LA,GW}$. Nous avons observé que le coût de livraison des données augmente avec $d_{LA,GW}$. Le gain de performance observé a été de 62%. La latence de relève a été évaluée en variant m et un gain de performance de 61.5% a été observé. Le taux de perte de paquets a été évalué en fonction de λ_p et aussi en fonction de la vélocité v du MN. En variant λ_p , le gain observé est similaire à celui des résultats numériques. Lorsque la vélocité du noeud augmente dans une configuration de réseau densifié, la perte de paquets à été réduite d'environ 90%.

La charge active du GW a été évaluée en fonction du nombre de MNs et m d'une part, mais aussi en considérant le nombre de MNs et la densité du réseau λ_d . Une fréquence des relèves, qui se traduit par une augmentation de m , entraine une charge élevée sur le GW dans le cas du protocole LISP-MN. La réduction de cette charge par le protocole LISP-LSA est évaluée à 90%. Lorsque le réseau se densifie, la charge sur le GW est réduite de 84%.

3.3 Phase 3 : Conception de l'architecture de gestion de la mobilité

L'explosion du trafic temps réel, la prolifération d'applications sensibles au délai, la croissance du nombre de périphériques mobiles, le besoin d'évolutivité et de flexibilité dans la gestion de l'infrastructure du réseau, ainsi que les nombreuses exigences des réseaux 5G, forcent la mise en oeuvre de solutions novatrices au sein des réseaux mobiles.

La troisième phase de ce projet a consisté en la proposition d'une architecture de gestion de la mobilité. La conception de ladite architecture a nécessité la définition de ses composantes et leur structure interne, ainsi que la définition des mécanismes de contrôle. Par la suite, nous avons proposé les modèles analytiques du débit et de la probabilité d'interruption d'une session durant une relève comme métriques complémentaires à l'évaluation de performance.

Les nouveaux modèles analytiques et les composantes de l'architecture proposée ont été implémentés afin de procéder à l'évaluation de performance. De ce fait, les objectifs spécifiques 3 et 4 ont été atteints. Les résultats obtenus à cette phase ont fait l'objet du Chapitre 6, lequel présente l'article intitulé « *An Architecture for Mobility Management in 5G Identifier Locator Split Dense Small Cells* ».

3.3.1 Composantes de l'architecture

A l'Étape 1 du projet, nous avons introduit le concept du LSA dans un domaine local LISP pour la gestion de la mobilité. A cet effet, les composantes ci-dessous ont été définies :

- *La passerelle (GW)* : C'est l'entité du réseau qui permet de relier le domaine local au réseau global Internet. Le GW implémente plusieurs caches internes pour la gestion de localisation des MNs. Il joue également le rôle de proxy lors de la communication avec des sites distant qui ne prennent pas en charge le mécanisme d'ID/Locator Split.
- *Le point d'ancrage local (LA)* : Le LA est l'entité au coeur du concept de LSA. Il coordonne l'ensemble des SCs déployées dans sa zone d'administration et sert de point d'ancrage aux usagers à la périphérie du réseau. Il participe à la gestion de la mobilité locale ainsi qu'aux différents types de relèves.
- *Les routeurs tunnels (xTR)* : Ce sont les entités du réseau auxquels les MNs sont directement connectés. Ils forment le plan de transmission des données. Les xTRs encapsulent les paquets de données avant de les transmettre ou les désencapsulent à la réception de ceux-ci.
- *Le réseau overlay de signalisation* : Cette composante du réseau permet de gérer le trafic de signalisation et regroupe en son sein l'ensemble des LAs du domaine local, ainsi que le GW. Le réseau de signalisation est distribué et fonctionne sur la base du protocole LISP.

3.3.2 Mécanismes de contrôle

Les mécanismes de contrôle définissent les procédures de fonctionnement par lesquelles l'architecture proposée est capable d'assurer la gestion de la mobilité des utilisateurs. Ils impliquent une interaction entre plusieurs composantes de l'architecture pour l'accès et à la mise à la jour des caches. Les mécanismes de contrôle que nous avons définis sont les suivants :

- *L'enregistrement* : Cette procédure permet au MN, lors de sa première attache au réseau de s'authentifier afin d'accéder aux services qui correspondent à son profil. Cette procédure est également exécutée lors d'une relève réactive.
- *La recherche* : Le mécanisme de relève permet de localiser un MN au sein de l'infrastructure du réseau afin de lui acheminer ses paquets de données.
- *Le routage de données* : La procédure de routage définit le mécanisme d'acheminement des paquets de données de bout-en-bout entre une source et une destination.
- *La relève* : La procédure de relève définie permet de gérer la mobilité des MNs. Elle est assurée par le protocole LISP-LSA défini lors des étapes 1 et 2 précédentes.

3.3.3 Métriques et modèles d'évaluation de performance

Le débit et la probabilité d'interruption d'une session durant une relève ont été considérés en complément des métriques évaluées à l'Étape 2.

Le débit a été évalué en faisant le ratio du nombre de paquets reçus durant une session par la durée totale de session. Le nombre total de paquets reçus a été évalué en faisant la différence entre le nombre de paquets transmis durant une session et le nombre de paquets perdus. La durée totale de session quant à elle, inclut le temps total de transmission, le temps total des relèves, et le temps total d'acheminement des données.

Le rayon de couverture R , d'une SC, est un facteur important dans un réseau dense. La probabilité d'interruption d'une session a été évaluée en considérant le temps de résidence d'un MN dans le rayon de couverture d'une SC, ainsi que le temps de signalisation d'une relève. Une session est interrompue lorsque le temps de résidence du MN dans une SC, suite à une relève (intra-LSA ou inter-LSA), est inférieure au temps total de la signalisation engendrée par celle-ci.

3.3.4 Implémentation

Les modèles analytiques développés ont ensuite été implémentés avec le logiciel MATLAB pour une évaluation numérique. Ces modèles ont également été implémentés avec le logiciel de simulation de réseaux ns-3.26 afin de valider les performances du protocole LISP-LSA. Les résultats de simulation ont été présentés avec un intervalle de confiance à 95%. Les paramètres de simulation utilisés ont été adoptés de la littérature. Nous avons considéré la version améliorée du protocole LISP-MN comme référence en vue de l'évaluation de performance. Pour chacune des métriques étudiées, le ratio d'amélioration des performances du protocole LISP-LSA, a été évalué en comparaison à celles du protocole LISP-MN amélioré.

3.3.5 Évaluation de performance

Pour chacune des métriques considérées, nous avons mesuré le gain de performance du protocole LISP-LSA en comparaison au protocole LISP-MN. Nous avons considéré les approches numériques et de simulation pour l'évaluation de performance.

3.3.5.1 Résultats numériques

Le débit a été évalué en considérant divers scénarios. En variant le nombre de paquets par sessions, $E(S)$, et $d_{LA,GW}$, nous avons remarqué de manière générale, une augmentation du

débit lorsque les sessions sont plus longues. Le gain de performance est d'autant plus élevé lorsque $d_{LA,GW}$ est grande. Par exemple le gain de performance mesuré est de 20% lorsque $d_{LA,GW} = 20$ msec et de 32% lorsque $d_{LA,GW} = 100$ msec. Par ailleurs, en variant le débit en fonction de $E(S)$ et K , on note également une augmentation du débit. Le gain de performance observé est de 22%. En variant le débit en fonction de $E(S)$ et λ_p , nous avons également observé une croissance du débit avec un gain maximal de performance d'environ 22%.

La probabilité d'interruption d'une session a été évaluée en considérant divers scénarios. En variant la vitesse, v , des MNs et R , nous avons remarqué que le protocole LISP-LSA réduit de manière significative la probabilité d'interruption des sessions. Plus le réseau est densifié, plus élevé est le gain de performance observé. Par exemple la probabilité d'interruption est réduite de 31% lorsque $R = 15$ m, il est de 40% lorsque $R = 5$ m. En variant v et K , on constate également des gains de performance significatifs. De manière générale, la probabilité d'interruption croît avec la vitesse des MNs. Cependant, lorsque K augmente, le gain de performance observé diminue. Ce gain est de 33% lorsque $K = 2$, et il passe à 26% lorsque $K = 4$. Des résultats tout aussi satisfaisants ont été observés en variant v et $d_{LA,GW}$. Plus le MN est éloigné du réseau coeur, plus importante est le gain de performance observé. Par exemple, lorsque $d_{LA,GW} = 10$ msec, le gain de performance observé est de 35%. Ce gain passe à 55% lorsque $d_{LA,GW} = 100$ msec.

3.3.5.2 Résultats de simulation

Les métriques étudiées ont également été évaluées par une approche de simulation. En évaluant le débit du réseau en fonction de $E(S)$ et $d_{LA,GW}$, de significatifs gains de performance ont été observés. Le gain observé est d'autant plus élevé que le MN est éloigné du GW. Par exemple, le débit est amélioré de 21% lorsque $d_{LA,GW} = 50$ msec, et ce gain passe à 36% lorsque $d_{LA,GW} = 150$ msec. En variant $E(S)$ et λ_p , des gains de performance intéressants ont été observés, même si le nombre de paquets perdus augmente avec λ_p . Le gain observé est par exemple de 28% et 12% pour $\lambda_p = 40$ packets/sec et $\lambda_p = 150$ packets/sec respectivement. En variant v ainsi que la densité du réseau λ_d , nous avons remarqué une augmentation du gain de performance lorsque le réseau se densifie et que la vitesse du MN augmente. Par exemple, lorsque $\lambda_d = 15$ m, le gain observé est de 37% lorsque $v = 3.5$ m/sec. Ce gain augmente à plus de 90% lorsque $v = 5.5$ m/sec.

La probabilité d'interruption des relèves a été évaluée en variant d'une part v et $d_{LA,GW}$, et d'autre part v et le délai dans le réseau de transit δ_{core} . Cette probabilité est réduite par le protocole LISP-LSA de manière significative. Lorsque $d_{LA,GW} = 100$ msec, le gain de performance observé est de 33% lorsque $v = 3.5$ m/sec. Ce gain de performance augmente

significativement à 70% lorsque $v = 7.5$ m/sec. Nous avons également observé que la probabilité d'interruption des sessions tend généralement à augmenter lorsque δ_{core} augmente. Néanmoins, les performances du protocole LISP-LSA sont toujours meilleures lorsque v et δ_{core} augmentent.

CHAPITRE 4 ARTICLE 1 : LOCAL ANCHOR BASED ARCHITECTURE FOR MOBILITY SUPPORT IN IDENTIFICATOR LOCATOR SPLIT NETWORKS

Eric Mayeul O. Fafolahan and Samuel Pierre

Mobile Computing and Network Research Laboratory (LARIM)
Department of Computer and Software Engineering, Ecole Polytechnique de Montreal
C.P. 6079, succ. Centre-ville, Montreal, Quebec, H3C 3A7, Canada
Email : {eric.fafolahan, samuel.pierre}@polymtl.ca.

Conférence : In Proceedings of the Second International Conference on Internet of things,
Data and Cloud Computing. ACM, Mar. 2017.

Abstract

Although wireless networks have experienced tremendous improvements, mobility management remains an issue in a context of massive deployment of small cells, especially in the context of Internet of things and future 5G networks. In this paper, using the Identifier and locator Split concept, we propose a local anchor based network architecture for mobility support in dense small cells networks, where the local domain is partitioned into several location service areas (LSA).

We developed a mathematical framework to analyse the performance of the proposed scheme and Markov models are used to obtain the close form expression of the intra-LSA and inter-LSA handover parameters. Numerical results show that the proposed scheme achieve up to 38% and 21% of signaling cost saving per mobile node for the intra-LSA and inter-LSA handovers respectively ($K=3$), compared to the existing LISP-MN protocol.

Keywords : ID/locator Split, local anchor, mobility architecture, handover, performance evaluation, small cells.

4.1 Introduction

Mobile wireless communication has experienced explosive growth over the past decade, fueled by the popularity of smartphones and tablets [19]. This tremendous evolution has paved the way to important improvements along with a plethora of value added services. As pointed out by Cisco, there is an overall growth in the number of mobile devices and connections. Moreover, these devices are getting smarter with an increasing number of devices with higher computing resources, network connection capabilities that create a growing demand for more capable and intelligent networks [124]. In the presence of small cells, mobility management is one of the most challenging issues, owing to the dense network layout, the short cell radii and the potentially unplanned deployment [125]. Furthermore, in denser networks, mobility management implies more frequent handovers due to the cell size. Hence, it may no longer be exclusively triggered by radio quality, but also by network management decisions [9].

Identifier and Locator Split (ILS) and Core Edge Separation (CES) are two concepts introduced over the recent years in order to solve the shortcomings of the IP addressing scheme and enable routing scalability in Internet. LISP Mobile Node (LISP-MN) [126], a host based solution, has been proposed to support mobility in mobile LISP networks. LISP-MN introduces a lightweight router functionality on MNs and relies on a central mapping system which is subjected to single failure point problems, with high signaling cost and long handover delays. In a dense network deployment configuration, a frequent handover due to short cell radii, will result in high location update signaling cost with increased processing costs at gateways. Moreover, if MN's ID-to-LOC mapping information is cached at multiple remote ITRs, frequent handover will induce outdated mapping informations at remote ITRs. Therefore, packets destined to MN will be sent to old ETR that MN is no more attached to. This may lead to packets being discarded by the old ETR. Although many studies have been dedicated to mobility management in ILS networks, the proposed solutions may not be suitable for dense network deployment and frequent handovers scenarios. Local anchor-based mobility schemes were studied in [127] [128] [129] and show improvements in network performance.

In this work, we consider the Identifier and Locator Split concept and propose a local anchor (LA) based network architecture for location management in dense mobile ILS networks. The contributions of the paper are summarized as follow :

- We propose a Local Anchor (LA) based architecture for location management in mobile Identifier and Locator Separation networks where the local domain is partitioned into several location service areas (LSA).
- We analytically model the Location Service Areas (LSA) to study mobile nodes mo-

bility.

- We provide an expression of the intra-LSA, inter-LSA and total location update signaling cost for the proposed mobility architecture.
- We present numerical results to show the improvement of the proposed architecture in comparison to LISP-MN.

The remaining of this paper is organized as follows. In Section 2, we describe the proposed network architecture. The analytical model developed to study the proposed architecture is presented in Section 3. Performance evaluation of intra-LSA and inter-LSA handovers metrics are presented in Section 4, followed by numerical results and analysis in Section 5. Finally Section 6 concludes this paper.

4.2 Overview of the proposed Local anchor based mobility architecture

In this section, we describe the proposed local anchor based location management architecture in localized LISP networks (LISP-LSA). The proposed network architecture is presented in Fig. 6.1.

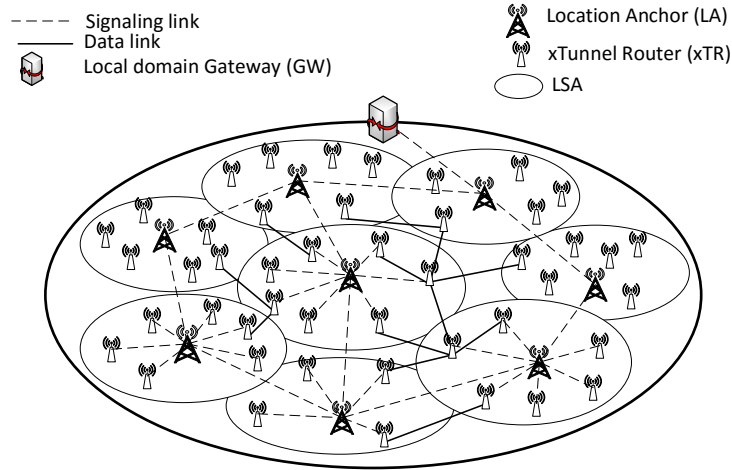


Figure 4.1 Local anchor based network architecture in localized LISP networks

4.2.1 Local anchor based location management in localized LISP networks

In the proposed architecture, LISP-LSA, a mobile LISP domain is partitioned into several location service areas (LSAs). Each LSA contains several tunnel routers (xTRs) which are the access routers (ARs). They are connected to the local location anchor (LA) of the LSA.

LAs and xTRs have several ingress/egress IP addresses. Local anchors are connected to the domain gateway (GW) through multihops links. To avoid single point of failure shortcomings at LAs, local anchors are connected through a single hop distributed mesh overlay network with one of their dedicated egress IP address. xTRs implement ITR/ETR functions for data routing inside and outside their LSAs. Tunnel routers have at least one egress IP address dedicated to signaling with the LA, and one egress IP address dedicated to data network. LAs also have at least one egress IP address dedicated to signaling with the domain gateway. It is important to mention that only signaling traffic is sent through the LA. Signaling and data networks are separated. Data network is composed of xTRs. Location service areas, are identified by an LSA_{ID} and it is broadcasted by its access points. The identifier and locator separation is processed on xTRs. When a handover occurs, the previous tunnel router temporarily serves as anchor router for ongoing selected sessions by MN for handover, until data delivery path is optimized.

4.2.2 Identifiers, Locators and the Mapping System

In the proposed architecture, three name spaces are introduced. Each mobile node (MN) is identified by a Global Endpoint Identifier (GEID), a Location Service Area LOCator (LOC_{LSA}), and a Local LOCator (LLOC). GEIDs are invariable, globally unique and represent the identity of the MN in the DNS system. LOC_{LSA} are IPv6 addresses whereas LLOC are IPv6 or IPv4 addresses. LOC_{LSA} are 128-bits long, globally routable, dynamic and used to identify the current LSA where the MN resides. LOC_{LSA} are composed of P_{LSA} and LA_{ID} which respectively, are the network prefix of the LSA and the identifier of the local anchor associated to the LSA where MN resides. LLOCs are 128-bits or 32-bits long, routable and dynamic. They are composed of P_{LSA} , ID_{xTR} , the ID of the xTR to which MN is attached and EID_{MN} .

Each domain gateway (GW) has a $GEID_{MN}$ -to- $ELOC_{LA}$ mapping cache. This mapping cache is used to map the identifiers of every node of the domain, to one of the egress IP addresses of the LA of the LSA where MN resides. Besides, each LA has two mapping caches : $GEID_{MN}$ -to- $ELOC_{xTR}$ and $GEID_{CN}$ -to- $ELOC_{LA}$. The former is used to keep, for MNs in the LSA, the mapping between nodes identifiers and one of the egress IP address of the MN's currently attached xTR. We assume that Egress addresses of xTRs in LSA are preconfigured and known by the LA. The later is used to map active correspondent nodes identifiers, to the egress IP addresses of their LA through the overlay network.

Indeed, each xTRs have two mapping caches : $GEID_{MN}$ -to- $LLOC_{MN}$ and $GEID_{CN}$ -to- $ELOC_{xTR}$. The former is used to map MN's identifier to its local locator, whereas the later is used to

map active correspondent nodes identifiers to one of the egress IP address of their attached xTR.

4.2.3 Location update signaling

In this section, we present signaling procedures for location update for both intra-LSA and inter-LSA handovers. In the proposed local anchor based scheme, L2 and L3 informations provided at network attachment stage are used to enable fast handovers for intra-LSA and inter-LSA handovers. Indeed, we define new types of messages, *Map Discovery Request*, *Map Select*, *MAP Forwarding* and *MAP Update* option, to be used for collaborative handovers in the proposed scheme.

4.2.3.1 Intra-LSA location update scheme

The signaling for the inter-LSA location update scheme is presented in Fig. 4.2. When L2 anticipated triggers are generated, MN selects the targeted access network and configures its new LLOC. It then selects which air interface to turn on for access network discovery and handoff preparation. MN then notifies its movement to LA by sending a *Map Select Request* message. This message contains its old and new LLOCs, IP ingress address of the previous (PxTR) and new (NxTR) tunnel routers, the selected list of $GEID_{CN}$ whose sessions need handover. LA starts a fast handoff signaling by sending *MAP Forwarding Request* message, containing MN's new and old LLOCs and the list of GEIDs of CNs, to both PxTR and NxTR. PxTR updates its $GEID_{MN-to-LLOC_{MN}}$ cache by setting $LLOC_{MN}$ to $ELOC_{NxTR}$ and sends a *MAP Forwarding Request* message to NxTR. This message contains the list of $GEID_{CN}$ and $ELOC_{CN}$ for CN previously indicated by MN, so that NxTR can update its $GEID_{CN-to-ELOC_{xTR}}$ cache. NxTR sends a *MAP Forwarding Response* to PxTR and the previous tunnel router starts packet forwarding to NxTR. NxTR sends a *MAP Forwarding Response* message to LA. The LA responds to MN with a *Map Select Response* message and NxTR buffers these forwarded packets until MN becomes attached to NxTR. Meanwhile, when MN is attached to NxTR, the buffered packets are delivered to it. Moreover, after the completion of the handover, subsequent packets are sent from NxTR with MN's new ELOC. This information is updated at xTR_{CN} so that data delivery route is optimized from MN to CN.

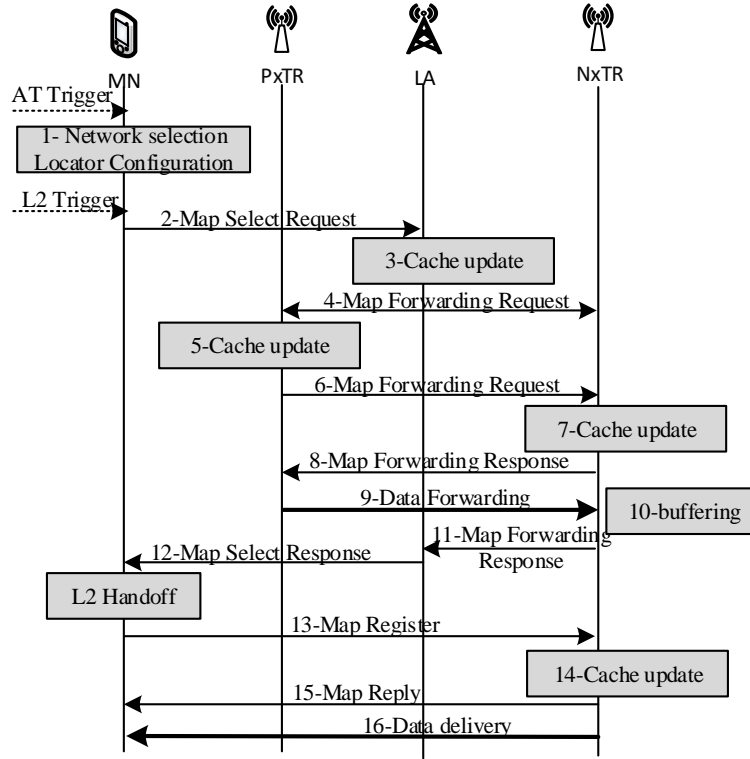


Figure 4.2 Signaling for the intra-LSA location update scheme

4.2.3.2 Inter-LSA location update scheme

For Inter-LSA handover, MN moves from old LSA controlled by previous location anchor (PLA) to a new location service area controlled by the new location anchor (NLA). The signaling for the inter-LSA location update scheme is presented in Fig. 4.3. The MN gathers the new LSA_{ID} and LA informations (ingress, egress IP address) from the L2 information received from neighboring xTRs and detect that it's moving outside the current LSA. At this step, MN notifies its movement to PLA by sending a *Map Select Request* message as in intra-LSA handover along with the new LSA_{ID} . The PLA then infers the NLA egress IP address and sends it a *Map Update* message through the overlay network. This message contains the list of $GEID_{CN}$ and the corresponding $ELOC_{LA}$ so that NLA updates its $GEID_{CN}$ -to- $ELOC_{LA}$ cache. Afterwards, NLA sends the *Map Update Response* message to PLA. Both PLA and NLA respectively send a *MAP Forwarding Request* message to PxTR and NxTR. This message contains MN's new and old LLOCs and the list of GEIDs of CNs. PxTR updates its $GEID_{MN}$ -to- $LLOC_{MN}$ cache and sets $LLOC_{MN}$ to $ELOC_{NxTR}$ and sends a *MAP Forwarding Request* message to NxTR. This message contains the list of $GEID_{CN}$ and $ELOC_{CN}$ for CN previously indicated by MN, so that NxTR can update its $GEID_{CN}$ -to- $ELOC_{xTR}$ cache. NxTR sends

a *MAP Forwarding Response* to both PxTR and NLA. PxTR sends a *MAP Forwarding Response* message to PLA and starts packet forwarding to NxTR. The previous LA responds to MN with a *Map Select Response* message. NxTR buffers these forwarded packets from PxTR until MN becomes attached to it. Meanwhile, when MN connects to NxTR, *Map Register* and *Map Notify* messages are exchanged between MN and NxTR. The buffered packets are then delivered to MN. Subsequent data packets are sent from NxTR with MN's new ELOC. This information is updated at xTR_{CN} so that data delivery route is optimized from MN to CN.

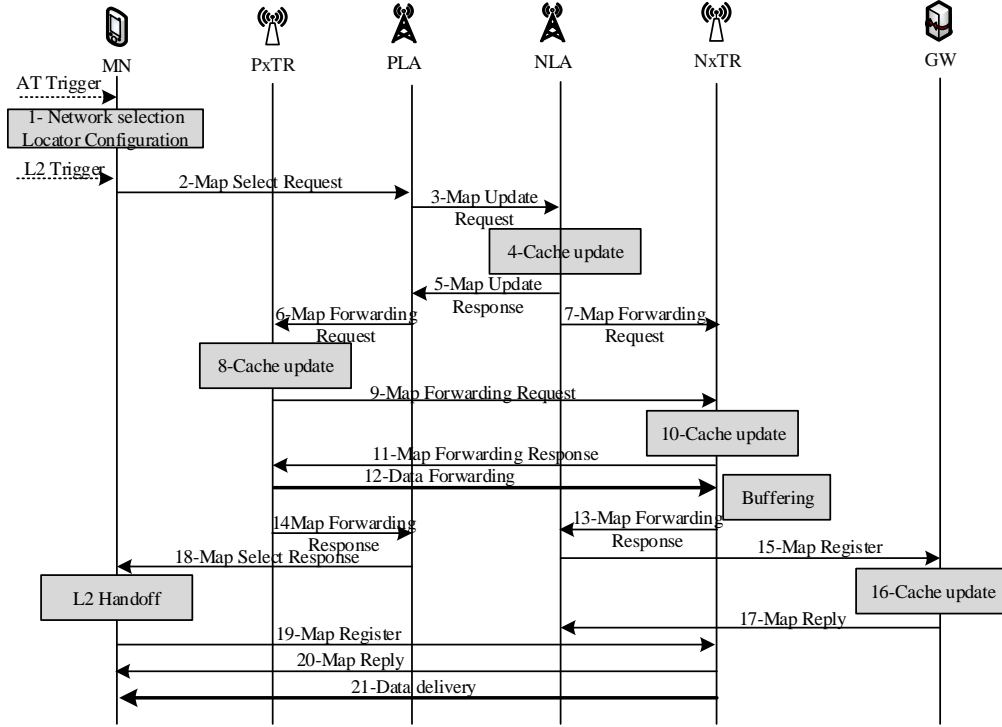


Figure 4.3 Signaling for the inter-LSA location update scheme

4.3 Analytical model

User Mobility and traffic models are crucial for efficient system design and performance evaluation [130]. To evaluate the performance of the proposed anchor based location management scheme in ILS networks, we utilize a discrete-time Markov model to capture the user behavior in the location service area containing the local anchor and the surrounding access routers. The stationary probability for a MN in the coverage of each xTR are derived from the developed analytical model. We use the results to obtain closed-form expressions for the evaluated handover performance metrics.

4.3.1 Model Description

We consider a square-based 2D grid network model as proposed in previous works [122] [131] to represent the LSA. Typically, a location service area with a side length of l is divided into $n \cdot n$ square elements that form the grid. Each element of the grid corresponds to a subnet covered by an xTR in the local service area. The LA is position at the center of the grid and all the xTRs are deployed surrounding the local anchor in tiers. The topology consists of K rings of small cells around the LA. Besides, Each subnet has four neighboring cells where MN can move with equal probability and each ring i contains $8i$ grid elements. Fig 4.4 represent an example of the grid model used. Each subnet of the grid is represented as a state variable A_i^j of the markov model, where i is the ring number and j is the index position inside that ring. The central grid element containing the LA is denoted by A_0^0 . Additionally, we use the state S_{idle} to indicate that MN has no active sessions.

A_3^{22}	A_3^{23}	A_3^{24}	A_3^1	A_3^2	A_3^3	A_3^4
A_3^{21}	A_2^{15}	A_2^{16}	A_2^1	A_2^2	A_2^3	A_3^5
A_3^{20}	A_2^{14}	A_1^8	A_1^1	A_1^2	A_2^4	A_3^6
A_3^{19}	A_2^{13}	A_1^7	A_0^0	A_1^3	A_2^5	A_3^7
A_3^{18}	A_2^{12}	A_1^6	A_1^5	A_1^4	A_2^6	A_3^8
A_3^{17}	A_2^{11}	A_2^{10}	A_2^9	A_2^8	A_2^7	A_3^9
A_3^{16}	A_3^{15}	A_3^{14}	A_3^{13}	A_3^{12}	A_3^{11}	A_3^{10}

Figure 4.4 2D Grid model

We model the evolution of the MN's activity as a stochastic process that occurs in a sequence of discrete steps Δt . Moreover, we assume that sessions arrive as a Poisson process with rate λ ; session duration and subnets residence time follow exponential distribution with mean rate $1/\mu$ and mean value $1/m$ respectively. It follows that $P_\lambda = \lambda\Delta t$, $P_\mu = \mu\Delta t$ and $P_m = m\Delta t$ are respectively the session arrival, session terminating, and the MN leaving the current subnet probabilities during a time slot Δt . When the time slot is sufficiently small, the evolution of the system can reflect real-life characteristics. The resolution of the Markov chain is time consuming when the number of rings K around the LA increases, and will result in a state space explosion. To avoid this issue, we make a state aggregation using the symmetric property of the 2-D model. All the states that exhibits the same properties are then aggregated. Let A_i^s denote the new aggregated state model where s is the state label

inside the ring i . After state aggregation, we have K rings of $i + 1$ states at each ring i . The boundary states of the LSA with respect to the current LA are denoted $A_{\overline{K}}^s$. The MN enters a new LSA at any of the boundary states. Consequently, we model the MN movement to enter the original states again from the boundary states once the mobile node starts moving within the new LSA.

$\overline{A}_3^{(5)}$	$\overline{A}_3^{(4)}$	$\overline{A}_3^{(3)}$	$\overline{A}_3^{(2)}$	$\overline{A}_3^{(1)}$	$\overline{A}_3^{(2)}$	$\overline{A}_3^{(3)}$	$\overline{A}_3^{(4)}$	$\overline{A}_3^{(5)}$
$\overline{A}_3^{(4)}$	$A_3^{(4)}$	$A_3^{(3)}$	$A_3^{(2)}$	$A_3^{(1)}$	$A_3^{(2)}$	$A_3^{(3)}$	$A_3^{(4)}$	$\overline{A}_3^{(4)}$
$\overline{A}_3^{(3)}$	$A_3^{(3)}$	$A_2^{(3)}$	$A_2^{(2)}$	$A_2^{(1)}$	$A_2^{(2)}$	$A_2^{(3)}$	$A_3^{(3)}$	$\overline{A}_3^{(3)}$
$\overline{A}_3^{(2)}$	$A_3^{(2)}$	$A_2^{(2)}$	$A_1^{(2)}$	$A_1^{(1)}$	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	$\overline{A}_3^{(2)}$
$\overline{A}_3^{(1)}$	$A_3^{(1)}$	$A_2^{(1)}$	$A_1^{(1)}$	A_0^0	$A_1^{(1)}$	$A_2^{(1)}$	$A_3^{(1)}$	$\overline{A}_3^{(1)}$
$\overline{A}_3^{(2)}$	$A_3^{(2)}$	$A_2^{(2)}$	$A_1^{(2)}$	$A_1^{(1)}$	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	$\overline{A}_3^{(2)}$
$\overline{A}_3^{(3)}$	$A_3^{(3)}$	$A_2^{(3)}$	$A_2^{(2)}$	$A_2^{(1)}$	$A_2^{(2)}$	$A_2^{(3)}$	$A_3^{(3)}$	$\overline{A}_3^{(3)}$
$\overline{A}_3^{(4)}$	$A_3^{(4)}$	$A_3^{(3)}$	$A_3^{(2)}$	$A_3^{(1)}$	$A_3^{(2)}$	$A_3^{(3)}$	$A_3^{(4)}$	$\overline{A}_3^{(4)}$
$\overline{A}_3^{(5)}$	$\overline{A}_3^{(4)}$	$\overline{A}_3^{(3)}$	$\overline{A}_3^{(2)}$	$\overline{A}_3^{(1)}$	$\overline{A}_3^{(2)}$	$\overline{A}_3^{(3)}$	$\overline{A}_3^{(4)}$	$\overline{A}_3^{(5)}$

Figure 4.5 LSA with bordering states

From the Markov model, with a probability $1 - P_\lambda$, MN stays in state S_{idle} . Let X_r and Y_r be the remaining subnet residence time and inter-session arrival time respectively. Then, X_r and Y_r are exponentially distributed with rate m and λ respectively due to the memoryless property of an exponentially distributed time. Let's θ be the probability that during the intersession arrival time, MN's locator hasn't changed in cache. Then [132], $\theta = \text{prob}(Y_r \leq X_r)$ and $\theta = \lambda / (\lambda + m)$. When a session arrives, MN wakes up from S_{idle} state into any of the aggregated states $A_i^{(s)}$ with probability $P_\theta = 8i\theta/N$, according to the cell density in each ring and the total number of cells, N , in the LSA. With probability P_μ , MN returns to state S_{idle} and remains in any state $A_i^{(s)}$ with an active session with probability $(1 - P_\mu)(1 - P_m)$. Moreover, with probability $P_m(1 - P_\mu)P(A_i^s, A_j^s)$, MN undergoes a handover with an active session from state A_i^s to state A_j^s . $P(A_i^s, A_j^s)$ is obtained from the user mobility characteristics and the number of rings in the LMA. Let π_{idle} and π_i^s $\{0 \leq i \leq K, 1 \leq s \leq i + 1\}$ the stationnary probability distribution of MN being at the states S_{idle} and A_i^s respectively. We define the following parameters as :

$$\alpha = \begin{cases} 1, & \text{if } K = 2 \\ 0, & \text{otherwise} \end{cases}; \quad \beta = \begin{cases} 1, & \text{if } i = K - 1 \\ 0, & \text{otherwise} \end{cases};$$

$$\gamma = \begin{cases} 1, & \text{if } i = j - 1; \\ 1/2, & \text{otherwise} \end{cases}; \quad \varphi = \begin{cases} 1, & \text{if } j = K - 1; \\ 1/2, & \text{otherwise} \end{cases};$$

The balance equations are given as follow :

$$\pi_{idle} = (1 - P_\lambda)\pi_{idle} + P_\mu \left(\sum_{i=1}^K \sum_{j=1}^{i+1} \pi_i^j + \pi_0 \right) \quad (4.1)$$

$$\pi_0 = \frac{\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_0 + \frac{1}{4}P_m(1 - P_\mu)\pi_1^1 \quad (4.2)$$

$$\pi_1^1 = \frac{8\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_1^1 + P_m(1 - P_\mu) \left(\pi_0 + \frac{1}{2}\pi_1^2 + \frac{1}{4}\pi_2^1 + \frac{\alpha}{2}\pi_1^1 \right) \quad (4.3)$$

$$\begin{aligned} \pi_i^1 &= \frac{8i\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_i^1 + P_m(1 - P_\mu) \left(\frac{1}{4}\pi_{i-1}^1 + \frac{1}{4}\pi_{i+1}^1 + \frac{1}{4}\pi_i^2 \frac{\beta}{4}\pi_K^1 \right), \\ &\forall K \geq 3; \forall 2 \leq i \leq K-1 \end{aligned} \quad (4.4)$$

$$\begin{aligned} \pi_K^1 &= \frac{8K\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_K^1 + P_m(1 - P_\mu) \left(\frac{1}{4}\pi_{K-1}^1 + \frac{1}{4}\pi_K^2 + \frac{1}{4}\pi_K^2 \right), \\ &\forall K \geq 2 \end{aligned} \quad (4.5)$$

$$\pi_1^2 = \frac{8\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_1^2 + P_m(1 - P_\mu) \left(\frac{1}{2}\pi_1^1 + \frac{1}{4}\pi_2^2 + \frac{\alpha}{4}\pi_2^2 \right) \quad (4.6)$$

$$\begin{aligned} \pi_i^{i+1} &= \frac{8i\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_i^{i+1} + P_m(1 - P_\mu) \left(\frac{1}{4}\pi_i^i + \frac{1}{4}\pi_{i+1}^{i+1} + \frac{\beta}{4}\pi_i^{i+1} \right), \\ &\forall K \geq 3; \forall 2 \leq i \leq K-1 \end{aligned} \quad (4.7)$$

$$\begin{aligned} \pi_K^{K+1} &= \frac{8K\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_K^{K+1} + P_m(1 - P_\mu) \left(\frac{1}{4}\pi_K^K + \frac{1}{4}\pi_K^K \right), \\ &\forall K \geq 2 \end{aligned} \quad (4.8)$$

$$\begin{aligned} \pi_i^2 &= \frac{8i\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_i^2 + P_m(1 - P_\mu) \left(\frac{1}{2}\pi_i^1 + \frac{\gamma}{2}\pi_{i-1}^2 + \frac{\gamma}{2}\pi_i^3 + \frac{\gamma}{4}\pi_{i+1}^2 \right. \\ &\quad \left. + \frac{\beta}{4}\pi_K^2 \right), \forall K \geq 3, \forall 2 \leq i \leq K-1; j=2; \end{aligned} \quad (4.9)$$

$$\begin{aligned} \pi_i^{j+1} &= \frac{8i\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_i^{j+1} + P_m(1 - P_\mu) \left(\frac{1}{4}\pi_i^j + \frac{\gamma}{2}\pi_{i-1}^{j+1} + \frac{\gamma}{2}\pi_i^{j+2} \right. \\ &\quad \left. + \frac{1}{4}\pi_{i+1}^{j+1} \frac{\beta}{4}\pi_K^{j+1} \right), \forall K \geq 4; \forall 2 \leq j \leq K-2 \text{ and } j+1 \leq i \leq K-1; \end{aligned} \quad (4.10)$$

$$\begin{aligned} \pi_K^2 &= \frac{8K\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_K^2 + P_m(1 - P_\mu) \left(\frac{1}{2}\pi_K^1 + \frac{1}{4}\pi_{K-1}^2 + \frac{1}{4}\pi_K^3 \right. \\ &\quad \left. + \frac{1}{2}\pi_K^1 + \frac{1}{4}\pi_K^3 \right) \end{aligned} \quad (4.11)$$

$$\begin{aligned} \pi_K^{j+1} &= \frac{8K\theta}{N}\pi_{idle} + (1 - P_\mu)(1 - P_m)\pi_K^{j+1} + P_m(1 - P_\mu) \left(\frac{1}{4}\pi_K^j + \frac{\varphi}{2}\pi_{K-1}^{j+1} + \frac{\varphi}{2}\pi_K^{j+2} \right. \\ &\quad \left. + \frac{\varphi}{2}\pi_K^{j+2} + \frac{1}{4}\pi_K^j \right), \forall K \geq 3; \forall 2 \leq j \leq K-1 \end{aligned} \quad (4.12)$$

$$\pi_K^1 = \frac{1}{4}\pi_K^1, \forall K \geq 2; \forall 1 \leq j \leq K-1 \quad (4.13)$$

$$\pi_K^{j+1} = \frac{1}{4}\pi_K^{j+1} \quad (4.14)$$

$$\pi_K^{K+1} = \frac{1}{2}\pi_K^{K+1} \quad (4.14)$$

The normalization equation is given by :

$$\pi_{idle} + \pi_0 + \sum_i^K \sum_{j=1}^{i+1} \pi_i^j + \sum_{j=1}^{K+1} \pi_K^j = 1 \quad (4.15)$$

We use the proposed mathematical framework and the stationary probability distribution developed above to derive the close form expression of intra-LSA and inter-LSA handover location update costs.

4.4 Performance Evaluation and Analysis

In this section, the performance of the proposed architecture is compared to LISP-MN. According to the grid, MN is attached to $PxTR$ on ring r_i before handover. When a handover occurs, MN may stay on the same ring, move forward or backward the LA and will, respectively, be on ring r_i , r_{i-1} or r_{i+1} . MN and CN are assumed to be located in the same local domain.

Hereafter are the general parameters to be used in the expressions of the costs.

- λ_s : Session arrival rate.
- μ_s : Session duration parameter.
- m : Mobility rate.
- K : Number of rings in LSA.
- $\dot{C}_{x,y}^h$ (resp. $\dot{C}_{x,y}^d$) : Control packet transmission cost between two nodes x and y distanced by h (resp. d) hops on wired (resp. wireless) network.
- $C_{x,y}$: Total control packet transmission cost between two nodes x and y. ($C_{x,y} = \dot{C}_{x,y}^h + \dot{C}_{x,y}^d$).
- τ : Unit transmission cost over wired link.
- κ : Weighting factor for the wireless link.
- Ψ_i^j : signaling cost when MN handovers from any subnet on ring r_i to a new subnet on ring r_j for intra-LSA handover.
- $\dot{\Psi}_i^j$: signaling cost when MN handovers from any subnet on ring r_i to a new subnet on ring r_j for inter-LSA handover.
- π_i^j : stationary probability distribution of state A_i^j .
- ξ : probability that MN undergoes handover with an active session ; $\xi = P_m(1 - P_\mu)$.
- M : The average size of control message.
- PC_x : cache lookup or cache update cost on network nodes.

4.4.1 Intra-LSA handover cost

In case of intra-LSA handover, the location is only updated at the LA. The intra-LSA handover signaling cost, C_{intra} , is given in equation 4.16 and 4.17.

— if $K = 1$

$$C_{intra} = \pi_0^0 \Psi_0^1 \xi + \pi_1^1 \xi \left(\frac{1}{2} \Psi_1^1 + \frac{1}{4} \Psi_1^0 \right) + \frac{1}{2} \pi_1^2 \Psi_1^1 \xi \quad (4.16)$$

— if $K \geq 2$

$$\begin{aligned} C_{intra} = & \pi_0^0 \Psi_0^1 \xi + \sum_{i=1}^{K-1} \pi_i^1 \xi \left(\frac{1}{4} \Psi_i^{i-1} + \frac{1}{2} \Psi_i^i + \frac{1}{4} \Psi_i^{i+1} \right) + \sum_{i=2}^{K-1} \sum_{j=2}^i \pi_i^j \xi \left(\frac{1}{4} \Psi_i^{i-1} \right. \\ & \left. + \frac{1}{2} \Psi_i^i + \frac{1}{4} \Psi_i^{i+1} \right) + \sum_{i=1}^{K-1} \pi_i^{i+1} \xi \left(\frac{1}{2} \Psi_i^i + \frac{1}{2} \Psi_i^{i+1} \right) + \sum_{j=1}^K \pi_K^j \xi \left(\frac{1}{4} \Psi_K^{K-1} \right. \\ & \left. + \frac{1}{2} \Psi_K^K \right) + \frac{1}{2} \pi_K^{K+1} \xi \Psi_K^K \end{aligned} \quad (4.17)$$

4.4.2 Inter-LSA handover cost

This happens when MN is located at ring K in the LSA and it moves in the direction that increase the number of rings, considering the current LSA. Based on that, C_{inter} is given in equation 6.6.

$$C_{inter} = \frac{1}{4} \pi_K^1 \dot{\Psi}_K^{K+1} \xi + \sum_{i=2}^K \pi_K^i \xi \frac{1}{4} \dot{\Psi}_K^{K+1} + \frac{1}{2} \pi_K^{K+1} \dot{\Psi}_K^{K+1} \xi \quad (4.18)$$

4.4.3 Cost definition

Ψ_i^j and $\dot{\Psi}_i^j$ incur packet transmission cost and processing cost on network entities for the successful completion of a handover. The packet transmission cost is proportional to the distance in hops between source and destination. The transmission cost of a control packet between nodes X and Y belonging to the wired and wireless part of a network can be respectively expressed as [130] : $\dot{C}_{x,y}^h = \tau \cdot h \cdot M$ and $\dot{C}_{x,y}^d = \tau \cdot \kappa \cdot d \cdot M$.

1. LISP-LSA scheme : the signaling costs are given as :

$$\Psi_i^j = \begin{cases} 4C_{MN,LA} + 2C_{AR,LA} + 2C_{MN,AR} + PC_{AR} + PC_{LA}; & \text{if } (i,j) \in \{(0,1);(1,0)\} \\ 2C_{MN,LA} + C_{AR,LA} + 2C_{AR,LA} + 2C_{AR,AR} + 2C_{MN,AR} + 2PC_{AR} \\ + 2PC_{LA}; & \text{otherwise.} \end{cases} \quad (4.19)$$

$$\begin{aligned} \dot{\Psi}_K^{K+1} &= 2C_{MN,LA} + 2C_{LA,LA} + 4C_{LA,AR} + 2C_{AR,AR} + 2C_{LA,GW} \\ &\quad + 2C_{MN,AR} + 2PC_{AR} + PC_{LA} + PC_{GW} \end{aligned} \quad (4.20)$$

2. LISP-MN scheme : In case of LISP-MN, there is no local anchor deployment. When a handover occurs, the location of the MN is updated on the gateway. The location update signaling cost in case of LISP-MN is given by equation 4.21.

$$\begin{aligned} \Psi_i^j = \dot{\Psi}_i^j &= 8C_{MN,AR} + 5C_{AR,LA} + 5C_{LA,GW} \\ &\quad + 3C_{GW,AR} + 3PC_{AR} + PC_{GW}. \end{aligned} \quad (4.21)$$

4.5 Numerical Results

In this section we present the numerical evaluation of the proposed scheme and LISP-MN based on the performance metrics presented in the previous section. The default parameters setting used in the simulations are presented in Table 4.1.

We first evaluate the average signaling cost during an intra-LSA handover in term of transmission and processing costs of the messages from the perspective of a single MN. Fig. 4.6 and Fig. 4.7 show the intra-LSA signaling cost of both LISP-LSA and LISP-MN schemes, plotted as a function of the number of rings (K) and mobility rate (m) in the location service

Table 4.1 System parameters

Parameter	Value	Parameter	Value
λ	0.001/s	PC_{LA}	12
μ	0.01/s	PC_{GW}	24
m	0.1/s	$d_{MN,AR}$	1
Δt	0.01s	$h_{LA,GW}$	2
K	1-10	$h_{GW,ARcn}$	2
$M(bytes)$	60	PC_{AR}	8
τ	1	κ	1.5

area.

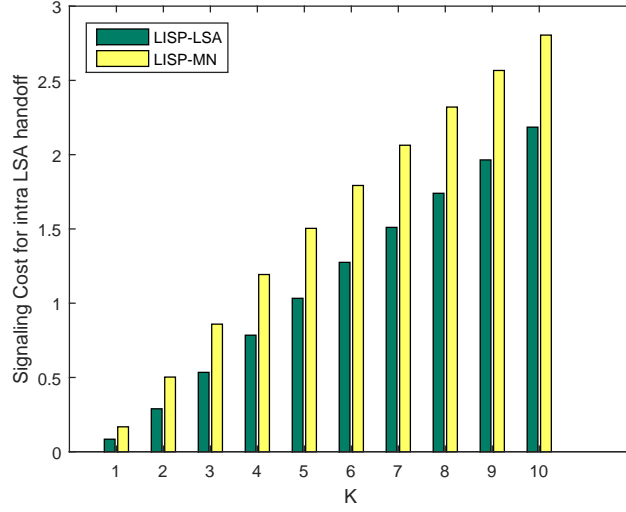


Figure 4.6 Intra-LSA handover signaling cost versus K ($m=0.1$)

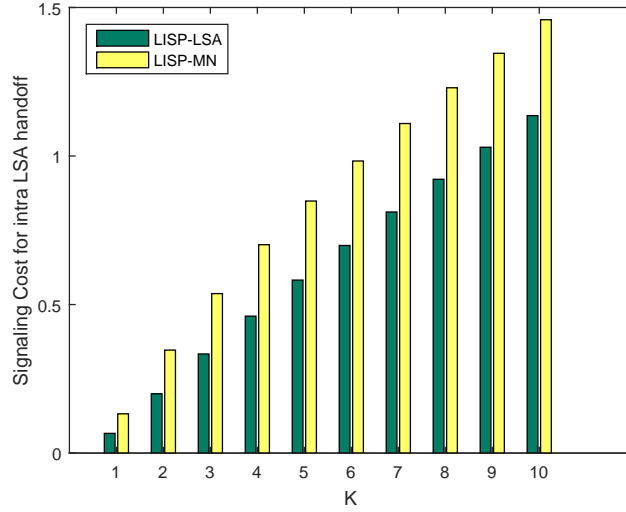


Figure 4.7 Intra-LSA handover signaling costs versus K ($m=0.05$)

We observe that for both schemes, the signaling cost increase as the number of rings (K) in LSA increases. The generated cost is higher when the cell residence time ($1/m$) is small, i.e frequent handovers happens. We observe that the cost saving of the proposed scheme decrease from 49% for $K=1$ to 22% for $K=10$. This is explained by the fact that small values of K implies frequent handovers in both protocols, therefore the proposed LISP-LSA protocol achieve a higher cost saving. The cost saving is about 38% for intra-LSA handover for $K=3$. These results validates the fact that, in dense network deployment scenarios with frequent

handover due to high node mobility, the proposed protocol minimizes the location update signaling load during intra-LSA handover so that network performance can be improved.

In Fig. 4.8 and Fig. 4.9, the average inter-LSA handover signaling cost is plotted as a function of the number of rings (K) and mobility rate (m) in the location service area.

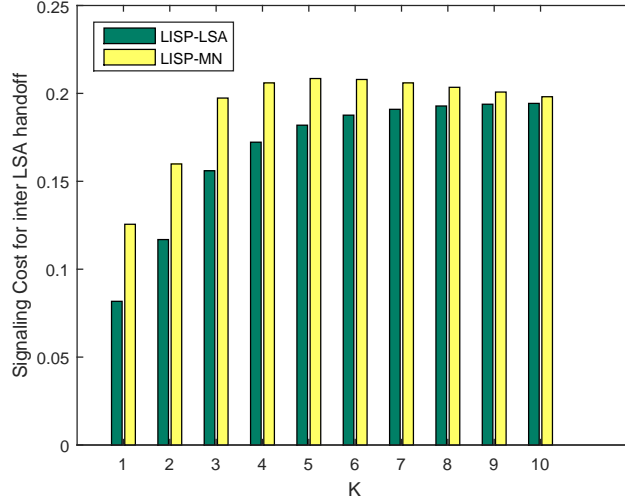


Figure 4.8 Inter-LSA handover signaling cost versus K ($m=0.1$)

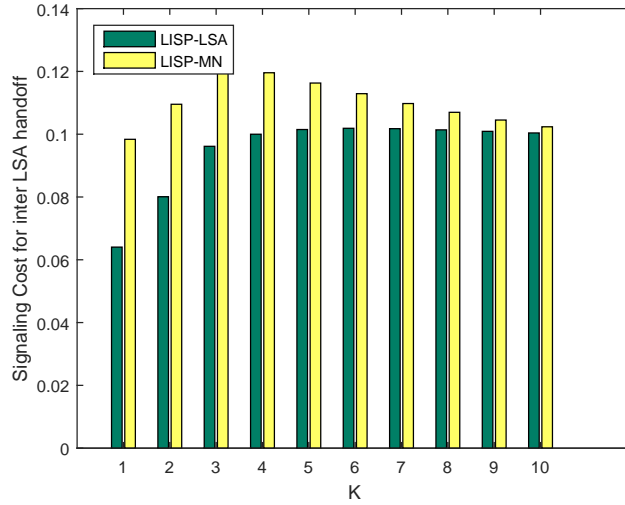


Figure 4.9 Inter-LSA handover signaling cost versus K ($m=0.05$)

It is shown that the signaling cost increases when the number of rings (K) increases in LSA and frequent handovers ($m=0.1$) also incurs more signaling cost in the network. However, as the number of rings increase in LSA, the cost saving realized by LISP-LSA decreases from 35% ($K=1$) to less than 5% ($K=10$). This shows that, the cost saving is lost for high values of K since the inter-LSA handover incurs more signaling compared to the intra-LSA

handover. The cost saving is about 21% for inter-LSA handover for $K=3$. This suggest that up to certain value of K , the proposed scheme still can achieve good cost savings. However, there is a tradeoff to be find between the number of ring (K) in the LSA and MN mobility rate (m) so that the proposed scheme can always acheive better cost savings.

4.6 Conclusion

In this work, we introduced LISP-LSA, a local anchor based architecture for mobility support in Identifier and Locator Split Networks. For this, we developped a mathematical framework to analyse the proposed scheme using markov models. Based on this model, we derive close form expression of the intra-LSA and inter-LSA handover signaling cost and compared with LISP-MN. The numerical results shows that the proposed scheme can acheive up to 38% and 21% of signaling cost saving per node for the intra-LSA and inter-LSA handovers. For future work, we intend to study more handover metrics such as the total signaling cost ; as function of the number of handovers to local anchor and gateway for both schemes per sessions ; handoff delay and packet lost under different probabilty of cell residence time and system parameters, in order to provide a more thorough framework for analysing the proposed handover scheme. The proposed architecture will then be integrated in to a Software Defined Network.

CHAPITRE 5 ARTICLE 2 : A SEAMLESS MOBILITY MANAGEMENT PROTOCOL IN 5G LOCATOR IDENTIFICATOR SPLIT DENSE SMALL CELLS

Eric Mayeul O. Fafolahan and Samuel Pierre

Mobile Computing and Network Research Laboratory (LARIM)
Department of Computer and Software Engineering, Ecole Polytechnique de Montreal
C.P. 6079, succ. Centre-ville, Montreal, Quebec, H3C 3A7, Canada
Email : {eric.fafolahan, samuel.pierre}@polymtl.ca.

Revue : Accepté pour publication dans le journal
IEEE Transactions on Mobile Computing, Avril 2019.

Abstract

Network densification with Small Cells (SCs) has emerged as a key technique to increase the 5G network capacity. However, in a densified network, fast mobile nodes will experience frequent handovers with a high signaling load, handover latency, and packet loss, due to the short cell radius. Indeed, Distributed Mobility Management (DMM) protocols aim to solve the shortcomings of centralized mobility management solutions such as poor scalability and non-optimal routing. However, when the cell residence time is short, DMM protocols might suffer from increased costs and limited performance. This paper proposes a localized mobility management protocol in 5G dense SCs, based on the locator identifier separation protocol, local mobility anchoring, and fast handovers concepts. The proposed scheme divides a local domain into several location service areas, each controlled by a local anchor. We provide the analytical models of several handover metrics, namely the average total signaling cost, the data delivery cost, the handover latency, and the packet loss. Numerical and simulation results show significant cost savings, up to 30% in signaling overhead, up to 53% of packet loss, and up to 90% of processing load reduction at the core of the network compared to the existing lisp mobile node protocol.

Keywords : 5G, lisp, small cells, mobility protocol, local anchor, handover, performance evaluation.

5.1 Introduction

The proliferation of smart mobile devices, equipped with powerful processing and multimedia capabilities, combined with the latest technological innovations in wireless networks, have led us to the paradigm of the globally connected society. The increasing demand for high data transmission rates and network capacity, new wireless technologies with backward compatibility, ultra-low latency, flexible and dynamic network resource allocation mechanisms highlight the need to rethink the wireless networks and give birth to the 5G networks paradigm. In 2016, more than 63% of all traffic from mobile-connected devices (almost 84 exabytes) has been offloaded to the fixed network using Wi-Fi devices and femtocells each month. Indeed, of all IP traffic (fixed and mobile) in 2021, 50% will be Wi-Fi, 30% will be wired, and 20% will be mobile [133]. Therefore, Small Cells (SCs) and ultra dense network architectures have emerged as one of the promising key techniques to support the design of the 5G networks [5,6]. However, because of its early stage of development, the network architecture design for the emerging 5G networks under ultra-dense cellular layouts, requires more investigation [8]. Moreover, mobility management which allows reachability and maintains ongoing communication during roaming of mobile users in different networks [134], still remains an issue to be solved in future networks, because of the density of the layout, the short cell radii which also may be deployed without any formal prior planification [125]. Indeed, the fast Mobile Nodes (MNs) may experience frequent handovers under SCs, and several parameters owing to network management conditions may also trigger handover decisions [9]. Identifier/Locator Separation (ILS) architectures are being actively discussed and pursued at the Internet Engineering Task Force (IETF) Working Groups to solve the problems of mobility, multihoming, and security in the current Internet architecture [135].

Locator and Identity Split Protocol (LISP) [117] is a promising solution to the emerging 5G networks [136]. LISP supports the separation of the IPv4 and the IPv6 address spaces following a network-based map-and-encapsulate scheme. It brings scaling benefits by splitting the current IP addresses into Endpoint Identifiers (EIDs) and Routing Locators (RLOCs). The EID is a non-globally routable IP address, and the RLOC is a globally routable IP address attached to routers. Routers consult a mapping system which maps the EID prefixes onto site router's RLOCs in order to locate the MNs. Although LISP can natively support mobility and multihoming, it has been designed for fixed environments.

This work is an extension of our early work [137], which mainly introduced the concept of the Location Service Areas (LSAs) into a mobile LISP domain. It also evaluated the average intra-LSA and the average inter-LSA handover signaling costs. In this work, we consider the fast handover schemes, the analytical models of the total signaling cost, the data

delivery cost, the handover latency, the packet loss, and the simulation-based evaluation of the proposed protocol. We assume that, in the coming years, the MNs will be equipped with more sophisticated features enabling them to proactively detect, anticipate, and start fast handovers with an accurate precision. However, the design of the handover algorithm and the related network selection mechanisms are out of the scope of the proposed protocol. The main contributions of the paper are summarized as follows :

- We propose a localized anchoring protocol for mobility management in the future 5G dense SCs using the ILS mechanism to bring more flexibility in mobility management. Moreover, the proposed protocol considers the intra-LSA and the inter-LSA fast handover schemes in localized mobility scenarios in order to minimize the handover costs. In the proposed network, the local domain is divided into several LSAs, each controlled by a Local Anchor (LA). Moreover, the proposed protocol uses a collaborative approach for mobility management where the MN initiates the handover which is assisted by the network. This approach combines the configured user preferences on mobile devices with the knowledge of the network state at the network nodes.
- We propose a two-dimensional grid model of the LSA, following the guidelines of the Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS) [120], and the 3rd Generation Partnership Project (3GPP) [138] for ultra-dense networks deployment and femtocell-related performance analysis.
- We propose a discrete-time Markov model to study the mobility and the traffic models of the MNs.
- We provide the analytical expressions of several handover metrics for the LISP-LSA and the LISP Mobile Node (LISP-MN) protocols, namely the total signaling cost, the data delivery cost, the handover latency, and the packet loss rate.
- We implement the proposed mobility management protocol using the network simulator ns-3.26 to show its feasibility from a practical point of view. Moreover, we define several simulation scenarios to evaluate the proposed protocol.
- We present numerical and simulation results to show the performance gains of the proposed LISP-LSA protocol in comparison to the LISP-MN protocol.

The remaining of this paper is organized as follows : Section 2 briefly introduces the background and the motivation to this study. Section 3 presents the proposed architecture and the mobility schemes of the LISP-LSA protocol. Section 4 describes the analytical models developed for this study. Section 5 describes the expressions of several handover metrics. Section 6 presents the numerical and the simulation results showing the performance gain of the proposed protocol. Finally, Section 7 concludes this paper.

5.2 Background and Motivation

The current Internet routing and addressing architecture uses a single namespace (the IP address) to simultaneously express a device's identity, and its location within the network. The LISP protocol introduces a change to the Internet architecture by replacing the IP addresses with EIDs and RLOCs. The EIDs are used to identify the MN's sessions whereas the RLOCs define the global routing addresses through the Internet. The LISP protocol aims to improve the routing scalability while also facilitating flexible address assignment for multi-homing, provider independence, and mobility.

LISP-MN [126], a host-based protocol, was introduced to support mobility in LISP networks by adding a routing function to the MNs. The MN is statically provisioned with an EID and behaves as a LISP site in accordance to the LISP architecture [139]. During a handover, the MN receives a new RLOC and updates its EID-to-RLOC mapping information to its current Mapping-System (MS) to maintain reachability at its new location [139]. The LISP-MN protocol relies on a central MS, which may experience an increased signaling load and become a single point of failure when the number of MNs increases. Some improvements to the LISP-MN protocol are proposed in [140] with the use of a Local Mapping Server. In [141], a distributed mapping control scheme called LISP-DMC was proposed based on the existing LISP concepts. In LISP-DMC, the control traffic is processed at a distributed mapping system. According to the authors, LISP-DMC reduces the control traffic overhead and the total signaling delay. The work in [142] proposed a network-based distributed mobility control in localized mobile LISP networks. According to the authors, the proposed distributed scheme achieves better performance than the existing centralized scheme in terms of the signaling loads, binding update/query, and handover delays.

Several mobility schemes using a LA have been previously proposed to reduce the signaling load at the core of the network. The work in [127] proposed two location management schemes in SCs networks. The position of the MN inside a SC or a tracking area is registered to the LA, which is selected among the SCs, in such a way that the position of the MN is accessed with a finer granularity. The location update is performed using the X2 interface in order to reduce the signaling overhead at the core network. In [128], a LA based architecture using SCs clustering was proposed for handover management in coordinated SCs. The LA in this scheme is a SC that maintains a link with a group of neighboring SCs. The proposed scheme minimizes the handover interruption as well as the handover-related costs. The work in [129] proposed two local mobility management in SCs based on the X2 traffic forwarding chain. When a handover happens, the targeted SC establishes a local path with the previous serving SC without sending a path switch request to the core network. The proposed scheme has

shown improved performance compared to the 3GPP standard scheme. In [143], the authors proposed a novel mobility management framework that addresses critical issues including frequent handovers, handover failure, and excessive energy consumption for seamless handovers in the emerging dense cellular networks with mmWave capabilities. The work in [144] proposed a mobility management scheme to improve the location management efficiency in terms of total signaling cost and intersystem paging delay. The work in [145] proposed a new mobility model in 5G SCs networks using the concept of individual mobility model (IMM). The IMM considers the impact of human tendency and clustering behaviors which are not considered in the existing mobility models. While this work mainly proposed a new mobility model, our work proposes a new mobility management protocol in 5G dense SCs. Moreover, the concept of community is used in [145] to represent a specific region where a large number of MNs are assembled. Our work considers clusters of SCs in localized regions, and the MNs are associated to the SCs. Although significant contributions have been done to solve the mobility management issue, the emerging 5G dense SCs networks still require new mobility schemes.

Firstly, Centralized Mobility Management (CMM) schemes suffer from several well known limitations. In fact, the centralization of the network architecture results in a significant increase of the signaling load on the anchoring point, low scalability, suboptimal routing, the existence of a single point of failure which would lead to a considerable network performance degradation, security vulnerabilities, and non-optimal use of the network resources [146]. To solve these limitations, the Distributed Mobility Management (DMM) protocols [98, 147–149] aim to distribute the MN's traffic to the closest access points to it, contributing to the decentralization of the mobile network architecture. However, recent performance studies have shown considerable limitations in their performance, especially when the MN's cell residence time is shorter, or when the number of active sessions initiated within the visited networks is high [26, 33, 34]. When the online time of the MN is long enough, flows may survive through many handovers, requiring to set up multiple levels of tunnels with an increased signaling overhead. Therefore, the efficiency of such DMM schemes partly depends on how long the flows last, compared to the MN's cell residence time within the coverage of the SCs [34].

Secondly, network scalability and service continuity are two stringent performance requirements for 5G networks. With the advent of the Internet-of-Things, CMM protocols cannot handle the mobility signaling generated by billions of connected mobiles devices, because of the increased processing load at the LA which becomes a bottleneck. Moreover, 5G networks are expected to support several emerging applications, which require ultra-low latency, such as the tactile Internet, the virtual reality, and multi-player gaming [5, 6]. However, the increased handover latency of CMM protocols is not suitable for these applications. Indeed, the

DMM schemes design the mobility management procedure for a one-time handover [34]. As the MN undergoes frequent handovers in dense SCs networks, when the proportion of long flows is large, DMM schemes may incur high handover latency as well as increased signaling overhead [22]. The increased handover latency of both CMM and DMM schemes also induces a high packet loss rate for ongoing sessions.

Finally, the proposed mobility management schemes in ILS networks [126, 140–142] suffer from high signaling overhead, long handover latency during location location updates, and high packet loss. Therefore, they are not suitable for the future 5G dense SCs deployments. With these schemes, it is assumed that the mapping update service is located at the domain Gateway (GW) or at a remote MS. Accordingly, whenever the MN attaches to a new access point, the new locator must be updated at the remote GW or MS. This approach does not scale well because in a dense SCs network layout, fast MNs will experience frequent handovers due to the short cell radii. The resulting high signaling cost will increase the processing cost at the GW or at the MS. Moreover, if the MN's mapping information is cached at multiple remote Ingress Tunnel Routers (ITRs), frequent handovers will induce outdated mapping information at the remote ITRs, and this situation will require additional queries to the MS. Therefore, the data packets that are destined to the MN will be sent to the old Egress Tunnel Routers (ETR) to which the MN is not anymore attached to [38]. This will cause several data packets to be discarded by the old ETR resulting in a significant packet loss rate.

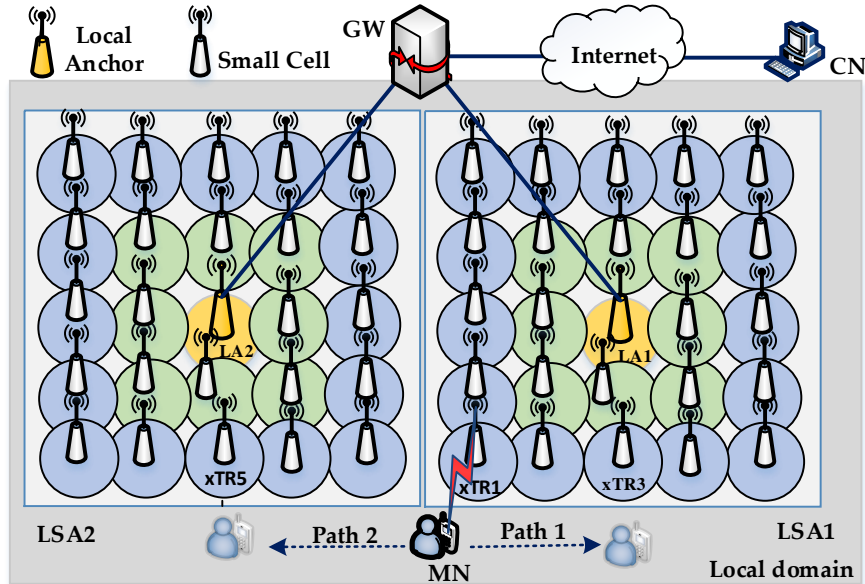


Figure 5.1 Local anchor based network architecture in localized LISP networks

5.3 Overview of the proposed Local anchor based network architecture

This section describes the proposed LISP-LSA mobility management protocol. Moreover, we present the mobility schemes on which the analytical study is based.

The proposed architecture is presented in Fig. 6.1. The local LISP domain is partitioned into several LSAs, where each LSA is controlled by a LA, and contains several SCs. These SCs act as tunnel routers (xTRs) according to LISP concepts. Indeed, the LAs are connected to the GW through a distributed overlay signaling network whereas the SCs are wired to the LA forming the underlay data network. In this architecture, the network nodes, namely the GW, the LAs, and the xTRs, use local caches to handle mobility informations. For more details on the network architecture, we refer the reader is referred to our previous work [137].

During a handover, the MN detaches from its serving SC or Previous xTR (PxTR) and attaches to the targeted SC or New xTR (NxTR). Moreover, a mobility context is exchanged between the nodes involved in a handover. It contains the MN's old and new ip addresses along with the IP addresses of the PxTR, the NxTR, and the Correspondent Nodes (CNs).

5.3.1 Intra-LSA Handover Scheme

The intra-LSA handover occurs when the PxTR and the NxTR belong to the same LSA. In Fig. 6.1, the MN undergoes an intra-LSA handover when it moves along path 1, detaching from xTR1 to connect to xTR3 in LSA1. Fig. 5.2 describes the intra-LSA handover signaling.

5.3.1.1 Proactive Intra-LSA Handover

When the L2 triggers are generated on the wireless link, the MN proactively select the targeted SC and configures its new local locator or IP address accordingly. Then, it sends a Map Select Request message with its mobility context to the LA. The latter forwards the mobility context both to the PxTR and to the NxTR through a Map Forwarding Request message. The PxTR updates its local caches and sends a Map Forwarding Request message to the NxTR to establish a new data forwarding path during the handover. Afterwards, the NxTR acknowledges the request by sending a Map Forwarding Response back to the PxTR before it starts forwarding the handover data packets. Meanwhile, the NxTR sends a Map Forwarding Response message to the LA which responds back to the MN with a Map Select Response message. The forwarded packets from the PxTR are buffered on the NxTR which later delivers them to the MN. It is also worth pointing out that the forwarded packets through the PxTR during the handover, follow an indirect or Non-Route-Optimized (NRO) path. After the completion of the handover, the MN sends a route optimization message to

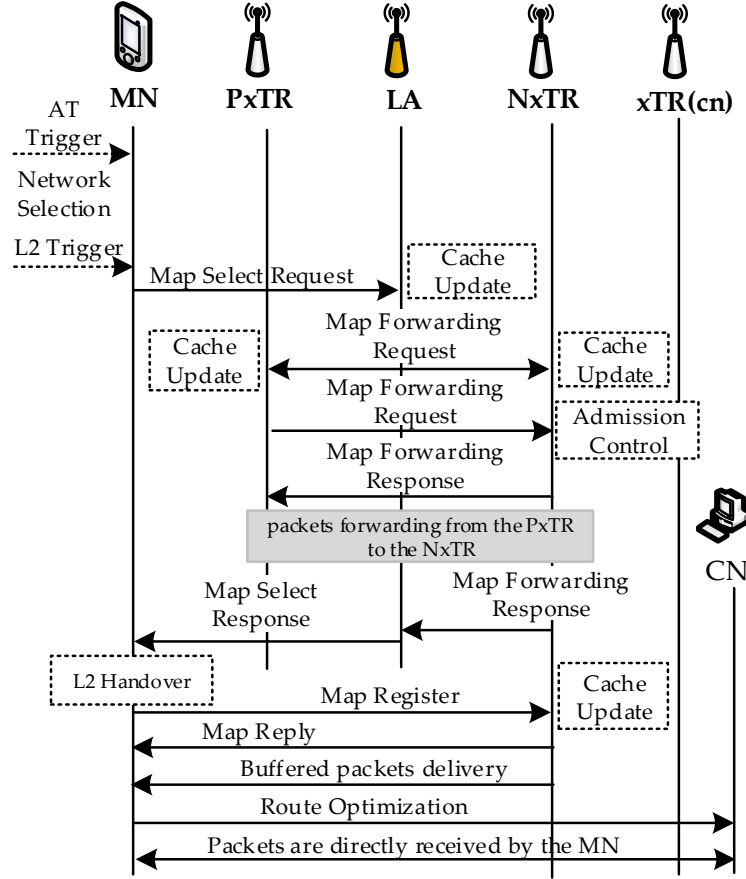


Figure 5.2 Intra-LSA handover

the CN so that subsequent data packets are directly sent to the NxTR following a direct or Route-Optimized (RO) path.

5.3.1.2 Reactive Intra-LSA Handover

The reactive mode is executed whether there is a wrong L2 handover anticipation or when the MN leaves its serving SC without receiving the Map Select Response message from the LA. In this mode, the MN performs a new registration to the network before starting a new communication session with the CNs.

5.3.2 Inter-LSA Handover Scheme

The inter-LSA handover occurs when the PxTR and the NxTR do not belong to the same LSA. In Fig. 6.1, the MN undergoes an inter-LSA handover when it moves along the path 2 detaching from xTR1 to connect to xTR6. Fig. 5.3 describes the inter-LSA handover signaling.

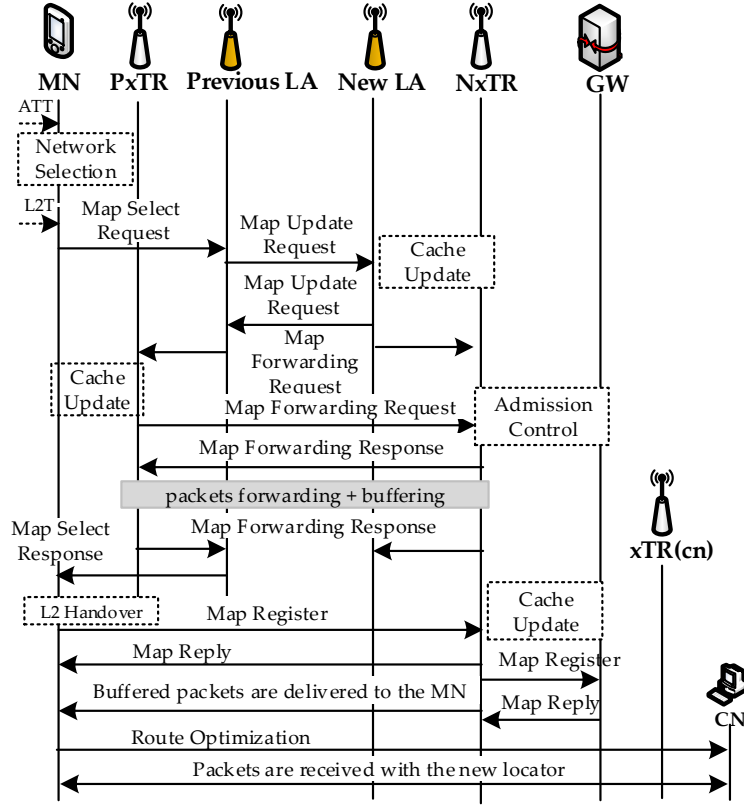


Figure 5.3 Inter-LSA handover

5.3.2.1 Proactive Inter-LSA Handover

Upon L2 triggers generation, the MN detects that it is moving outside of its current LSA. It proactively notify a fast handover movement to the previous LA by sending a Map Select Request message containing its mobility context. The previous LA infers an inter-LSA handover and sends a Map Update message to the new LA. Afterwards, the new LA sends a Map Update Response message back to the previous LA. Both LAs respectively send a Map Forwarding Request message to their respective xTR along with the mobility context. The PxTR establishes a new data forwarding path with the NxTR by sending a Map Forwarding Request message. The NxTR acknowledges the request and sends a Map Forwarding Response message both to the PxTR and to the new LA. The PxTR sends a Map Forwarding Response message back to the previous LA before it start forwarding the handover data packets to the NxTR. The previous LA then responds back to the MN with a Map Select Response message for completing the proactive handover signaling. The NxTR buffers the forwarded packets and later delivers them to the MN when it becomes attached to it. Upon the completion of the handover, the remaining steps are similar to the proactive intra-LSA

signaling.

5.3.2.2 Reactive Inter-LSA Handover

The reactive mode is executed when there is a wrong L2 handover anticipation or when the MN leaves its serving SC without receiving the Map Select Response message from the previous LA. The MN then performs a new registration procedure before starting a new session with its remote peers. In this case, the new LA updates the MN's new locator both at the previous LA and at the GW, in order to reflect its new position in the new LSA.

5.4 Analytical Models

We evaluate the performance of the proposed protocol using the guidelines from the METIS [120], and the 3GPP [138] projects, for ultra-dense networks deployment and femtocell-related performance analysis. The models of this section are used later to obtain the closed-form expressions of several handover metrics. In Table 6.1, we introduce several parameters that we use for the analytical models and the performance analysis.

5.4.1 Network Model

The LSA is represented in Fig. 5.4a as a square-based two-dimensional grid network model following the guidelines from [120] and [138]. We assume that the block centrally located in the grid corresponds to a planned LA. The surrounding blocks, deployed in multi-tiers, correspond to the SCs which have direct or indirect links to the planned LA. It is worth mentioning that, the distance expressed in number of hops between a SC and the LA, impacts the signaling cost during a handover. Each tier i of SCs is labelled in a clockwise direction and contains $8i$ blocs of SCs. For example, in Fig. 5.4a, the topology consists of two tiers or layers ($K = 2$) of SCs around the LA. The first tier is colored in light-green and the second tier is colored in light-blue. The grid like model is used in previous works [122] [131].

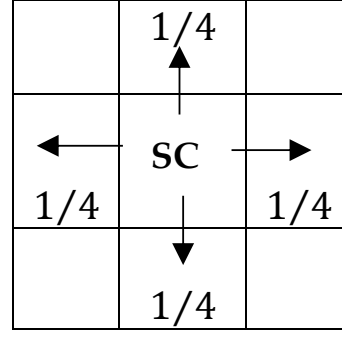
For identifying the presence of the MN with an active session within the area of a SC, state variables A_i^j are used to represent the position of the SC within the LSA. The state variable A_i^j indicates that the MN with an active session is connected to the SC at the position j within the tier i . The LA is denoted by the state variable A_0^0 . Furthermore, we use the state A_{idle} to indicate that the MN has no active sessions independent of the SC to which it is connected to. The next section presents the user mobility and the traffic models which are crucial for an efficient system design and the performance evaluation. [130].

Table 5.1 Notation

Parameters	
i	the position of the SC's tier inside the LSA
K	the maximum number of SCs's tiers in the LSA
j	the position of the SC inside its tier
Δt	the discrete time slot duration
λ	the session arrival rate
μ	the session duration parameter
m	the cell crossing rate
π_i^j	the stationary probability distribution of SC at position j within tier i
$\xi = P_m(1 - P_\mu)$	the probability of handover with an active session
Ψ_i^j	the intra-LSA handover cost function to a SC in any state A_i^j
$\dot{\Psi}_i^j$	the inter-LSA handover cost function to a SC in any state A_i^j
P_s	the probability of handover anticipation
PC_X	the processing cost on the node X
$C_{X,Y}$	the signaling cost between two nodes X and Y
τ	the unit signaling cost over a wired link
s_c (resp. s_d)	the control (resp. the data) packet size
$d_{X,Y}$	the number of hops between two nodes X and Y
κ	the weighting factor of the wireless link
α (resp. β)	the tunneling (resp. dropping) weighting factor
λ_p	the packet arrival rate
t_{L2}	the L2 handover latency
t_{MD}	the IP connectivity latency
t_{AC}	the MN's IP address configuration delay
t_{LA}	the LA's IP address configuration delay
t_{SR}	the packet reception latency with the new IP address
t_{BU}^X	the binding update latency to the node X
t_{LU}	the location update latency
$RTT_{X,Y}$	the round-trip-time between the nodes X and Y
λ_N (resp. λ_H)	the new (resp. the handover) mean session arrival rate
$E(H)$	the average number of handovers per session
P_N	the probability of new traffic reception in the SC
P_H	the probability of handover traffic reception in the SC
t_H	the elapsed time between L2 trigger and L2 link disconnection
t_{FW}	the packet forwarding latency after the L2 trigger
$t_{REG}^{i,X}$	the registration latency to the node X from any SC on tier i
$t_{X,Y}$	the one-way signaling delay between nodes X and Y
q	the wireless link failure probability
s_{RS}	the router solicitation packet size
s_{RA}	the router advertisement packet size
$T_{i,j}^X$	the update latency to the node X from any SC in any state A_i^j
B_w (resp. B_{wl})	wired (resp. wireless) link bandwidth
L_w (resp. L_{wl})	the wired (resp. wireless) link delay
$\bar{\omega}_q$	the average router queueing delay
λ_d	the network densification parameter

A_2^{15}	A_2^{16}	A_2^1	A_2^2	A_2^3
A_2^{14}	A_1^8	A_1^1	A_1^2	A_2^4
A_2^{13}	A_1^7	A_0^0	A_1^3	A_2^5
A_2^{12}	A_1^6	A_1^5	A_1^4	A_2^6
A_2^{11}	A_2^{10}	A_2^9	A_2^8	A_2^7

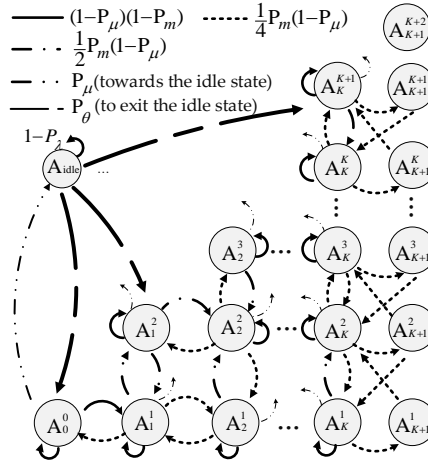
(a) The 2D grid model



(b) The movement direction probabilities

A_2^3	A_2^2	A_2^1	A_2^2	A_2^3
A_2^2	A_1^2	A_1^1	A_1^2	A_2^2
A_2^1	A_1^1	A_0^0	A_1^1	A_2^1
A_2^2	A_1^2	A_1^1	A_1^2	A_2^2
A_2^3	A_2^2	A_2^1	A_2^2	A_2^3

(c) The aggregated 2D grid model



(d) The discrete time Markov Model

Figure 5.4 2D Grid-based network models and the Markov Model

5.4.2 Mobility and Traffic Models

We consider a two-dimensional random walk mobility model [121]. The MN can move from a SC to its neighbors SCs with an equal probability [121] as presented in Fig. B.1. The random walk mobility is shown to be suitable for user movements where the mobility is generally confined to a limited geographical area [122]. Indeed, the two-dimensional random walk mobility model is based on the properties of regular and absorbing Markov chains [123]. Therefore, we model the evolution of the MN's activity as a stochastic process that occurs in a sequence of discrete steps Δt . The MN changes state only at the end of Δt .

We assume that the MN's sessions arrive as a Poisson process [120] [121] with rate λ , each session's duration time follows an exponential distribution with mean rate $1/\mu$, and the SC residence time follows an exponential distribution with mean value $1/m$ [121]. Then, it follows

that $P_\lambda = \lambda\Delta t$, $P_\mu = \mu\Delta t$, and $P_m = m\Delta t$ are respectively the session arrival probability, the session terminating probability, and the MN's movement probability from the current SC during the time slot Δt . When the time slot is sufficiently small, the evolution of the system can reflect real-life characteristics. Let $X(\Delta t)$ be the MN's location within each SC A_i^j at time Δt . $\{X(\Delta t), \Delta t \geq 0\}$ is therefore, a Markov process with discrete time and finite space $E = \{A_{\text{idle}}, A_i^j | 0 \leq i \leq K, 1 \leq j \leq 8i, \forall i \geq 1\}$. When K increases, the resolution of the Markov Chain $X(\Delta t)$ as defined above is time consuming and will result in a state space explosion. To avoid this issue, we make a state aggregation using the symmetric property of the two-dimensional model. The states that exhibit the same properties are therefore aggregated. The aggregated grid model is presented in Fig. 5.4c and the Markov model based on the aggregated states model is presented in Fig. 5.4d.

Let X_r and Y_r be the remaining subnet residence time and the inter-session arrival time respectively. Then, X_r and Y_r are exponentially distributed with rate m and λ respectively due to the memoryless property of an exponentially distributed time. Let θ be the probability that during the intersession arrival time, the MN's locator has not changed in cache. Then we have [132] : $\theta = \text{prob}(Y_r \leq X_r)$ and $\theta = \lambda/(\lambda + m)$. When a session arrives, the MN wakes up from the state A_{idle} into any of the aggregated states A_i^j with probability $P_\theta = 8i\theta/N$, according to the SCs's density in each tier and the total number of cells, N , in the LSA.

It returns from any active state to the state A_{idle} with probability P_μ . The MN stays in state A_{idle} with probability $1 - P_\lambda$. It remains in any state A_i^j with an active session with probability $(1 - P_\mu)(1 - P_m)$. Moreover, the MN undergoes a handover with an active session from any state A_i^j to any state A'^j_i with probability $P(A_i^j, A'^j_i)\xi$. The probability $P(A_i^j, A'^j_i)$ is obtained from the user mobility characteristics and the number of tiers in the LSA. For example, with reference to Fig. 5.4c, $P(A_1^1, A_1^2) = 1/2$ and $P(A_1^1, A_2^1) = 1/4$.

Let π_{idle} be the stationary probability distribution of the MN being idle and π_i^j be the stationary probability distribution that the MN is connected to the SC with position j within tier i with an active session. Using the balance Equations (1)-(15) and the normalization Equation (16) defined in [137], we derive the stationary probability distributions π_i^j and π_{idle} of the Markov model. They are then used to obtain several handover performance metrics.

5.5 Performance Evaluation and Analysis

In this section, the proposed mathematical framework is used to derive the close form expressions of several handover metrics namely, the total signaling cost, the data delivery cost, the handover latency, and the packet loss rate.

5.5.1 Total Signaling Cost

The total signaling cost, C_S^{HO} , is computed as the sum of the average intra-LSA signaling cost (C_{intra}) and the average inter-LSA signaling cost (C_{inter}). Their expressions are given in Equations (5.1) and (6.6) respectively. Indeed, the signaling cost functions Ψ_i^j and $\dot{\Psi}_i^j$ include the packet transmission and processing costs on the network nodes. In IP networks, the packet transmission cost between two nodes is proportional to the distance in hops between the source and the destination [130]. The transmission cost of a control packet between two nodes belonging to the wired and the wireless parts of a network are respectively expressed as : $C_{X,Y} = \tau s_c d_{X,Y}$ and $C_{X,Y} = \tau \kappa s_c d_{X,Y}$.

$$C_{\text{intra}} = \begin{cases} \pi_0^0 \Psi_0^1 \xi + \pi_1^1 \xi \left(\frac{1}{2} \Psi_1^1 + \frac{1}{4} \Psi_1^0 \right) + \frac{1}{2} \pi_1^2 \Psi_1^1 \xi & \text{if } K = 1. \\ \pi_0^0 \Psi_0^1 \xi + \sum_{i=1}^{K-1} \pi_i^1 \xi \left(\frac{1}{4} \Psi_i^{i-1} + \frac{1}{2} \Psi_i^i + \frac{1}{4} \Psi_i^{i+1} \right) & \\ + \sum_{i=2}^{K-1} \sum_{j=2}^i \pi_i^j \xi \left(\frac{1}{4} \Psi_i^{i-1} + \frac{1}{2} \Psi_i^i + \frac{1}{4} \Psi_i^{i+1} \right) & \\ + \sum_{i=1}^{K-1} \pi_i^{i+1} \xi \left(\frac{1}{2} \Psi_i^i + \frac{1}{2} \Psi_i^{i+1} \right) + \sum_{j=1}^K \pi_K^j \xi \left(\frac{1}{4} \Psi_K^{K-1} + \frac{1}{2} \Psi_K^K \right) & \\ + \frac{1}{2} \pi_K^{K+1} \xi \Psi_K^K & \text{if } K \geq 2. \end{cases} \quad (5.1)$$

$$C_{\text{inter}} = \frac{1}{4} \pi_K^1 \dot{\Psi}_K^{K+1} \xi + \sum_{i=2}^K \pi_K^j \xi \frac{1}{4} \dot{\Psi}_K^{K+1} + \frac{1}{2} \pi_K^{K+1} \dot{\Psi}_K^{K+1} \xi. \quad (5.2)$$

5.5.1.1 LISP-MN Scheme

With this scheme, $C_{\text{intra}} = C_{\text{inter}}$. Therefore, $\Psi_i^j = \dot{\Psi}_i^j$ and the signaling cost functions are given in Table 5.2.

5.5.1.2 LISP-LSA Scheme

With this scheme, C_{intra} is expressed as :

$$C_{\text{intra}} = P_s C_{\text{intra}}^{\text{p}} + (1 - P_s) C_{\text{intra}}^{\text{r}}, \quad (5.3)$$

where $C_{\text{intra}}^{\text{p}}$ and $C_{\text{intra}}^{\text{r}}$ are respectively the proactive and the reactive average intra-LSA signaling costs. Indeed, $C_{\text{intra}}^{\text{p}}$ and $C_{\text{intra}}^{\text{r}}$ are computed by replacing Ψ_i^j in Equation (5.1) by the proactive and the reactive signaling cost functions $\Psi_{i,\text{intra}}^{j,\text{p}}$ and $\Psi_{i,\text{intra}}^{j,\text{r}}$ respectively. Their expressions are given in Table 5.2. In the same way, C_{inter} is expressed as :

expressions are given in Table 5.2.

5.5.2 Data Delivery Cost

The data delivery cost, C_D^{HO} , is computed as the linear combination of the data encapsulation cost (C_{enc}) and the data loss cost (C_{loss}) [130]. The transmission cost of a data packet is η greater than the transmission cost of a control packet with $\eta = s_d/s_c$. This cost involves the packet transferring costs through the network and the processing costs on nodes. The data delivery cost is expressed as :

$$C_D^{HO} = \alpha C_{enc} + \beta C_{loss}. \quad (5.5)$$

5.5.2.1 LISP-MN Scheme

The handover timing diagram of the LISP-MN protocol is presented in Fig. 5.5. The data encapsulation cost to the NxTR after a handover is expressed as :

$$C_{enc}^{lispmn} = \lambda_p C_{d,enc} (t_{L2} + t_{MD} + t_{AC} + t_{BU}^{GW} + t_{BU}^{CN} + t_{SR}), \quad (5.6)$$

where $C_{d,enc} = \eta \cdot (C_{CN,NxTR} + C_{NxTR,MN})$ is the cost of transferring the data packets from the CN to the PxTR. Indeed, $t_{SR} = (1/2)RTT_{MN,CN}$. The data loss cost is expressed as :

$$C_{loss}^{lispmn} = \lambda_p C_{d,loss} (t_{L2} + t_{MD} + t_{AC} + t_{BU}^{GW} + t_{BU}^{CN}), \quad (5.7)$$

where $C_{d,loss} = \eta \cdot C_{CN,PxTR}$ is the cost of sending the data packets from the CN to the PxTR. Using the Equations (5.6) and (5.7), the data delivery cost for the LISP-MN protocol is expressed as :

$$C_D^{HO}(\text{LISP-MN}) = \alpha C_{enc}^{lispmn} + \beta C_{loss}^{lispmn}. \quad (5.8)$$

5.5.2.2 LISP-LSA Scheme

The handover timing diagram of the LISP-LSA protocol is presented in Fig. 5.6. Considering the intra-LSA ($C_{D,intra}^{HO}$) and the inter-LSA ($C_{D,inter}^{HO}$) data delivery costs, the total data delivery cost for the LISP-LSA protocol is expressed as :

$$C_D^{HO}(\text{LISP-LSA}) = C_{D,intra}^{HO} + C_{D,inter}^{HO}. \quad (5.9)$$

The expressions of $C_{D,intra}^{HO}$ and $C_{D,inter}^{HO}$ are presented below. The overall traffic arrival rate coming to the MN's current SC after a handover is $\lambda_N + \lambda_H$. The probabilities P_N and P_H

are expressed as follow [150] [151] : $P_N = \lambda_N/(\lambda_N + \lambda_H)$ and $P_H = \lambda_H/(\lambda_N + \lambda_H)$. Moreover, $\lambda_H = \lambda_N E(H)$ [150] [151]. It has been shown in [151] that, regardless of the distribution of the cell residence time and the session duration, $E(H) = m/\mu$. Therefore, the probabilities of new and handover traffic are : $P_N = \mu/(\mu + m)$ and $P_H = m/(\mu + m)$. With the proactive intra-LSA handover scheme, $t_H \geq t_{FW}$, $t_{LA} = 0$, and the data encapsulation cost through the NRO path is :

$$C_{enc, intra}^{p,nro} = \lambda_p C_{d,intra}^{s,nro} (t_{L2} + t_{MD} + t_{BU}^{NxTR} + t_{SR}), \quad (5.10)$$

where $C_{d,intra}^{s,nro} = P_H \cdot \eta \cdot (C_{CN,PxTR} + C_{PxTR,NxTR} + C_{NxTR,MN})$ and $t_{SR} = RTT_{MN,CN}$.

The data encapsulation cost through the RO path during the proactive intra-LSA handover is :

$$C_{enc,intra}^{p,ro} = \lambda_p C_{d,intra}^{s,ro} (t_{L2} + t_{MD} + t_{LU} + t_{SR}), \quad (5.11)$$

where $C_{d,intra}^{s,ro} = P_N \cdot \eta \cdot (C_{CN,NxTR} + C_{NxTR,MN})$. The packet loss cost with the proactive intra-LSA handover scheme is :

$$C_{loss,intra}^{p,nro} = \lambda_p C_{d,intra}^{f,nro} \max(t_{FW} - t_H, 0), \quad (5.12)$$

where $C_{d,intra}^{f,nro} = P_H \cdot \eta \cdot (C_{CN,PxTR} + C_{PxTR,MN})$ is the cost of transferring data packets from the CN to the MN when $t_{FW} \leq t_H$. Using the Equations (5.10), (5.11), and (5.12), the data delivery cost of the proactive intra-LSA handover scheme is :

$$C_{D,intra}^{HO,p} = \alpha (C_{enc, intra}^{p,nro} + C_{enc,intra}^{p,ro}) + \beta C_{loss,intra}^{p,nro}. \quad (5.13)$$

With the reactive intra-LSA handover scheme, the data encapsulation cost is :

$$C_{enc,intra}^{r,ro} = \lambda_p C_{d,intra}^{s,ro} (t_{L2} + t_{MD} + t_{REG}^{i,LA} + t_{SR}). \quad (5.14)$$

The data loss cost is expressed as follows :

$$C_{loss,intra}^{r,nro} = \lambda_p C_{d,intra}^{f,nro} (t_{L2} + t_{MD} + t_{REG}^{i,LA} + t_{SR}), \quad (5.15)$$

where $C_{d,intra}^{f,nro} = P_H \cdot \eta \cdot (C_{CN,PxTR} + C_{PxTR,NxTR})$ is the cost of sending data packets from the CN to a wrong NxTR through the PxTR due to an incorrect handover anticipation.

Using the Equations (5.14) and (5.15), the data delivery cost in the reactive intra-LSA handover mode is expressed as :

$$C_{D,intra}^{HO,r} = \alpha C_{enc, intra}^{r,ro} + \beta C_{loss,intra}^{r,nro}. \quad (5.16)$$

Using the Equations (5.13) and (5.16), the average data delivery cost for the intra-LSA handover scheme is expressed as :

$$C_{D,intra}^{HO} = P_s C_{D,intra}^{HO,p} + (1 - P_s) C_{D,intra}^{HO,r}. \quad (5.17)$$

Following the same reasoning, the proactive inter-LSA data delivery costs are expressed as follows :

$$\begin{aligned} C_{enc,inter}^{p,nro} &= \lambda_p C_{d,intra}^{s,nro} (t_{L2} + t_{MD} + t_{LA} + t_{BU}^{NxTR} + t_{SR}), \\ C_{enc,inter}^{p,ro} &= \lambda_p C_{d,intra}^{s,ro} (t_{L2} + t_{MD} + t_{LA} + t_{BU}^{NxTR} + t_{SR}), \text{ and} \\ C_{loss,inter}^{p,nro} &= \lambda_p C_{d,intra}^{f,nro} \max(t_{FW} - t_H, 0). \end{aligned}$$

The proactive inter-LSA data delivery cost is :

$$C_{D,inter}^{HO,p} = \alpha (C_{enc,inter}^{p,nro} + C_{enc,inter}^{p,ro}) + \beta C_{loss,inter}^{p,nro}. \quad (5.18)$$

In the reactive inter-LSA handover scheme, the data encapsulation and the packet loss costs are given by :

$$\begin{aligned} C_{enc,inter}^{r,ro} &= \lambda_p C_{d,intra}^{s,ro} (t_{L2} + t_{MD} + t_{LA} + t_{SR} + t_{LU}^{prev-LA} + t_{REG}^{i,GW}), \text{ and} \\ C_{loss,inter}^{r,nro} &= \lambda_p C_{d,intra}^{f,nro} (t_{L2} + t_{MD} + t_{LA} + t_{SR} + t_{LU}^{prev-LA} + t_{REG}^{i,GW}). \end{aligned}$$

The data delivery cost for the reactive inter-LSA handover scheme is :

$$C_{D,inter}^{HO,r} = \alpha C_{enc,inter}^{r,ro} + \beta C_{loss,inter}^{r,nro}. \quad (5.19)$$

Using the Equations (5.18) and (5.19), the average data delivery cost for the inter-LSA handover is computed as :

$$C_{D,inter}^{HO} = P_s C_{D,inter}^{HO,p} + (1 - P_s) C_{D,inter}^{HO,r}. \quad (5.20)$$

5.5.3 Handover Latency and Packet Loss

The handover latency, T^{HO} , is the time interval during which the MN can not send or receive any packets during the handover. Indeed, T^{HO} is expressed as the sum of the total transmission delays and the total processing time during the handover. The transmission delay is

Table 5.3 Expressions of the binding update latencies

t_{BU}^{GW}	$=$	$2(t_{MN,NxTR} + t_{NxTR,GW}) + T_{Proc}^{NxTR} + T_{Proc}^{GW}$
t_{BU}^{CN}	$=$	$3(t_{MN,NxTR} + t_{NxTR,xTRcn} + t_{xTRcn,CN} + 3T_{Proc}^{xTR})$
t_{BU}^{NxTR}	$=$	$2t_{MN,NxTR} + T_{Proc}^{NxTR}$
$t_{REG}^{i,LA}$	$=$	$2(t_{MN,NxTR} + t_{NxTR,LA}) + T_{Proc}^{NxTR} + T_{Proc}^{LA}$
$t_{REG}^{i,GW}$	$=$	$2(t_{MN,NxTR} + t_{NxTR,LA} + t_{LA,GW}) + T_{Proc}^{NxTR} + T_{Proc}^{LA} + T_{Proc}^{GW}$
$t_{BU}^{prev-LA}$	$=$	$2t_{new-LA,prev-LA} + T_{Proc}^{LA}$

expressed as follows [152] :

$$t_{X,Y} = c + (d_{X,Y} - 1) \left(s_c / B_w + L_w + \bar{\omega}_q \right), \text{ with :}$$

$$c = \begin{cases} (1 - q) / (1 + q) \cdot (s_c / B_{wl} + L_{wl}), & \text{if } X=MN \\ 0, & \text{otherwise.} \end{cases}$$

5.5.3.1 LISP-MN scheme

The handover latency of the LISP-MN protocol is expressed as :

$$T^{HO} = t_{L2} + t_{MD} + t_{AC} + t_{BU}^{GW} + t_{BU}^{CN}. \quad (5.21)$$

The expressions of the binding update latencies t_{BU}^{GW} and t_{BU}^{CN} are given in Table 6.2.

5.5.3.2 LISP-LSA scheme

The handover latency of the LISP-LSA protocol is expressed as :

$$T^{HO} = T_{intra}^{HO} + T_{inter}^{HO}, \quad (5.22)$$

where T_{intra}^{HO} and T_{inter}^{HO} are respectively the average intra-LSA handover latency and the inter-LSA handover latency. Based on the grid model, T_{intra}^{HO} is expressed as :

$$T_{intra}^{HO} = \sum_{i=1}^{K-1} \sum_{j=i}^{i+1} T_{i,j}^{LA} + \frac{3}{4} \sum_{j=1}^K T_{K,j}^{LA} + \frac{1}{2} T_{K,K+1}^{LA}, \quad (5.23)$$

where $T_{i,j}^{\text{LA}}$ is expressed as [153] :

$$T_{i,j}^{\text{LA}} = \pi_i^j (P_s t_{\text{LA}}^p + (1 - P_s) t_{\text{LA}}^r). \quad (5.24)$$

In Equation (6.12), t_{LA}^p and t_{LA}^r are the proactive intra-LSA handover latency and the reactive intra-LSA handover latency respectively. They are expressed as follows : $t_{\text{LA}}^p = t_{\text{L2}} + t_{\text{MD}} + t_{\text{BU}}^{\text{NxTR}}$ and $t_{\text{LA}}^r = t_{\text{L2}} + t_{\text{MD}} + t_{\text{REG}}^{i,\text{LA}}$. The expressions of the binding update latencies $t_{\text{BU}}^{\text{NxTR}}$ and $t_{\text{REG}}^{i,\text{LA}}$ are given in Table 6.2.

Based on the grid model, $T_{\text{inter}}^{\text{HO}}$ is evaluated as :

$$T_{\text{inter}}^{\text{HO}} = \frac{1}{4} \sum_{j=1}^K T_{K,j}^{\text{GW}} + \frac{1}{2} T_{K,K+1}^{\text{GW}}, \quad (5.25)$$

where $T_{K,j}^{\text{GW}}$ is expressed as [153] :

$$T_{K,j}^{\text{GW}} = \pi_i^j (P_s t_{\text{GW}}^p + (1 - P_s) t_{\text{GW}}^r). \quad (5.26)$$

In Equation (6.14), t_{GW}^p and t_{GW}^r are the proactive inter-LSA handover latency and the reactive inter-LSA handover latency respectively. They are expressed as : $t_{\text{GW}}^p = t_{\text{L2}} + t_{\text{MD}} + t_{\text{LA}} + t_{\text{BU}}^{\text{NxTR}}$ and $t_{\text{GW}}^r = t_{\text{L2}} + t_{\text{MD}} + t_{\text{LA}} + t_{\text{REG}}^{i,\text{GW}} + t_{\text{BU}}^{\text{prev-LA}}$. The expressions of the binding update latencies $t_{\text{REG}}^{i,\text{GW}}$ and $t_{\text{BU}}^{\text{prev-LA}}$ are given in Table 6.2. For both schemes, the average packet loss rate $P_{\text{loss}}^{\text{HO}}$ is proportional to the handover latency T^{HO} and the packet arrival rate λ_p .

5.6 Analytical and Simulation Results

In this section we evaluate the performance of both protocols using the analytical models previously developed. Indeed, we also perform simulations in a more realistic environment than the one described by the analytical models using the network simulator ns-3.26 [154].

5.6.1 Numerical Evaluation

Table B.1 summarizes the default parameters that we use from [128–130, 152] based on the recommendations in [155]. For each handover metric, we define the improvement ratio or gain of the proposed scheme as the ratio of the costs difference over the cost of the LISP-MN scheme under the same parameters. The links are supposed to be full-duplex in term of capacity and delay.

Table 5.4 Default values

Parameters	Value	Parameters	Value	Parameters	Value
λ	0.001/s	t_{L2}	50 ms	λ_p	10 packets/s
μ	0.01/s	B_w	100 Mbps	L_w	2 ms
B_{wl}	11 Mbps	Δt	0.01 s	L_{wl}	10 ms
q	0.5	s_{RS}	52 bytes	s_d	1024 bytes
m	0.1/s	τ	1	κ	1.5
β	0.8	PC_{xTR}	8	PC_{LA}	12
$d_{MN,xTR}$	1	$d_{xTR,CN}$	5	$d_{LA,GW}$	7
T_{Proc}^{xTR}	10 ms	T_{Proc}^{LA}	10 ms	T_{Proc}^{GW}	20 ms
$\bar{\omega}_q$	$50 \cdot 10^{-6}$ s	K	1-6	s_c	80 bytes
s_{RA}	80 bytes	α	0.2	PC_{GW}	24
$d_{xTR,CN}$	7	P_s	0.98	-	-

5.6.2 Simulation Setup

The simulated scenario is presented in Fig. 6.3. It is composed of two LSAs consisting of 50 SCs with two LAs. The SCs are wired to their LA, whereas the LAs are wired to the GW. The MNs are represented by the thick dots and they are randomly deployed in the network. The packets are sent through the intermediate router and the remote access router xTR_{CN} . The distance between two neighboring cells is λ_d . The network becomes denser with small values of λ_d .

The LTE access network is configured on the SCs and the MNs. The latter experience mobility based on the Gauss-Markov mobility model [156]. In this model, also used in [128] and [157], each MN chooses a velocity and a direction to follow at every discrete time step. In its discrete version, the velocity v and the direction d of the MN at any time step $t + 1$, are computed based on the values at time step t as given by :

$$v_{t+1} = \alpha v_t + (1 - \alpha)\bar{v} + \sqrt{(1 - \alpha^2)}s_{v_t},$$

$$d_{t+1} = \alpha d_t + (1 - \alpha)\bar{d} + \sqrt{(1 - \alpha^2)}s_{d_t},$$

where α , $0 \leq \alpha \leq 1$, is the velocity memory factor, \bar{v} and \bar{d} are respectively the asymptotic mean of v_{t+1} and d_{t+1} . Moreover, s_{v_t} and s_{d_t} are independent and uncorrelated stationary Gaussian process with zero mean and standard deviation σ_v and σ_d respectively, where σ_v and σ_d are the asymptotic standard deviation of v_{t+1} and d_{t+1} . The traffic parameters are adopted from the numerical analysis. The network configuration and the simulation parameters are given in Table S1 of the supplemental material. We assume that each MN can attach to any

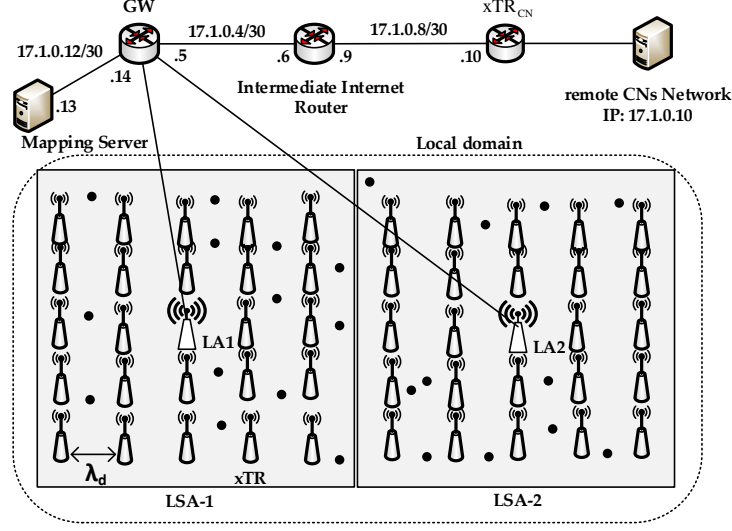


Figure 5.7 Simulated Network in Network Simulator NS3

randomly selected SC to initiate a communication with its CN. The simulation is run for an average of 1500-time steps such that each time step has a duration of 1s. The results from independent simulations trials have been conducted to have a confidence interval of 95%. The evaluated metrics are listed as follows :

- The total signaling cost : it is defined as the total number of signaling packets transmitted by the MNs and the network nodes because of mobility.
- The data delivery cost : it is computed as the linear combination of the data encapsulation cost and the data loss cost during handovers.
- The packet loss : it is the difference between the number of packets sent by the CNs and the one received by the MNs during a session.
- The handover latency : it is the average time that a MN needs to update its new locator in the mapping system after a handover and
- The processing load at the GW : it is the average number of mobility packets processed at the core per time unit because of mobility.

For each metric, we compute the improvement ratio achieved by the proposed LISP-LSA mobility protocol compared to LISP-MN.

5.6.3 Numerical and Simulation Results

5.6.3.1 Numerical Results

Fig. 5.8a plots the total signaling cost as a function of K for different values of m . For both schemes, the total signaling cost increases with K . Moreover, the total signaling cost increases

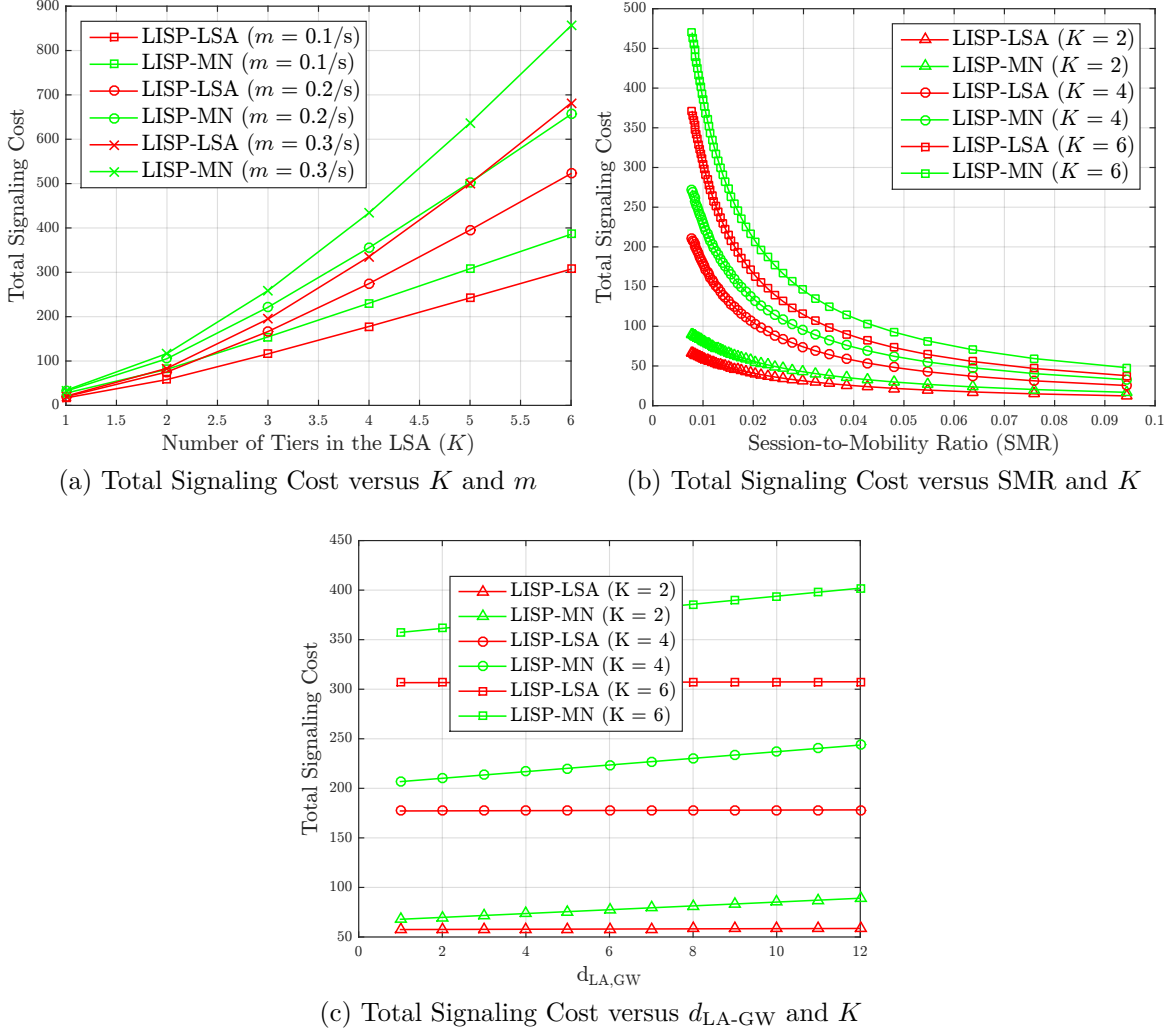


Figure 5.8 Total Signaling Cost versus different parameters

with the mobility rate and the LISP-LSA scheme has a lower signaling load. For example, when $K = 2$, the LISP-LSA scheme reduces the total signaling cost by 30%. However, the cost saving decreases to less than 20% when $K \geq 6$. This is explained by the fact that, a higher mobility rate tend to generate frequent handovers in a denser SCs networks configuration, leading to an increased overall signaling cost. Since the signaling cost is proportional to the distance from the MN to the mobility agent, the LISP-MN scheme induces an increased signaling load because of frequent location updates towards the core of the network, whereas the LISP-LSA scheme reduces the signaling cost because of the proximity of the LA to the network edge. However, because the intra-LSA handover signaling cost for LISP-LSA increases with K , there is a tradeoff to be found between K and m in such a way that the proposed anchoring scheme can always achieve good cost savings.

The cell residence time combined to the MN's activity is an important factor that impacts the performance of mobility management protocols. The session-to-mobility ratio (SMR) represents the relative ratio of the session arrival rate over the user mobility rate. Small values of the SMR indicate that the MN's mobility rate is higher than its session arrival rate.

Fig. 5.8b illustrates the impact of the SMR on the total signaling cost for different values of K . Again, the LISP-LSA scheme outperforms the existing LISP-MN scheme. Smaller values of the SMR means that the MN's movement triggers frequent location updates. We note that the proposed scheme achieves up to 26%, 22% and 20% cost saving respectively for $K = 2$, $K = 4$ and $K = 6$. These improvements are due to several reasons. With the LISP-MN scheme, the MN which is also a mobile LISP site, initiates the binding update with all its CNs after every handovers. This is likely to generate additional signaling load and hinder the scalability of the LISP-MN scheme when frequent handovers occur in denser SCs networks. On the other hand, the LISP-LSA scheme uses the Map Update Option to reduce the binding update signaling during handovers. For higher values of the SMR the total signaling cost is reduced for both protocols.

Fig. 5.8c illustrates the impact of $d_{LA,GW}$, on the total signaling cost. The results show that, LISP-MN scheme generates more signaling load in the network when $d_{LA,GW}$ increases. However, the anchoring scheme LISP-LSA is less sensitive to $d_{LA,GW}$ because the MN updates its position at the LA which is closer to the edge. The results show that the proposed scheme can save up to 64%, 47% and 32% signaling cost respectively for $K = 2$, $K = 4$ and $K = 6$, with $d_{LA,GW} = 10$. This result points out the benefit of having a LA closer to the MN at the edge of the network so that less signaling is generated during handovers towards the core of the network.

Fig. 5.9a plots the data delivery cost as a function of λ_p . The data delivery cost increases proportionally to λ_p for both schemes and the LISP-LSA scheme outperforms the LISP-MN scheme by reducing the delivery cost up to 66%. We explain this by the data forwarding and the buffering mechanisms that the LISP-LSA scheme implements to reduce the packet loss rate during handovers. Since the LISP-MN scheme does not implement any of the aforementioned mechanisms, its data delivery cost is likely to increase when the binding update delay on the remote CNs increases due to long delays in the transit networks. The observed improvement will also be better if the MN maintains several concurrent sessions with CNs.

Fig. 5.9b plots the handover latency against K and m . The results show for both schemes that, the handover delay varies with the mobility rate. However, the LISP-LSA scheme outperforms the LISP-MN scheme and reduces the handover latency up to 54% for $K = 2$. Using its fast handover features, the LISP-LSA scheme starts its handover signaling before the MN leaves its

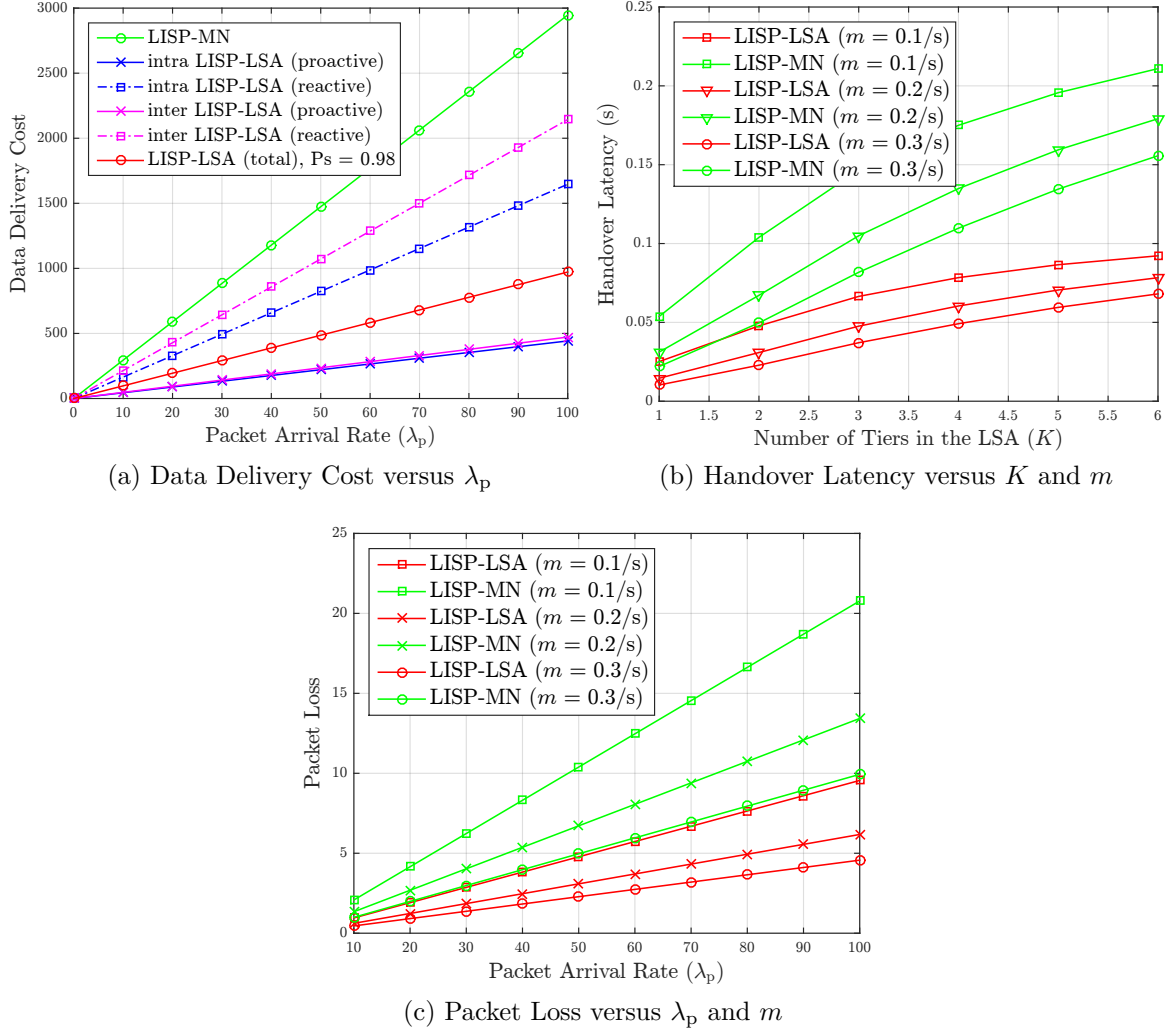


Figure 5.9 Data Delivery Cost, Handover Latency and Packet Loss

previous link on the PxTR, then reducing the handover latency. However, this improvement may be reduced by the quality of the wireless link which directly has an effect on the MN's ability to successfully anticipate the handover.

Fig. 5.9c plots the packet loss against λ_p , for $K = 2$. The packet loss linearly increases with λ_p for both schemes and the LISP-LSA reduces the packet loss up to 54% for $\lambda_p = 10$ packets/s and $m = 0.1/s$. As mentioned above, the absence of any data forwarding, data buffering, and fast handover signaling mechanisms in the LISP-MN scheme, increases both the handover latency and the packet loss rate. Moreover, the binding update latency towards the CNs increases the session recovery time for this scheme inducing a higher packet loss rate at the same time. However, the LISP-LSA also requires an efficient buffering mechanisms.

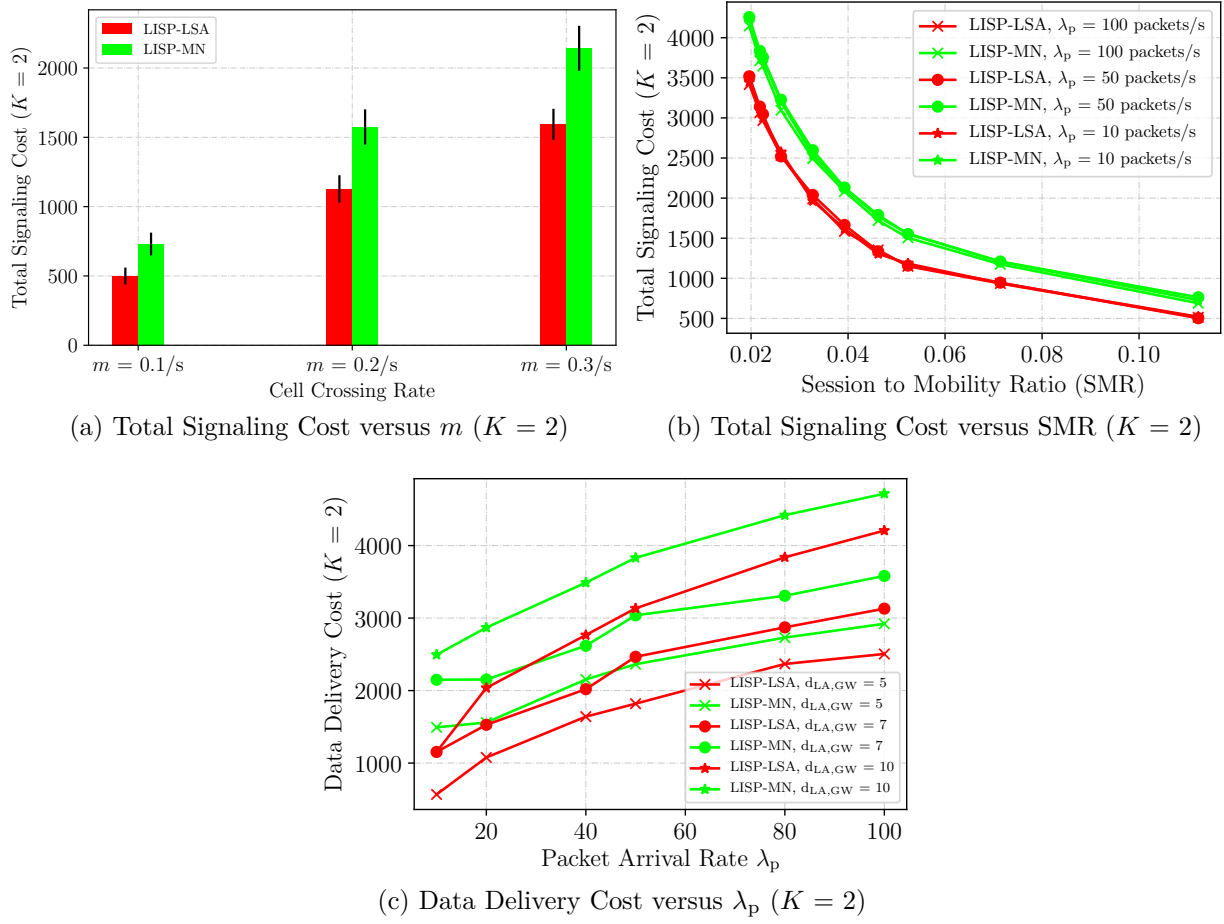


Figure 5.10 Total Signaling Cost and Packet Loss versus different parameters

5.6.3.2 Simulation Results

Fig. 5.10a plots the simulated signaling cost against m . The signaling cost increases with m for both schemes and the LISP-LSA scheme outperforms the LISP-MN scheme by reducing the signaling cost up to 31.6%.

Indeed, Fig. 5.10b shows that the total signaling cost decreases with the SMR for both protocols and the proposed scheme achieves up to 28.5% of cost reduction.

Fig. 5.10c plots the data delivery cost against λ_p and $d_{LA,GW}$. The cost increases with λ_p for both schemes. It is observed that the more distant the MN is from the core network, the higher is the data delivery cost because the distance is likely to increase the amount of lost packets during a handover, resulting to an increased delivery cost. However, the LISP-LSA outperforms the existing protocol with up to 62% of cost reduction.

Fig. 5.11a plots the handover latency of both schemes against the mobility rate. It is observed

that the proposed protocol outperforms the LISP-MN scheme and reduces the handover latency up to 61.5%. This reduction increases when the MN is more distant from the core network. The result shows the performance gain that the LA, which is located at the edge of the network, brings by reducing the handover latency.

Fig. 5.11b plots the packet loss, which is proportional to the handover latency, against λ_p . It is observed that the packet loss linearly increases with λ_p for both schemes and the LISP-LSA scheme outperforms the LISP-MN scheme.

The aforementioned results are consistent with the results from the analytical models and show that the proposed protocol improves the network performance. To further validate our results, we analyse both protocols considering the scalability and the network densification parameters.

Fig. 5.11c plots the packet loss against the MN's velocity and λ_d . When the network becomes denser, the MNs experience frequent handovers with a higher probability of packet loss for ongoing sessions. The proposed protocol achieves over 90% savings in packet loss for the considered density parameters. This improvement is due to the fast handover schemes, the packet forwarding, and the buffering features implemented in the LISP-LSA scheme to improve handover performance.

Fig. 5.12a plots the signaling overhead against the number of MNs and m ($\lambda_d = 20$ m). For low to moderate values of the cell crossing rate ($m = 0.1/s$ and $m = 0.2/s$), it is observed that the proposed scheme reduces the signaling cost by 31%. The maximum signaling cost reduction is around 40%. However, when the cell crossing rate increases, frequent handovers induce a higher signaling overhead for both protocols and the cost reduction drops below 17%.

Fig. 5.12b plots the signaling overhead against the number of MNs and λ_d . The cost reduction is about 30% ($\lambda_d = 20$ m). However, there is no noticeable change in the signaling cost when λ_d increases because the MNs experience less handovers. It is worthwhile noting that, when the number of active MNs is increased in a denser network ($\lambda_d = 15$ m), the signaling cost is increased for both protocols.

Fig. 5.13a plots the processing load at the GW for different values of the MN's velocity. The load gain increases when m increases, because of the higher frequency of handovers. The proposed LISP-LSA scheme reduces the processing load at the core of the network up to 90% with moderate to high values of m . Furthermore, Fig. 5.13b plots the impact of λ_d on the processing load at the GW, by varying the number of active MNs. When the network becomes denser, there is a higher number of handovers. In case of the LISP-MN scheme,

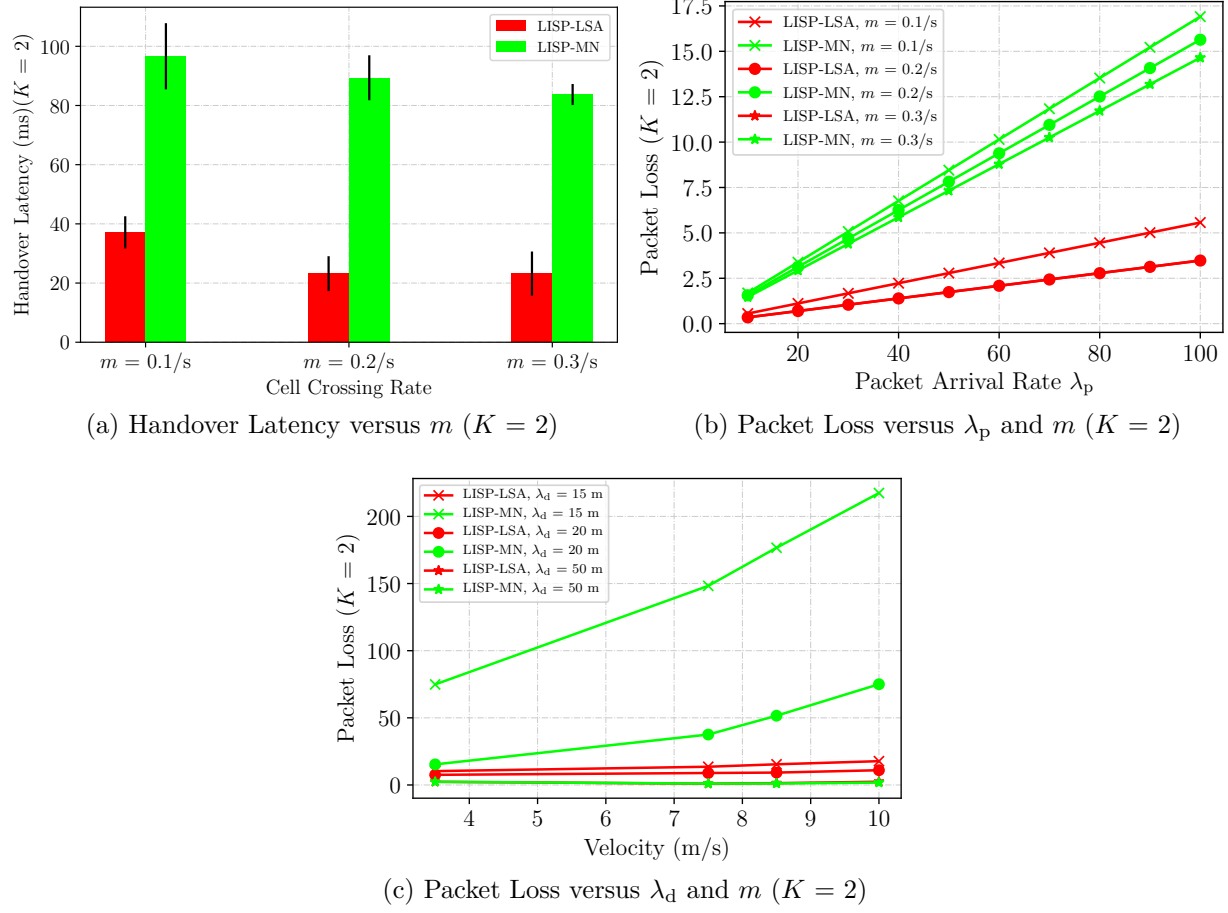


Figure 5.11 Handover Latency and Packet Loss versus different parameters

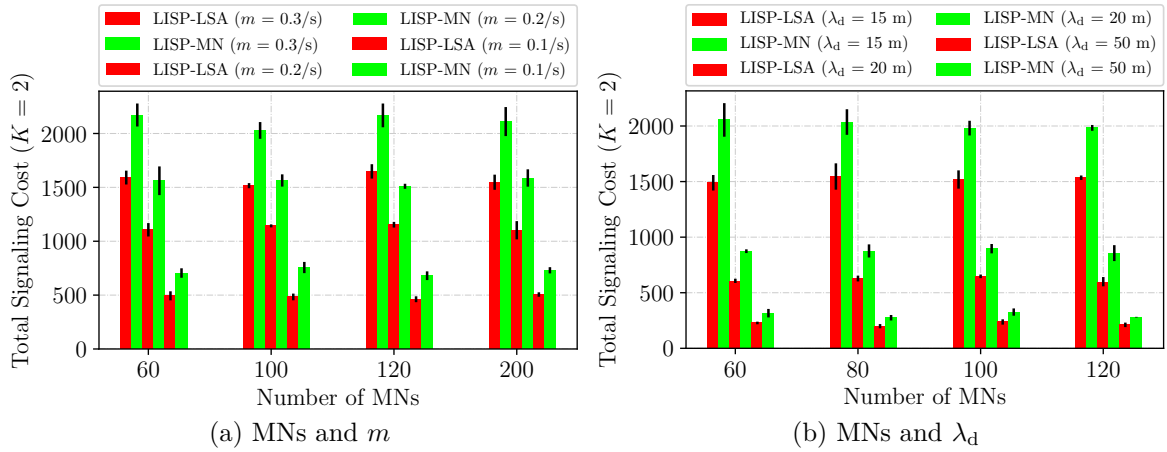


Figure 5.12 Total Signaling Cost versus different parameters

the processing load at the GW also increases. The proposed protocol reduces the processing load at the core of the network up to 84% because the load is shared among the LAs. The proposed LSA based mobility approach reduces the load at the core of the network enabling a better network scalability.

5.6.4 Discussion

Through numerical and simulation results, the proposed LISP-LSA mobility management protocol outperforms the LISP-MN protocol. In a highly dense network, the frequency of handovers is a major issue that impacts the quality of service perceived by the MNs of the network. The observed performance gain of the LISP-LSA scheme is due to the proximity of the LA to the edge of the network thus reducing the handover disconnection time. This approach also reduces the mobility signaling load as well as the packet loss of active sessions. Indeed, the proposed approach also reduces the processing load at the core of the network, thereby enhancing its scalability.

The 5GPP [158] Architecture Working Group has identified network slicing function (NSF) as a key mechanism in 5G networks enablers. The NSF enables the deployment of multitude end-to-end logical, self-contained and independent shared or partitioned networks on a common infrastructure platform. The proposed LISP-LSA scheme, along with its partitioned network configuration aligns with this vision. The physical network infrastructure can be used to share, inside the same LSA or slice, a set of physical resources with virtualized network functions (SCs, network links, LAs) across multiple mobile operators in an isolated manner. Moreover, the inter-slice or the intra-slice control mechanisms will use the intra-LSA and the

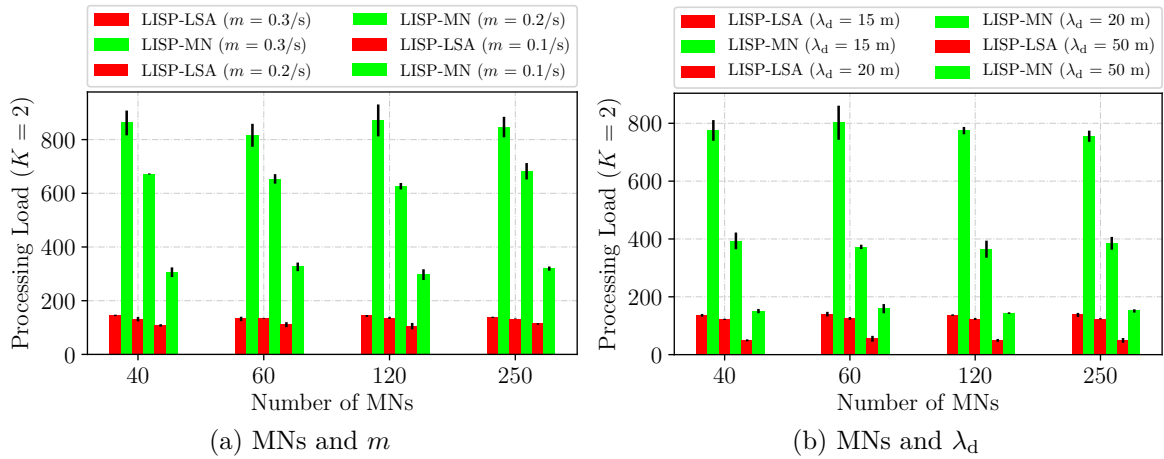


Figure 5.13 Processing Load on the GW versus different parameters

inter-LSA signaling mechanisms for mobility management scenarios.

Nevertheless, in a highly dense network, the MNs will move in the coverage of several SCs. Therefore, it is important to design an efficient admission control algorithm to avoid the exhaustion of available resources on the SCs. Several parameters such as the Signal-to-Noise Ratio (SNR), the transmission power, and the number of active MNs must be considered for an appropriate resource allocation in order to mitigate the handover blocking probability and the outage probability. The design of such algorithms is out of the scope of the proposed mobility management protocol in this work.

5.7 Conclusion

Dense SCs are one of the promising key technologies for 5G networks. They aim to offer better capacity and give more flexibility to operators, by using a spatial densification of low power nodes in the network architecture. The need of a better quality of user experience and seamless handovers in future 5G networks, requires mobility solutions that can take full advantage of the IP infrastructure to support adequate mobility, while maintaining a good level of quality of service for all types of data traffic.

This paper proposed a mobility management protocol with fast handover schemes called LISP-LSA. The proposed mobility protocol uses the ILS mechanism, dense SCs configuration, and a LA. To evaluate the performance of the proposed protocol, we developed a mathematical framework using discrete-time Markov models. We used this framework to obtain the close form expressions of several handover metrics. To further validate the numerical results the protocol was implemented using the network simulator ns-3.26.

The numerical and the simulation results show noticeable performance gains. The LISP-LSA scheme significantly reduces the signaling overhead, the data delivery cost, the packet loss, and the handover latency. Moreover, the proposed protocol also reduces the processing load at the core network thus promoting a good network scalability. However, there is a balance to find between the size of the LSA and the cell crossing rate of the MNs to keep better performance. Moreover, in a highly dense network, it is necessary to design an efficient admission control algorithm to avoid the exhaustion of available resources on the SCs.

CHAPITRE 6 ARTICLE 3 : AN ARCHITECTURE FOR MOBILITY MANAGEMENT IN 5G IDENTIFICATOR LOCATOR SPLIT DENSE SMALL CELLS

Eric Mayeul O. Fafolahan and Samuel Pierre

Mobile Computing and Network Research Laboratory (LARIM)
Department of Computer and Software Engineering, Ecole Polytechnique de Montreal
C.P. 6079, succ. Centre-ville, Montreal, Quebec, H3C 3A7, Canada
Email : {eric.fafolahan, samuel.pierre}@polymtl.ca.

Revue : Soumis dans le journal *IEEE Transactions on Mobile Computing*,
en Mars 2019.

Abstract

Recent network statistics reveal that the number of mobile subscribers has proliferated, and the bandwidth demand for video and data continues to increase. Service providers are looking for innovative approaches to improve their network performance. Network densification with small cells has emerged as a critical enabling technology to increase the network capacity while supporting the design of 5G networks. However, mobility management remains an issue in dense and ultra-dense small cells networks. This work proposes a new architecture for mobility management in 5G Identifier Locator Split dense small cells based on the concept of location service areas. We provide the analytical models of the throughput and the handover blocking probability. The performance of the proposed architecture is evaluated based on several parameters such as the backhaul link latency, the packet arrival rate, the velocity, and the network density. Numerical and simulation results show that the proposed architecture significantly increases the throughput gain by 40 % and reduces the handover blocking probability rate by 55% when the backhaul link latency increases. Indeed, as the velocity increases, the throughput gain is increased by more than 90% while the handover blocking rate is reduced by 70%.

Keywords : 5G, lisp, small cells, throughput, handover blocking probability, performance evaluation.

6.1 Introduction

There is a strong momentum in the global 5G Market, and service providers have announced that they will begin providing 5G services between late 2018 and mid-2019 [2]. 5G is the next phase of mobile technology with high bandwidth, broader coverage, and ultra-low latency [1]. 5G will be able to accommodate new uses cases such as Internet-of-Things (IoT), augmented and virtual reality, smart grids, smart cities, autonomous vehicles, and factory automation [4]. Beyond the connectivity and the advanced user applications deployments, 5G networks will provide new revenue opportunities to service providers when combined with the enhanced network edge capabilities, data analytics, machine learning, and artificial intelligence [1].

A look into recent network statistics reveal that the number of mobile subscribers has increased, and the bandwidth demand for video and data continues to rise. For instance, mobile data traffic will grow at a rate of 46% from 2017 to 2022 [1]. Indeed, the vast majority of mobile data traffic (99%) will originate from smart devices by 2022, up from 92% in 2017 [1]. By 2022, mobile video is expected to account for 79% of the total mobile data traffic. Video traffic in mobile networks is forecast to grow by around 35 percent annually through 2024 [2]. More importantly, a 5G connection will generate 2.6 times more traffic than the average 4G connection by 2022.

Service providers are looking for innovative approaches to improve their network performance as well as to reduce the operation and the maintenance costs [147]. Small Cells (SCs) and ultra-dense network architectures have emerged as one of the fundamental techniques to support the design of the 5G networks [5, 6]. A good proportion of traffic generated by mobile devices is offloaded from the cellular network onto Wi-Fi access points or operator-owned femtocells and picocells [1]. The offloaded percentage on 5G is estimated to be 71% by 2022 [1].

Although network densification with a large number of SCs is expected to bring more capacity, seamless coverage, flexibility and connectivity in future 5G cellular networks, they are still in an initial stage and need further investigation [8]. Moreover, the fast Mobile Nodes (MNs) may experience frequent handovers under SCs, and several parameters owing to network management conditions may also trigger handover decisions [9]. In fact, due to the small and disparate cell sizes, MNs will not be able to complete a handover by the time they trigger a new handover [143]. Consequently, mobility management is still a complex issue to be solved in future 5G dense cellular networks. To address the challenges of mobility management in IP networks, an extensive body of works has appeared in the literature. IP-based mobility schemes are designed to provide IP session continuity and IP reachability to MNs. IP session

continuity allows the user to maintain the ongoing sessions, while its device changes the attached network [159].

Centralized Mobility Management (CMM) schemes such as Mobile IP, Mobile IPv6, and Proxy-Mobile IPv6 rely on a centralized mobility anchor to manage both the mobile data and the MN's mobility context [159]. However, CMM schemes generally introduce a significant network overhead in terms of handover latency, packet loss, and signaling cost when the MNs change their point of attachment very frequently [91]. Distributed Mobility Management (DMM) schemes [98, 148, 149] look for solutions that improve mobility management for future network evolution trends, such as scalability, avoid single points of failure, enable transparent mobility support to upper layers, and provide mobility support just when needed [159]. Instead of having a centralized mobility anchor, DMM schemes distributed the mobility anchors among network entities and placed them as close as possible to the MN [147].

Many authors investigated mobility management in the context of heterogeneous networks. The work in [160] proposed a new mobility management architecture for ultra-dense cellular systems using a Local Access Server (LAS). The LAS is composed of a local service center (LSC) and a local data center (LDC). The LSC collects the topology status of the backhaul network while maintaining the backhaul connections. The LDC manages the data plane. On top of this architecture, two localized mobility management (LMM) schemes are proposed using centralized and distributed mechanisms. The work in [128] introduced a local anchor (LA) based mobility management scheme. The solution defines the LA as the SC maintaining links with a group of neighboring SCs forming a cluster. Compared to existing 3GPP schemes, the proposed approach minimizes the handover interruption as well as the total handover costs. The work in [161] introduced a new architecture, with two mobility mechanisms, to enhance mobility management in LTE-Advanced SCs networks in the context of Vehicular Ad Hoc Networks. A first scheme, which is a network-based approach, uses on a one-hop clustering to efficiently relay traffic from the nearby vehicles towards the serving SC. This work also proposed a k-hops LA-based mobility scheme with three different procedures. The proposed mechanisms are shown to reduce the overall signaling load towards the core network. The authors in [129] proposed two local mobility management in SCs based on the X2 traffic forwarding chain. Instead of switching the data path after each handover, an X2 based traffic forwarding chain is established between the current serving SC and the previous serving SC. A trade-off between the path switch cost and the traffic forwarding is found based on a defined threshold of the forwarding chain. The proposed scheme has shown improved performance compared to the 3GPP standard approach.

The work in [143] proposed a novel mobility management framework that addressed critical

issues including frequent handovers, handovers failure, and excessive energy consumption for seamless handover in the emerging dense cellular networks with mmWave capabilities. Based on the concept of individual mobility model (IMM), the authors in [145] proposed a new mobility model in 5G SCs networks. The IMM considers the impact of human tendency and clustering behaviors. The work in [162] proposed two intelligent handover skipping techniques to decrease the MN's handover rate. The first scheme considers a single base station (BS). Several topological characteristics of the network, namely the cell's area, the trajectory distance, and the distance of the BS from the cell edge are used to decide whether or not to skip a handover. The second technique uses the topological properties which result from BS cooperation. The handover is skipped when the next two consecutive cells along a user's trajectory have a BS in common with the common associated BSs. The work in [163] proposed a joint cell resource allocation algorithm which satisfies both coverage and mobility performance requirements in heterogeneous ultra-dense networks. The proposed algorithm is adaptable to dynamic network changes in parameters, such as traffic distribution, mobility pattern, network topology, and different values of an acceptable signal-to-interference-plus-noise ratio in the network tiers.

It is widely recognized that the current Internet routing system faces serious scaling problems and that the Identifier/Locator Separation (ILS) is a promising solution to solve these problems [164]. The ILS architectures are discussed and pursued at the Internet Engineering Task Force Working Groups to solve the problems of mobility, multihoming, and security in the current Internet architecture [135]. Moreover, the Locator and Identity Split Protocol (LISP) [117] is a mature solution. A host-based mobility protocol called LISP Mobile Node (LISP-MN) is proposed in [139] [126]. With the LISP-MN scheme, each MN maintains a map-cache and implements a lightweight tunnel router functionality [139]. The authors in [140] proposed some improvements to the LISP-MN scheme with the use of a Local Mapping Server in such a way that the MNs would be aware of their locations and add location types in the mapping entries of the mapping system, thus avoiding unnecessary double mapping lookups. The work in [142] proposed a network-based distributed mobility control in LISP networks. The solution provides better performance than the centralized LISP-MN scheme for localized domains in terms of the signaling loads, binding update/query, and handover delays. In [164] the authors use the concept of indirection from PMIPv6. The network is divided into many rendezvous domains (RDs) and assign a rendezvous point to an MN when it moves into an RD. The work in [165] proposed a distributed network architecture to support mobility management in ILS networks. On top of this architecture, this work proposed an effective location management solution.

Being extended from our early work [137], which introduced the concept of the Location

Service Areas (LSAs) into a mobile LISP domain, this work proposes a mobility management architecture in localized ILS dense SCs networks. We investigate the throughput and the handover blocking probability of the proposed LISP-LSA mobility management scheme and compare its performance with the LISP-MN scheme. The main contributions are summarized as follows :

- We propose a localized mobility management architecture for the future 5G dense SCs based on the ILS concept. Moreover, the proposed architecture uses the LISP mechanisms to enable communications between the network entities and the MNs.
- Using the proposed architecture, we provide the analytical expressions of the throughput and the handover blocking probability for the LISP-LSA and the LISP Mobile Node (LISP-MN) mobility schemes.
- We implement the main components of the proposed architecture using the network simulator ns-3.26 to show its feasibility from a practical point of view. Moreover, we define several simulation scenarios to evaluate the performance of the proposed architecture.
- We present numerical and simulation results to show the performance gains of the proposed architecture.

6.2 Background and Motivation

This section presents the background information on the LISP protocol and the motivation of this work.

6.2.1 Background

The LISP protocol is as a promising solution to the emerging 5G networks [136]. LISP introduced the fundamental notion of separating location and identity into two namespaces [166]. When network addresses are used as identities, they are called Endpoint Identifiers (EIDs), and they make up the EID namespace. The nodes in the network are assigned addresses in the Locator namespace and form what is known in LISP as the Routing Locators (RLOCs). The EID is a non-globally routable IP address, and the RLOC is a globally routable IP address attached to routers.

LISP maintains a directory of EIDs and their corresponding locations. In other words, the network routers consult a mapping directory which maps the EID namespace to the RLOC namespace. The MNs send packets to the Ingress Tunnel Router (ITR) using the EIDs. The ITR encapsulates packets with RLOCs and forwards them to an Egress Tunnel Router

(ETR). When receiving encapsulated packets, the ETR decapsulates the packets and sends them to the destination host [165]. A tunnel router (xTR), is a reference to an ITR or ETR when the direction of data traffic is not part of the context description. They are the network devices at the edges of the LISP network. These routers perform the encapsulation and de-encapsulation of EID traffic into RLOC addressed tunnels [166]. The xTRs have both the ITR and the ETR functionalities.

A LISP Proxy ETR (PETR) implements the ETR functions on behalf of non-LISP sites. A PETR is typically used when a LISP site needs to send traffic to non-LISP sites, but the LISP site is connected through a service provider network that does not accept non-routable EIDs as packet sources. A LISP Proxy ITR (PITR) implements the ITR mapping database lookups and LISP encapsulation functions on behalf of non-LISP-capable sites. PITRs are typically deployed in service providers networks to allow non-LISP customers from those networks to connect to LISP sites.

6.2.2 Motivation

Although DMM schemes show better performance compared to CMM schemes, most of them designed the mobility management procedure for one-time handover [34]. Indeed, when the MN's ongoing sessions are long enough, flows may survive through many handovers. Consequently, the DMM schemes which require to set up and maintain multiple levels of tunnels may not be as efficient as expected [34].

Moreover, beyond the single point of failure aspect of CMM schemes, handover performance also depends on the routing operation in such a way that the latency is also affected by the time required to route data packets to the MN's new position.

With the rising of the Internet-of-Things, where billions of mobiles devices will be connected to the network, there is a real need to solve the scalability issue of routing tables at the core of the network. Indeed, the emerging 5G networks are expected to support several emerging applications, which require ultra-low latency, such as the tactile Internet, the virtual reality, and multi-player gaming [5, 6].

While several works have been devoted to mobility management in ILS networks, they may not be suitable for the upcoming 5G dense SCs deployments. In most of the ILS approaches mentioned earlier, the position of the mapping service may hinder the effectiveness of those schemes in 5G dense and ultra-dense networks. It is assumed that the mapping update service is located at the domain Gateway (GW) or a remote Mapping System (MS). Therefore, whenever the MN handovers to a new SC, the new locator must be updated at the remote

node located in the core network. Consequently, the QoS of fast MNs, which will experience frequent handovers, is severely impacted. The resulting high signaling delay will also increase the blocking probability.

Moreover, if the MN's mapping information is cached at multiple remote ITRs, frequent handovers will induce outdated mapping information at the remote ITRs, and this situation will require additional queries to the MS. Therefore, the data packets that are destined to the MN will be sent to the old ETR to which the MN is not connected to [38]. Thus, the old ETR discards several data packets resulting in poor network throughput.

6.3 Proposed Architecture

This section presents the proposed localized mobility management architecture in ILS dense SCs. The proposed architecture is presented in Fig. 6.1. The main features of the architecture are described as follows :

- *Location Service Area based* : the proposed architecture is a localized mobility management solution based on the ILS mechanism. It introduces the concept of Locations Service Area (LSA). The LSA is a conceptual representation of a group of SCs under the control of a network node called a Local Anchor (LA). The LA is used to bring the mobility anchor of the MN as close as possible to the network edge. Fig. 6.1 contains two LSAs in light-blue. Each LSA contains several xTRs which form the underlay data network.
- *Distributed mobility management* : the mobility signaling is handled in a distributed overlay network containing the LAs and the GW. It is important to mention that only the signaling traffic passes through the overlay network.
- *Packet delivery based on the LISP features* : to bring scalability to the core network, the proposed architecture is based on the LISP concepts. Locator-based local and global routing is used for end-to-end communications.
- *Various Mobility Schemes* : the proposed architecture support both intra-LSA and inter-LSA mobility schemes based on the LISP-LSA [167] mobility protocol. The proposed mobility schemes also support proactive and reactive handover schemes.

In the next section, we describe the main components of the proposed architecture.

6.3.1 Network Components

The components of the architectures are described as follows :

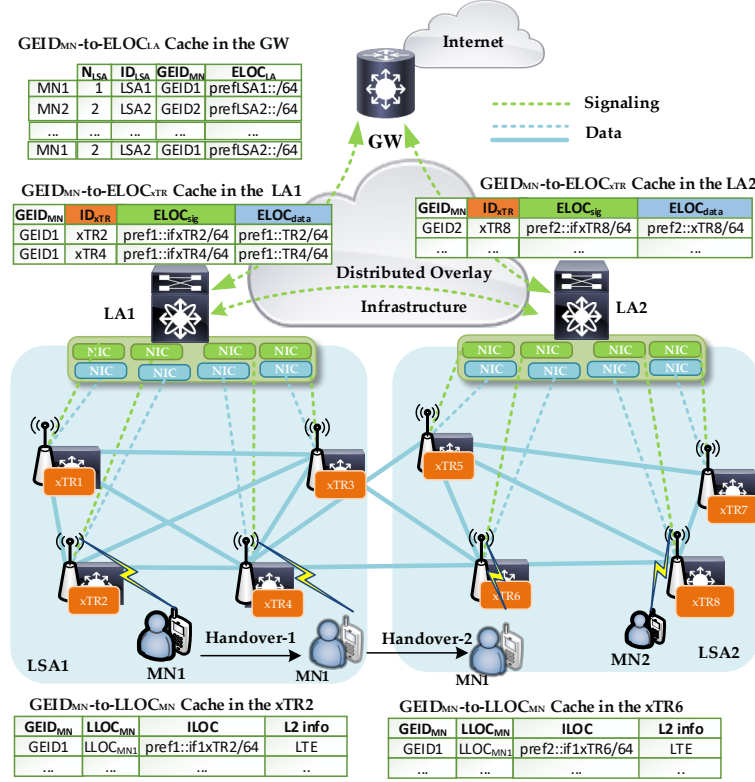


Figure 6.1 Overview of the proposed network architecture and components

6.3.1.1 Mobile Node

The MNs are the users equipped with devices having wireless access to the network. Each MN is identified with a Global Endpoint Identifier (GEID), a Location Service Area LOCator (LOC_{LSA}), and a Local LOCator (LLOC).

The MN's GEID (GEID_{MN}) is invariable, globally unique and represents the identity of the MN in the DNS system. It is used for end-to-end communications with remote peers. The LOC_{LSA} (128-bit) is routable, dynamic and used to identify the LSA where the MN is located. The leftmost 64-bit part indicates the network prefix of the LSA (P_{LSA}) whereas the remaining 64-bit part is the LA's identifier (LA_{ID}). Indeed, the LLOCs (128-bit) are locally routable and dynamic. The leftmost 64-bit part of the LLOC is the P_{LSA}. The following leftmost (32-bit) part identifies the current xTR (ID_{xTR}) to which the MN is attached to. The remaining part denotes the network interface of the MN.

6.3.1.2 Tunnel Router

The xTRs are the wireless access points located in the LSA. They provide wireless access to the MNs attached to them. In order to perform these functionalities, the xTRs maintains two routing caches namely the downlink-routing cache ($\text{GEID}_{\text{MN-to-LLOC}_{\text{MN}}}$) and the uplink-routing cache ($\text{GEID}_{\text{CN-to-ELOC}_{\text{xTR}}}$).

The downlink-routing cache is used to map the MN's identifier to its local locator. When the xTR receives data traffic on the egress interface, they de-encapsulate the packets and send them to the MN using the downlink-routing cache mapping.

The uplink-routing cache is used to map the CN's identifiers to their remote xTRs egress IP address. When the xTR receives data traffic on the ingress interface, using the uplink-routing cache entry, they encapsulate the packet before sending them to the remote CN. The uplink-routing cache is used for the end-to-end data delivery from the MN to the CN.

6.3.1.3 Local Anchor

The LA is an edge access router used for inter-networking between the xTRs and the core network. The architecture introduces the concept of the LSA which is controlled by the LA.

The LA provides several network functionalities such as the MN's registration, the intra-LSA handover binding update, and the resolution service in the LSA. In order to perform these functionalities, the LA maintains two routing caches namely the downlink-routing cache ($\text{GEID}_{\text{MN-to-ELOC}_{\text{xTR}}}$) and the uplink-routing cache ($\text{GEID}_{\text{CN-to-ELOC}_{\text{LA}}}$). The downlink-routing cache is used for mapping the MNs's GEIDs to the egress address of their currently attached xTR. We assume that the egress addresses of the xTRs in the LSA are preconfigured and known by the LA. The uplink-routing cache is used to map the CNs identifiers to the egress addresses of their LAs.

6.3.1.4 Local Gateway

The GW is the network node that is used for the internetworking of the local domain with the external network. The GW provides several network functionalities namely the MNs's registration, the inter-LSA handover binding update, the resolution service, and the routing services. Moreover, the GW also implements the PETR and the PITR functionalities to enable communications with non-LISP sites.

In order to perform these functionalities, the GW maintains a downlink-routing cache ($\text{GEID}_{\text{MN-to-ELOC}_{\text{LA}}}$). This cache is used to map the GEIDs of the MNs in the local domain to the

egress IP address of their current LA.

The signaling traffic from the LAs towards the core network passes through a LISP-based distributed overlay network which connects the LAs to the GW. The overlay network is used both for the lookup resolution service and the binding update services. The resolution service is used for the location discovery service before sending packets to remote CNs. The binding update service is used to update the MN's position during a handover.

6.3.2 Packet Routing

In the proposed architecture, the data packet routing is based on the ILS routing architecture. The GEIDs are used to enable end-to-end data communications. We assume that the MN initiates the session towards its CN.

The MN uses the name resolution service to find the CN's GEID ($GEID_{CN}$) based on the DNS system. Then, it encapsulates the packet using $GEID_{MN}$ and $GEID_{CN}$. The data packet is then sent to its attached xTR using its LLOC. Upon receiving the data packet, the xTR performs a destination lookup to find the CN's position in the network.

6.3.2.1 Destination Lookup

Using the $GEID_{CN}$ from the received packet, the MN's xTR performs a destination lookup using $GEID_{CN}$ to find the CN's xTR egress locator ($ELOC_{CN}$). It checks its local caches in order to determine whether an entry based on $GEID_{CN}$ exists.

Case 1 : if the CN and the MN are attached to the same xTR, an entry exists in the xTR's downlink-routing cache. An entry may also exist in the xTR's uplink-routing cache based on previous communications. In either situations, the $ELOC_{CN}$ is also locally available.

Case 2 : if the xTR cannot find any entry in its local caches, it sends a Map-Request message to its LA along with $GEID_{CN}$. When the LA finds an entry related to the CN in its the downlink-routing cache, it sends a Map-Response message back to the MN's xTR with the $ELOC_{CN}$. If the LA finds an entry related to the CN in its uplink-routing cache, it forwards the Map-Request message to the CN's current LA. The latter responds with a Map-Response message containing the $ELOC_{CN}$. Afterward, the LA delivers the response to the MN's xTR.

Case 3 : if the LA cannot find any cache entry related to the CN, the Map-Request message is forwarded to the GW. If the CN's mapping information exists in the GW's downlink-routing cache, the GW transmits the Map-Request message to the CN's current LA. The latter responds to the MN's LA with a Map-Response message containing $ELOC_{CN}$.

Case 4 : if the CN is not located in the local domain, the data packet needs to be routed to the external network. The GW sends a Map-Response message back to the LA with a nondefault destination in the local domain. Then, the LA forwards the response message to the xTR. Upon receiving this message, the xTR transmits the packet to the GW for global routing.

Depending on the CN's position in the network, the data packet routing may be local or global.

6.3.2.2 Data routing

The local routing is used when the CN and the MN are located in the same local domain. Indeed, when the CN and the MN are attached to the same xTR, the packet is forwarded to the CN without any further processing. Otherwise, after the destination lookup step, the MN's xTR encapsulates the data packet using its egress locator as source and $ELOC_{CN}$ as the destination. The packet is routed to the CN's xTR through the data network. The CN's xTR decapsulates the received packet and delivers it to the CN.

The global routing is performed when the CN and the MN are not located in the same domain. From the DNS record, the GW knows whether the CN's site supports ILS-based communications. When the CN's site supports LISP, the GW encapsulates the data packet with its egress locator as source and the egress locator of the remote site as the destination. The data packet is delivered using the LISP routing architecture. When the CN's remote site do not support LISP, the GW removes the header that was previously added by the MN. The packet is routed natively to the destination using the PETR functionality on the GW. The return traffic is handled with the PITR functionality of the GW.

6.3.3 Mobility Management

The proposed architecture supports the LISP-LSA [137] protocol for mobility management. During a handover, the MN detaches from its serving SC or Previous xTR (PxTR) and attaches to a new SC or New xTR (NxTR). Moreover, a mobility context containing the handover information is exchanged between the network nodes involved in a handover. The LISP-LSA supports both the intra-LSA and the inter-LSA handover management procedures.

The intra-LSA handover occurs when the PxTR and the NxTR belong to the same LSA. In Fig. 6.1, the MN undergoes an intra-LSA handover when it detaches from xTR2 to connect to xTR4 in LSA1. The inter-LSA handover occurs when the PxTR and the NxTR do not belong to the same LSA. In Fig. 6.1, the MN undergoes an inter-LSA handover when it

detaches from xTR4 to connect to xTR6. We refer the reader to our previous work in [167] for more details about the LISP-LSA protocol.

6.4 Analytical Models

This section presents our assumptions and the network model that we use. We base our analysis on the guidelines from the METIS [120], and the 3GPP [138] projects, for ultra-dense networks deployment and femtocell-related performance analysis.

6.4.1 Assumptions

Without loss of generality, we make the following assumptions :

- The MN's sessions arrive as a Poisson process [120] [121] with rate λ , each session's duration time follows an exponential distribution with mean rate $1/\mu$.
- t_r is the random variable defining the SC residence time with its probability density function $f_R(t)$. t_r follows an exponential distribution with mean value $1/m$ [121].
- t_s is the random variable defining the handover signaling delay of the LISP-MN protocol. The density function of t_s is $f_s(t) = \theta e^{-\theta t}$ and \bar{T}_S is the mean of t_s .
- t_s^{intra} is the random variable defining the intra-LSA handover signaling delay of the LISP-LSA scheme and \bar{T}_S^{intra} is the mean of t_s^{intra} .
- t_s^{inter} is the random variable defining the inter-LSA handover signaling delay of the LISP-LSA scheme and \bar{T}_S^{inter} is the mean of t_s^{inter} .

t_s , t_s^{intra} , and t_s^{inter} are exponentially distributed with cumulative density function $F_T(t)$, $F_T^{\text{intra}}(t)$, $F_T^{\text{inter}}(t)$ respectively. In Table 6.1, we introduce several parameters that we use for the performance analysis.

6.4.2 Network Model

We adopt the proposed network model from [167]. The network model is presented as follows :

- We represent the LSA as a square-based two-dimensional grid network model based on the guidelines from [120] and [138]. We assume that the block centrally located in the grid corresponds to a planned LA. The grid consists of K tiers of SCs around the LA.

This grid is used as a reference model for the Ultra Dense Small cells networks in 5G networks [120]. This scenario represents an indoor environment and contains several blocks of SCs in a low isolation scenario (LowIS). The scenario assumes that the SCs are deployed at the center of each bloc.

Table 6.1 Notation

Parameters	
i	the position of the SC's tier inside the LSA
K	the maximum number of SCs's tiers in the LSA
j	the position of the SC inside its tier
Δt	the discrete time slot duration
λ	the session arrival rate
μ	the session duration parameter
m	the cell crossing rate
π_i^j	the stationary probability distribution of SC at position j within tier i
E_S	the number of packets per sessions
σ	the interpacket transmission time
η	the number of handovers per sessions
P_s	the probability of handover anticipation
PC_X	the processing cost on the node X
$d_{X,Y}$	the end-to-end data delivery time between two nodes X and Y
s_d	the data packet size
s_t	the tunnel header size
$h_{X,Y}$	the number of hops between two nodes X and Y
α (resp. β)	the intra-LSA (resp. the inter-LSA) handover probability
λ_p	the packet arrival rate
t_{L2}	the L2 handover latency
t_{BU}^X	the binding update latency to the node X
P_N	the probability of new traffic reception in the SC
P_H	the probability of handover traffic reception in the SC
$t_{REG}^{i,X}$	the registration latency to the node X from any SC on tier i
$t_{X,Y}$	the one-way signaling delay between nodes X and Y
q	the wireless link failure probability
$T_{i,j}^X$	the update latency to the node X from any SC in any state A_i^j
B_w (resp. B_{wl})	the wired (resp. wireless) link bandwidth
L_w (resp. L_{wl})	the wired (resp. wireless) link delay
$\bar{\omega}_q$	the average router queueing delay
λ_d	the network densification parameter
p	the mobility protocol (LISP-MN or LISP-LSA)
$N_{rx}^{pkt}(p)$	the number of successfully received packets with p
$T(p)$	the total data transmission time of p
$D(p)$	the end-to-end data delivery time of p
$L_{HO}(p)$	the total session disruption time of p .
$Pr_{block}(p)$	the handover blocking probability of protocol p
t_{LA}	the LA's IP address configuration delay
t_{AC}	the MN's IP address configuration delay
t_{MD}	the IP connectivity latency

- The presence of the MN with an active session within the coverage of the SC is represented with a state variable A_i^j . This variable indicates that the MN with an active session is connected to the SC at the position j within the tier i in the LSA. The LA is denoted by the state variable A_0^0 . The state A_{idle} indicates that the MN has no active sessions regardless of its current attached SC.
- The MNs are randomly deployed in the LSA. We consider a two-dimensional random walk mobility model [121]. The MNs move from the SC to its neighbors SCs with an equal probability [121] as presented in Fig. S1 of the supplemental material.
- The evolution of the MN's activity is modelled as a stochastic process that occurs in a sequence of discrete steps Δt [167]. The MN changes state only at the end of Δt . Let $X(\Delta t)$ be the MN's location within each SC A_i^j at time Δt . $\{X(\Delta t), \Delta t \geq 0\}$ is therefore, a Markov process with discrete time and finite space $E = \{A_{\text{idle}}, A_i^j | 0 \leq i \leq K, 1 \leq j \leq 8i, \forall i \geq 1\}$.

6.5 Performance Evaluation and analysis

This section presents the closed-form expressions of the throughput and the handover blocking probability models for both the LISP-LSA and the LISP-MN protocols.

6.5.1 Throughput Model

We consider a session model where E_S data packets are sent from the CN to the MN. The throughput, $Thr(p)$, of the mobility protocol p is defined as the ratio of the number of successfully received data packets during a session over the total session duration time. It is expressed as follows :

$$Thr(p) = \frac{N_{\text{rx}}^{\text{pkt}}(p)}{T(p) + D(p) + L_{\text{HO}}(p)}. \quad (6.1)$$

6.5.1.1 Received Data Packets

The successfully received data volume for the LISP-MN protocol, $N_{\text{rx}}^{\text{pkt}}(\text{LISP-MN})$, is expressed as follows :

$$N_{\text{rx}}^{\text{pkt}}(\text{LISP-MN}) = E_S - \lambda_p \cdot L_{\text{HO}}(\text{LISP-MN}). \quad (6.2)$$

In the same way, $N_{\text{rx}}^{\text{pkt}}(\text{LISP-LSA})$, is expressed as follows :

$$N_{\text{rx}}^{\text{pkt}}(\text{LISP-LSA}) = E_S - \lambda_p \cdot L_{\text{HO}}(\text{LISP-LSA}). \quad (6.3)$$

6.5.1.2 Data Transmission Time

For both protocols, $T(p)$ is expressed as follows :

$$T(p) = \sigma \cdot (E_S - 1). \quad (6.4)$$

6.5.1.3 Data Delivery Time

The end-to-end data delivery time is computed by summing the packet delivery time among networks entities. If a packet of size s_d is transmitted between two nodes X and Y via both the wired and wireless links distanced by $h_{X,Y}$ hops, $d_{X,Y}$ is expressed as [152] :

$$d_{X,Y} = c + (d_{X,Y} - 1) \left(s_d / B_w + L_w + \bar{\omega}_q \right), \text{ with :}$$

$$c = \begin{cases} (1 - q) / (1 + q) \cdot (s_d / B_{wl} + L_{wl}), & \text{if } X=MN \\ 0, & \text{otherwise.} \end{cases}$$

With the LISP-MN scheme, $D(\text{LISP-MN})$ is expressed as :

$$D(\text{LISP-MN}) = E_S \cdot \left[(d_{\text{CN},\text{xTRcn}} + \text{PC}_{\text{xTRcn}}) + (d_{\text{xTRcn},\text{GW}} + \text{PC}_{\text{GW}}) + (d_{\text{GW},\text{NxTR}} + \text{PC}_{\text{xTR}}) + d_{\text{NxTR},\text{MN}} \right]. \quad (6.5)$$

The expressions of $d_{\text{CN},\text{xTRcn}}$, $d_{\text{xTRcn},\text{GW}}$, $d_{\text{GW},\text{NxTR}}$, and $d_{\text{NxTR},\text{MN}}$ are given in Table 6.3. For instance, in Equation (6.5), the data packets are tunnelled from the CN's tunnel router to the GW and the additional tunnelling header is considered in the expression of $d_{\text{xTRcn},\text{GW}}$.

With the LISP-LSA scheme, $D(\text{LISP-LSA})$ depends on whether the data packets are directly routed or indirectly routed from the CN to the MN.

The forwarded packets through the PxTR during the handover follow the indirect or Non-Route-Optimized (NRO) path. After the completion of the handover, the MN sends a route optimization message to the CN so that subsequent data packets are directly sent to the NxTR following a direct or Route-Optimized (RO) path.

Therefore, $D(\text{LISP-LSA})$ is expressed as follows :

$$D(\text{LISP-LSA}) = E_S \cdot \left(P_H \cdot D_{\text{LISP-LSA}}^{\text{nro}} + P_N \cdot D_{\text{LISP-LSA}}^{\text{ro}} \right), \quad (6.6)$$

where $D_{\text{LISP-LSA}}^{\text{nro}}$ and $D_{\text{LISP-LSA}}^{\text{ro}}$ are respectively the data delivery time through the indirect

and the direct data path. They are expressed as follows :

$$\begin{aligned}
 D_{\text{LISP-LSA}}^{\text{nfo}} = & (d_{\text{CN},\text{xTRcn}} + \text{PC}_{\text{xTRcn}}) + (d_{\text{xTRcn},\text{GW}} + \text{PC}_{\text{GW}}) \\
 & + (d_{\text{GW},\text{PxTR}} + \text{PC}_{\text{xTR}}) + (d_{\text{PxTR},\text{NxTR}} \\
 & + \text{PC}_{\text{xTR}}) + d_{\text{NxTR},\text{MN}}
 \end{aligned} \tag{6.7}$$

$$\begin{aligned}
 D_{\text{LISP-LSA}}^{\text{ro}} = & (d_{\text{CN},\text{xTRcn}} + \text{PC}_{\text{xTRcn}}) + (d_{\text{xTRcn},\text{GW}} + \text{PC}_{\text{GW}}) \\
 & + (d_{\text{GW},\text{NxTR}} + \text{PC}_{\text{xTR}}) + d_{\text{NxTR},\text{MN}}
 \end{aligned} \tag{6.8}$$

The expression of $d_{\text{PxTR},\text{NxTR}}$ is given in Table 6.3.

6.5.1.4 Session Disruption Time

In Equation (6.2), $L_{\text{HO}}(\text{LISP-MN})$ is proportional to η . It is expressed as $L_{\text{HO}}(\text{LISP-MN}) = \eta \cdot S(\text{LISP-MN})$, where $S(\text{LISP-MN})$ is the handover disruption time of the LISP-MN scheme.

After a handover, regardless of its position, the MN updates its new locator at the GW. Then, it starts the binding update with the CNs. Hence, $S(\text{LISP-MN})$ is expressed as :

$$S(\text{LISP-MN}) = t_{\text{L2}} + t_{\text{MD}} + t_{\text{AC}} + t_{\text{BU}}^{\text{GW}} + t_{\text{BU}}^{\text{CN}}. \tag{6.9}$$

The expressions of the binding update latencies $t_{\text{BU}}^{\text{GW}}$ and $t_{\text{BU}}^{\text{CN}}$ are given in Table 6.2.

In Equation (6.3), $L_{\text{HO}}(\text{LISP-LSA})$ is also proportional to η . It is expressed as $L_{\text{HO}}(\text{LISP-LSA}) = \eta \cdot S(\text{LISP-LSA})$, where $S(\text{LISP-LSA})$ is the handover disruption time of the LISP-LSA scheme.

$S(\text{LISP-LSA})$ is evaluated by summing the intra-LSA ($T_{\text{intra}}^{\text{HO}}$) and the inter-LSA ($T_{\text{inter}}^{\text{HO}}$) handover latencies of the LISP-LSA scheme. It is expressed as [167] :

$$S(\text{LISP-LSA}) = T_{\text{intra}}^{\text{HO}} + T_{\text{inter}}^{\text{HO}}. \tag{6.10}$$

Using the grid model, $T_{\text{intra}}^{\text{HO}}$ is expressed as :

$$T_{\text{intra}}^{\text{HO}} = \sum_{i=1}^{K-1} \sum_{j=i}^{i+1} T_{i,j}^{\text{LA}} + \frac{3}{4} \sum_{j=1}^K T_{K,j}^{\text{LA}} + \frac{1}{2} T_{K,K+1}^{\text{LA}}, \tag{6.11}$$

where $T_{i,j}^{LA}$ is expressed as [153] :

$$T_{i,j}^{LA} = \pi_i^j (P_s t_{LA}^p + (1 - P_s) t_{LA}^r). \quad (6.12)$$

In Equation (6.12), t_{LA}^p and t_{LA}^r are the proactive intra-LSA handover latency and the reactive intra-LSA handover latency respectively. They are expressed as follows : $t_{LA}^p = t_{L2} + t_{MD} + t_{BU}^{NxTR}$ and $t_{LA}^r = t_{L2} + t_{MD} + t_{REG}^{i,LA}$.

The expressions of the binding update latencies t_{BU}^{NxTR} and $t_{REG}^{i,LA}$ are given in Table 6.2.

Using the grid model, T_{inter}^{HO} is evaluated as [167] :

$$T_{inter}^{HO} = \frac{1}{4} \sum_{j=1}^K T_{K,j}^{GW} + \frac{1}{2} T_{K,K+1}^{GW}, \quad (6.13)$$

where $T_{K,j}^{GW}$ is expressed as [153] :

$$T_{K,j}^{GW} = \pi_i^j (P_s t_{GW}^p + (1 - P_s) t_{GW}^r). \quad (6.14)$$

In Equation (6.14), t_{GW}^p and t_{GW}^r are the proactive inter-LSA handover latency and the reactive inter-LSA handover latency respectively. They are expressed as : $t_{GW}^p = t_{L2} + t_{MD} + t_{LA} + t_{BU}^{NxTR}$ and $t_{GW}^r = t_{L2} + t_{MD} + t_{LA} + t_{REG}^{i,GW} + t_{BU}^{prev-LA}$.

The expressions of the binding update latencies $t_{REG}^{i,GW}$ and $t_{BU}^{prev-LA}$ are given in Table 6.2.

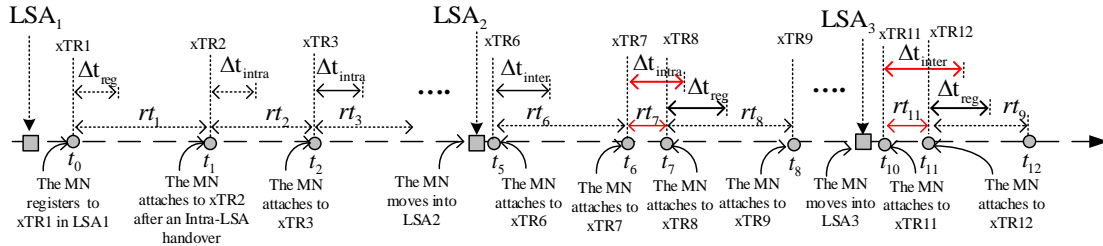


Figure 6.2 The model used for calculating the handover blocking probability

6.5.2 Handover Blocking Probability

The model that we consider for calculating the handover blocking probability is presented in Fig. 6.2.

At t_0 the MN attaches to the network and registers on xTR1 in the LSA₁. After being attached to xTR1 during a time interval denoted by rt_1 , it moves to xTR2 and performs the

Table 6.2 Expressions of the binding update latencies

t_{BU}^{GW}	$=$	$2(t_{MN,NxTR} + t_{NxTR,GW}) + T_{Proc}^{NxTR} + T_{Proc}^{GW}$
t_{BU}^{CN}	$=$	$3(t_{MN,NxTR} + t_{NxTR,xTRcn} + t_{xTRcn,CN} + 3T_{Proc}^{xTR})$
t_{BU}^{NxTR}	$=$	$2t_{MN,NxTR} + T_{Proc}^{NxTR}$
$t_{REG}^{i,LA}$	$=$	$2(t_{MN,NxTR} + t_{NxTR,LA}) + T_{Proc}^{NxTR} + T_{Proc}^{LA}$
$t_{REG}^{i,GW}$	$=$	$2(t_{MN,NxTR} + t_{NxTR,LA} + t_{LA,GW}) + T_{Proc}^{NxTR} + T_{Proc}^{LA} + T_{Proc}^{GW}$
$t_{BU}^{prev-LA}$	$=$	$2t_{new-LA,prev-LA} + T_{Proc}^{LA}$

Table 6.3 Expressions of the packet delivery time

$d_{CN,xTRcn}$	$=$	$(s_d/B_w) + L_w + \bar{w}_q$
$d_{xTRcn,GW}$	$=$	$(h_{xTRcn,GW} - 1) \cdot [L_w + \bar{w}_q + (s_d + s_t)/B_w]$
$d_{GW,NxTR}$	$=$	$(h_{GW,NxTR} - 1) \cdot [L_w + \bar{w}_q + (s_d + s_t)/B_w]$
$d_{NxTR,MN}$	$=$	$((1 - q)/(1 + q)) \cdot (L_w + s_d/B_w)$
$d_{PxTR,NxTR}$	$=$	$(h_{PxTR,NxTR} - 1) \cdot [L_w + \bar{w}_q + (s_d + 2 \cdot s_t)/B_w]$

intra-LSA handover signaling during a time interval of Δt_{intra} . After staying connected to xTR2 during a time interval of rt_2 , it handovers to xTR3.

This process continues and at t_6 the MN attaches to xTR7 in the LSA₂. However, shortly after starting the intra-LSA signaling, the MN moves to a new SC (xTR8) before the signaling completion ($rt_7 < \Delta t_{intra}$). Consequently, the handover is blocked and the MN has to perform a new registration to xTR8.

Later on, the MN enters the LSA₃ and attaches to xTR11 at t_{10} . However, before the inter-LSA handover signaling completion ($rt_{11} < \Delta t_{inter}$), the MN handovers to a new SC (xTR12) at t_{11} and the handover is also blocked.

From the above description, we define the handover blocking probability $Pr_{block}(p)$ of protocol p as the probability of the MN's cell residence time to be less than the handover signaling time. Consequently, the ongoing communications are forcefully terminated because of the loss of the link information or wireless channel.

For the LISP-MN scheme,

$$Pr_{block}(LISP-MN) = P(t_s > t_r) = 1 - P(t_s \leq t_r) \quad (6.15)$$

$$\begin{aligned}
P(t_s \leq t_r) &= \int_0^\infty P(t_s \leq t) f_R(t) dt \\
&= \int_0^\infty \left[\int_0^t f_s(\tau) d\tau \right] f_R(t) dt \\
&= \int_0^\infty \left[F_T(t) - F_T(0) \right] f_R(t) dt
\end{aligned} \tag{6.16}$$

Since, $F_T(0) = 1 - e^{-\theta 0} = 0$, and $\theta = 1/\bar{T}_S$, then :

$$\begin{aligned}
P(t_s > t_r) &= 1 - \int_0^\infty \left(1 - e^{-\theta t} \right) f_R(t) dt \\
&= \int_0^\infty e^{-\theta t} m e^{-mt} dt = \frac{m}{\theta + m}
\end{aligned} \tag{6.17}$$

$$Pr_{\text{block}}(\text{LISP-MN}) = \frac{m \cdot \bar{T}_S}{1 + m \cdot \bar{T}_S} \tag{6.18}$$

For the LISP-LSA scheme, $Pr_{\text{block}}(\text{LISP-LSA})$ is expressed as follows :

$$Pr_{\text{block}}(\text{LISP-LSA}) = \alpha \cdot Pr_{\text{block}}^{\text{intra}} + \beta \cdot Pr_{\text{block}}^{\text{inter}}, \tag{6.19}$$

where $Pr_{\text{block}}^{\text{intra}}$ and $Pr_{\text{block}}^{\text{inter}}$ are respectively the intra-LSA and the inter-LSA handover blocking probabilities. Following the same steps from Equations (16), (17), and (18), they are expressed as follows :

$$Pr_{\text{block}}^{\text{intra}} = P(t_s^{\text{intra}} > t_r) = 1 - P(t_s^{\text{intra}} \leq t_r). \tag{6.20}$$

$$P(t_s^{\text{intra}} > t_r) = 1 - \int_0^\infty \left[F_T^{\text{intra}}(t) - F_T^{\text{intra}}(0) \right] f_R(t) dt$$

then,

$$Pr_{\text{block}}^{\text{intra}} = \frac{m \cdot \bar{T}_S^{\text{intra}}}{1 + m \cdot \bar{T}_S^{\text{intra}}}. \tag{6.21}$$

$$Pr_{\text{block}}^{\text{inter}} = P(t_s^{\text{inter}} > t_r) = 1 - P(t_s^{\text{inter}} \leq t_r). \tag{6.22}$$

$$P(t_s^{\text{inter}} > t_r) = 1 - \int_0^\infty \left[F_T^{\text{inter}}(t) - F_T^{\text{inter}}(0) \right] f_R(t) dt$$

$$Pr_{\text{block}}^{\text{inter}} = \frac{m \cdot \bar{T}_S^{\text{inter}}}{1 + m \cdot \bar{T}_S^{\text{inter}}}. \quad (6.23)$$

Using the grid based configuration, α and β are expressed as follows :

$$\alpha = \sum_{i=1}^{K-1} \sum_{j=i}^{i+1} \pi_i^j + \frac{3}{4} \sum_{j=1}^K \pi_i^j + \frac{1}{2} \pi_i^j, \quad (6.24)$$

$$\beta = \frac{1}{4} \sum_{j=1}^K \pi_i^j + \frac{1}{2} \pi_i^j. \quad (6.25)$$

6.6 Analytical and Simulation Results

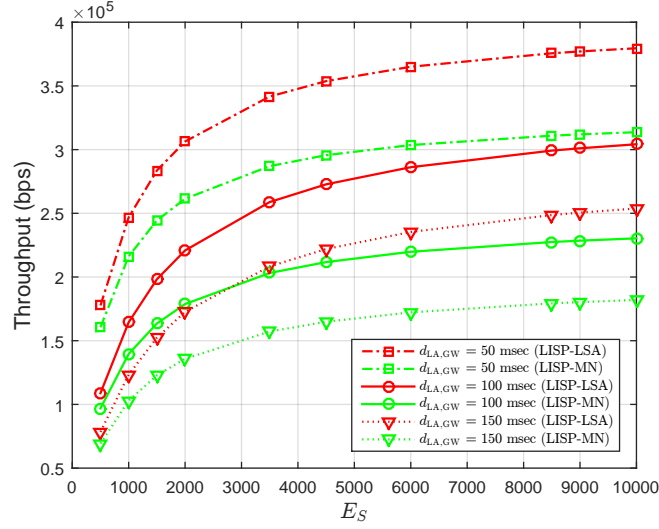
In this section, we evaluate the throughput and the handover blocking probability of both protocols using the analytical models previously developed. Indeed, both schemes have been implemented in a more realistic environment using the network simulator ns-3.26 [154].

6.6.1 Numerical Evaluation

Table S1 of the supplemental material summarizes the default parameters that we use from [128–130, 152] based on the recommendations in [155]. We compare the performance of both schemes using the improvement ratio or gain. We evaluate the gain as the ratio of the costs difference over the cost of the LISP-MN scheme under the same parameters.

6.6.2 Simulation Setup

Fig. 6.3 presents the simulation framework. It is composed of two LSAs consisting of 50 SCs with two LAs. The SCs are connected to their LA, and the LAs are wired to the GW. The MNs are represented by the thick dots and they are randomly deployed in the network. The LTE access network is configured on the SCs and the MNs. The MNs experience mobility based on the Gauss-Markov mobility model [156]. The packets are sent through the intermediate router and the remote access router xTR_{CN}. The distance between two neighboring cells is λ_d . The network becomes denser with small values of λ_d . The traffic parameters are adopted from the numerical analysis. Table S2 of the supplemental material presents the network configuration and the simulation parameters.

Figure 6.4 Throughput versus E_S and $d_{LA,GW}$

observed that the throughput for different values of λ_p tends to converge to the same values. This observation is due to the fixed number of handovers per sessions for the analytical results. The maximum improvement that we observed is 22%.

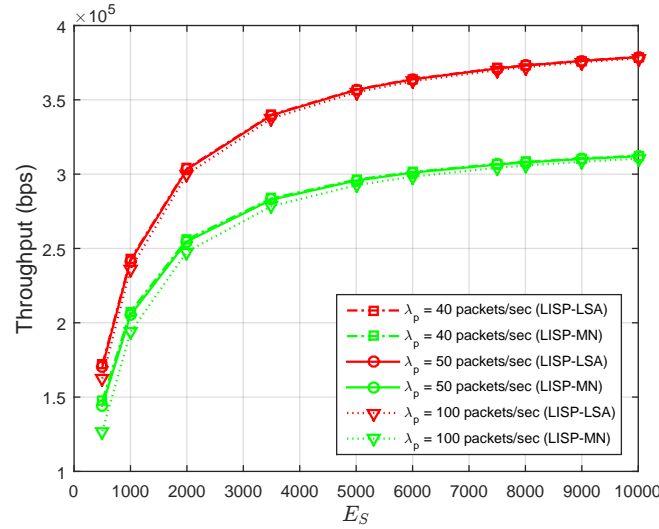
Figure 6.5 Throughput versus E_S and λ_p

Fig. 6.6 plots the throughput against E_S and K . It is observed that the throughput increases as the number of packets per sessions increases. The LISP-LSA scheme outperforms the LISP-MN scheme. Although we observed that the parameter K considerably impacts the signaling cost, the result shows that different values of K do not significantly impact the throughput.

The maximum throughput gain is 22%.

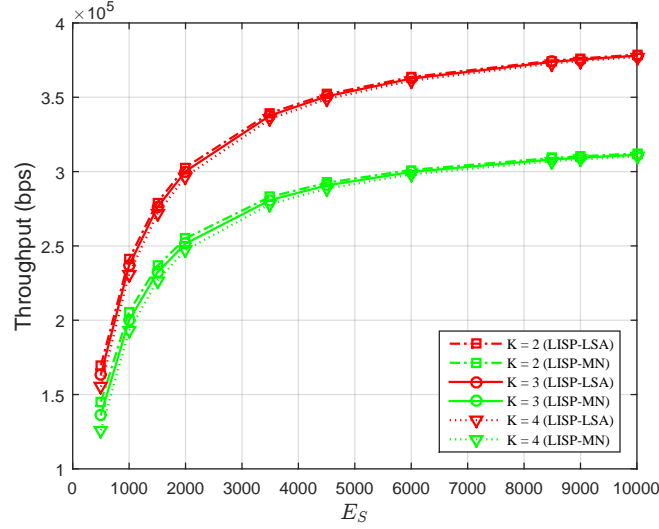


Figure 6.6 Throughput versus E_S and K

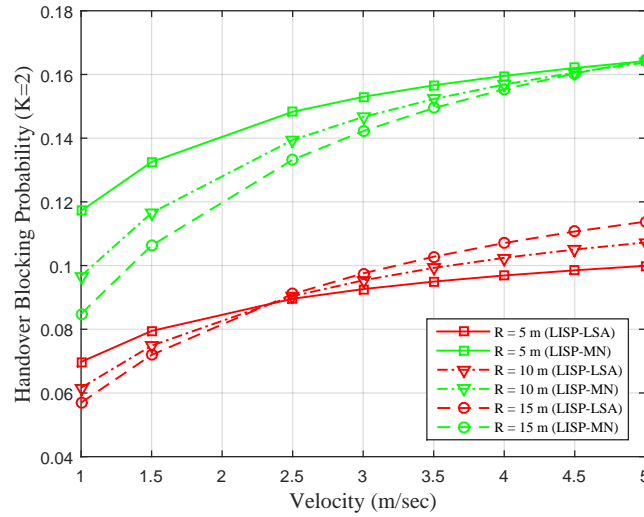


Figure 6.7 Handover Blocking Probability versus v and R

Fig. 6.7 plots the handover blocking probability as a function of v and R . When the velocity of the MN increases, the cell residence time is smaller, therefore leading to a relatively high mobility rate. Especially in a denser SC network configuration, the MN experiences frequent handovers. Since the LISP-LSA protocol mitigates the handover signaling time, the handover blocking probability of the LISP-LSA protocol is reduced. When the velocity increases, the handover blocking probability increases for both protocols. When $R = 5$ m, the LISP-LSA reduces the handover blocking probability by 40%. The blocking probability is reduced by

35% and 31% when the SC's radius is 10 m and 15 m respectively. When the network becomes denser, the LISP-LSA still achieves better performance.

Fig. 6.8 depicts the handover blocking probability against v and K . For both protocols, it is observed that the handover blocking probability increases with the MN's velocity. We showed in [167] that, when the size of the LSA increases, the average signaling cost also increases. Hence, increasing K increases the handover signaling time. Therefore, the handover blocking rate increases for both protocols with the combined effect of the MN's mobility rate and the increased signaling time. When $K=4$, we observed that the LISP-LSA protocol reduces the handover blocking probability by 26%. With smaller values of K , the blocking probability is reduced by 33% and 30% when $K=2$ and $K=3$ respectively.

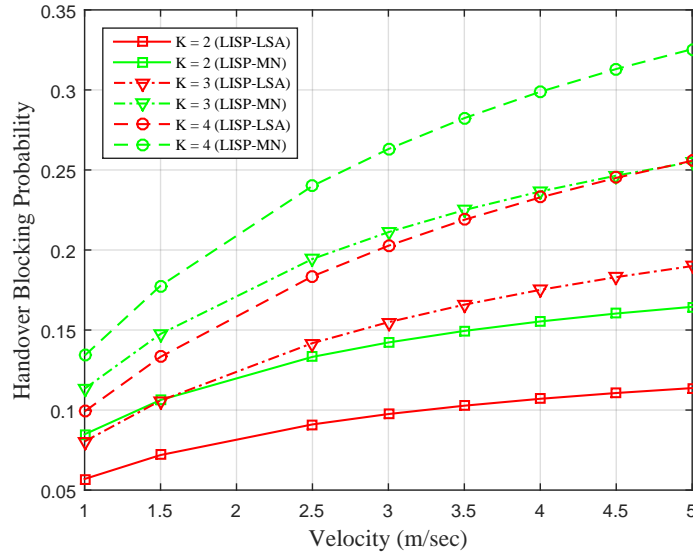


Figure 6.8 Handover Blocking Probability versus v and K

Fig. 6.9 depicts the handover blocking probability against v and $d_{LA, GW}$. For both protocols, we observe that the handover blocking probability increases as the MN's velocity increases. Indeed, $d_{LA, GW}$ has a significant impact on the handover signaling time. The effect of $d_{LA, GW}$ on the blocking probability is more obvious on the LISP-MN protocol while the LISP-LSA protocol is relatively less affected by $d_{LA, GW}$. On the one hand, the LISP-MN protocol always performed a location update at the GW after a handover. More importantly, the LISP-MN protocol performs the Solicit-Map-Request procedure for the binding update at the CNs. It significantly increases the handover signaling time. Therefore, when the network becomes denser, the handover blocking probability of the LISP-MN protocol is high. This situation is further exacerbated as the mobility rate increases. On the other hand, the LISP-LSA reduces the handover latency because of the proximity of the LA to the network edge. It is observed

that the blocking probability is reduced up to 35%, 47% and 55% by the LISP-LSA protocol when $d_{LA,GW} = 10$ msec, $d_{LA,GW} = 50$ msec, and $d_{LA,GW} = 100$ msec respectively. In short, as the MN becomes far from the core network LISP-LSA protocol considerably reduces the handover blocking probability.

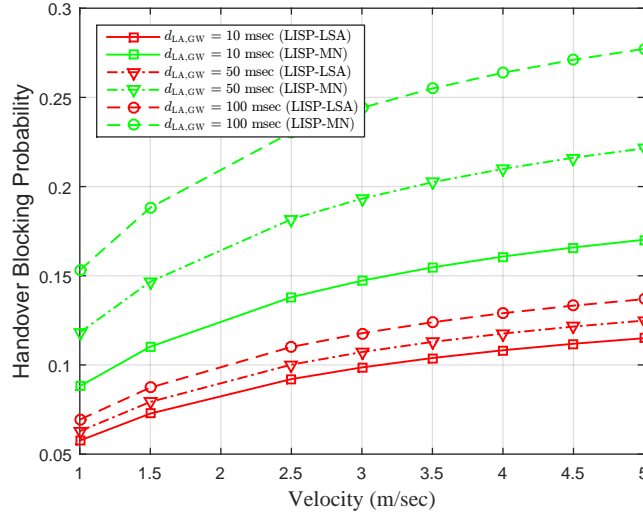


Figure 6.9 Handover Blocking Probability versus v and $d_{LA,GW}$

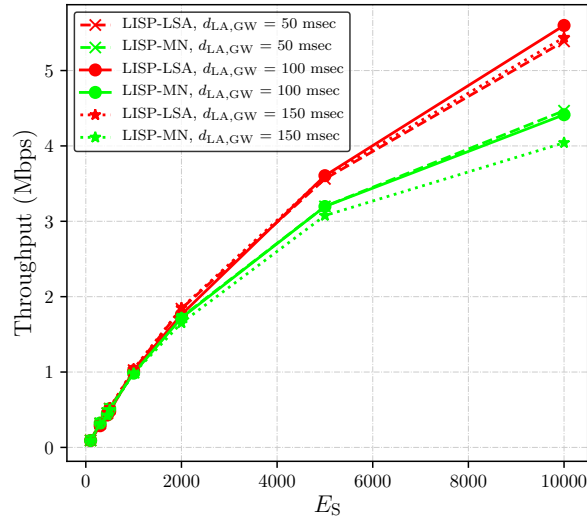


Figure 6.10 Throughput versus E_S and $d_{LA,GW}$

6.6.3.2 Simulation Results

Fig. 6.10 plots the simulated throughput against E_S and $d_{LA,GW}$. For both protocols, the throughput increases with the number of packets per session. However the LISP-LSA protocol

outperforms the LISP-MN protocol for the values of $d_{LA,GW}$. We observe that the throughput gain is 21%, 27%, and 36% when $d_{LA,GW} = 50$ msec, $d_{LA,GW} = 100$ msec, and $d_{LA,GW} = 150$ msec respectively.

Fig. 6.11 plots the simulated throughput against E_S and λ_p . The throughput increases with the number of packets per sessions for both protocols. The proposed LISP-LSA protocol shows improved performance. Indeed, we observe that the throughput increases when the data packet rate increases because the total transmission time decreases. However, when the data rate increases, the throughput gain decreases for both protocols because of the increased packet loss rate. The throughput gain is 12%, 15%, 24%, and 28% respectively when λ_p is 150, 100, 50 and 40 packets/sec respectively.

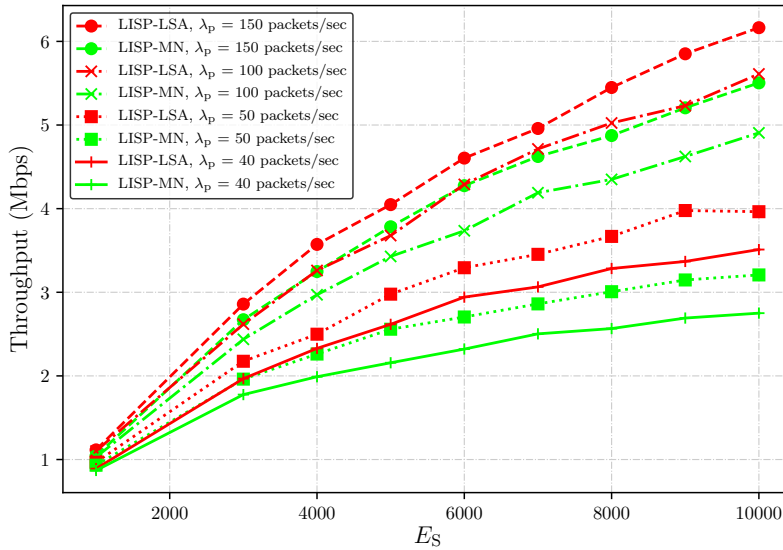


Figure 6.11 Throughput versus E_S and λ_p

Fig. 6.12 depicts the network throughput against the network density and the MN's velocity. When the network becomes denser ($\lambda_d = 15$ m), we observed that the throughput of the LISP-MN protocol decreases as the MN's velocity increases. The throughput gain is about 37% when $v = 3.5$ m/sec and it increases considerably to more than 90% when v is greater than 5.5 m/sec. On the one hand, the densification of the network leads to frequent handovers and the increased handover latency of the LISP-MN protocol induce a high packet loss rate. Therefore, the network throughput is significantly reduced. On the other hand, the LA at the network edge with the LISP-LSA protocol considerably reduces the handover latency. More importantly, the packet forwarding with the LISP-LSA scheme also reduces the packet loss rate, leading to the increased throughput gain. As the network become less denser ($\lambda_d > 20$ m), the throughput gain is about 10% when $v = 3.5$ m/sec, and the maximum gain is

about 31% when the velocity increases beyond 5.5 m/sec. In all cases, the LISP-LSA scheme performs better.

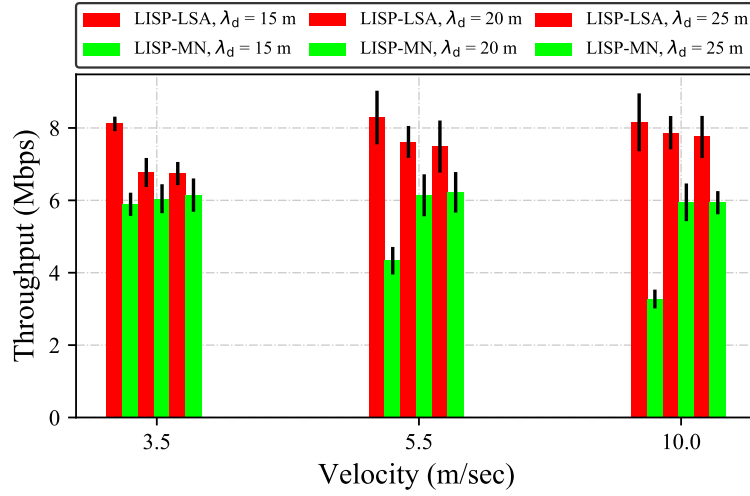


Figure 6.12 Throughput versus v and λ_d

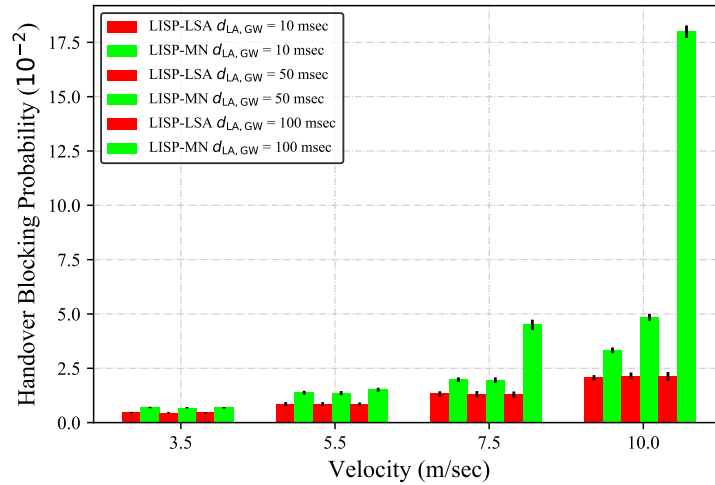


Figure 6.13 Handover Blocking Probability versus v and $d_{LA, GW}$

Fig. 6.13 depicts the handover blocking probability against the MN's velocity and the backhaul link latency towards the GW. As the MN's velocity increases, it is observed that the blocking probability of the LISP-MN protocol also increases with the link latency. However, we observe that the LISP-LSA protocol is less sensitive to the backhaul link latency. We observe that the LISP-LSA protocol reduces the blocking probability by 33% when $v = 3.5$ m/sec regardless of the link latency. When $v = 5.5$ m/sec, the blocking probability is reduced by 38% when $d_{LA, GW} = 10$ msec. The gain slightly increases up to 45% when $d_{LA, GW} = 100$

msec. As the MN's velocity increases ($v > 7.5$ m/sec and $d_{LA, GW} = 100$ msec), the blocking probability is considerably reduced by more than 70%. On the one hand, the LISP-MN protocol always performs the location update at the GW. The increased signaling time and the short cell residence time negatively impact the network performance and induce a higher handover blocking rate. On the other hand, the signaling delay is considerably reduced with the LISP-LSA scheme. Consequently, the handover blocking probability is also reduced.

6.7 Conclusion

SCs and ultra-dense network architectures have emerged as one of the essential techniques to support the design of the 5G networks. This paper proposed a new architecture for mobility management in 5G Identifier Locator Split dense SCs. The architecture uses the LISP-LSA protocol for mobility management. To evaluate the performance of the proposed architecture, we developed a mathematical model of the throughput and the handover blocking probability. To further validate the models, the proposed architecture was implemented using the network simulator ns-3.26.

Numerical and simulation results show that the proposed architecture significantly improves the network throughput. Moreover, the handover blocking probability is considerably reduced thus promoting a better user experience. Future research will be dedicated to integrating this new architecture into Software Defined Networking.

CHAPITRE 7 DISCUSSION GÉNÉRALE

Dans ce chapitre, nous ferons une synthèse de nos travaux au regard des objectifs de recherche initialement fixés. Par la suite, nous porterons un regard sur l’approche méthodologique que nous avons suivie. Nous terminerons ce chapitre par une analyse de l’ensemble des résultats que nous avons obtenus en vue de mettre en lumière leur portée pratique.

7.1 Synthèse des résultats

L’objectif principal de cette thèse consistait à concevoir une architecture de gestion de la mobilité dans les réseaux denses 5G. Pour atteindre cet objectif nous avons réparti le travail en trois étapes complémentaires.

Lors de la première phase, nous avons proposé une version initiale du protocole LISP-LSA afin de gérer la mobilité des MNs. Ce protocole est basé sur le mécanisme le mécanisme d’ID/Locator Split. Lors d’une phase préalable à la conception du protocole, nous avons proposé un modèle de densification du réseau qui devait servir de fondation à notre travail. Dans le but de procéder à l’évaluation de performance, un modèle analytique a été formulé afin de représenter le réseau densifié. Une première évaluation de performance nous a permis de mettre en évidence le gain de performance du modèle de densification proposé.

Dans une deuxième phase, le protocole initial a été étendu afin d’intégrer le support de relèves proactives à la gestion de la mobilité en nous basant sur les informations de la couche liaison de données (L2). Afin d’évaluer les performances desdits mécanismes, nous avons proposé les modèles analytiques et de simulation de plusieurs métriques que sont : le coût de signalisation, le coût de livraison des paquets, la latence des relèves, la perte de paquets.

Lors de la troisième phase du projet, nous avons proposé une architecture en définissant ses différentes composantes, ainsi que les mécanismes permettant d’assurer la gestion de la mobilité des usagers. Il s’agissait des mécanismes d’enregistrement, de localisation, de routage de paquets et de gestion des relèves. L’architecture proposée sépare le plan de signalisation du plan de données pour une plus grande flexibilité. L’architecture proposée intègre aussi un réseau overlay de signalisation basé sur le protocole LISP. Par la suite, nous avons défini les modèles analytiques du débit, et de la probabilité d’interruption des sessions. Par la suite, nous avons procédé à une nouvelle évaluation de performance de l’architecture proposée.

7.2 Méthodologie

Tout au long de nos travaux, nous avons basé notre évaluation de performance sur une modélisation analytique suivie d'une validation par des résultats de simulation.

Toutes les métriques étudiées dans le cadre de notre recherche ont été implémentées avec le logiciel MATLAB pour l'obtention des résultats numériques. Le protocole proposé, les composantes de l'architecture, les mécanismes de contrôle, ainsi que les modèles d'évaluation de performance ont tous été implémentés avec le logiciel de simulation de réseau ns-3.26 pour l'évaluation de performance. Cette phase d'implémentation a été la plus ardue et la plus laborieuse du projet de recherche.

7.3 Analyse des résultats

Les résultats analytiques ainsi que les résultats de simulation que nous avons obtenus ont révélé des gains de performance significatifs de l'architecture que nous proposons.

La mise en oeuvre du concept de la zone de localisation de service dans un réseau densifié ID/Locator Split offre plusieurs avantages. D'une part, la présence du point d'ancrage de mobilité près de la périphérie du réseau permet de réduire significativement la signalisation du réseau et la charge de la passerelle. Les temps de relèvement ont significativement été réduits, de même que la perte de paquets. De ce fait, le débit du réseau a été amélioré et la probabilité d'interruption des sessions a été aussi réduite.

D'un point de vue architectural, le protocole LISP-LSA réutilise des concepts du protocole LISP qui a été identifié comme un protocole mature dont le déploiement serait convenable aux réseaux 5G. L'utilisation du mécanisme ID/Locator Split favorise aussi une évolutivité du réseau par la réduction de la taille des tables de routage au niveau des routeurs tunnels, ainsi qu'au niveau du réseau coeur. D'un point de vue économique, le modèle de densification proposé permet de réduire le coût de déploiement de l'infrastructure du réseau par le raccordement des cellules de petite taille, aux points d'ancrage situés à la périphérie du réseau. De plus, les mécanismes proposés étant déployés au niveau de la couche réseau, ils peuvent facilement s'intégrer aux nouveaux types de réseaux d'accès qui seront déployés dans les réseaux 5G.

CHAPITRE 8 CONCLUSION ET RECOMMANDATIONS

Dans ce chapitre, nous procéderons à une récapitulation des travaux réalisés dans le cadre de cette thèse en présentant nos principales contributions. Par la suite, nous discuterons des limitations de notre approche et terminerons par une liste de recommandations pouvant faire l'objet de travaux futurs.

8.1 Synthèse des travaux

L'objectif global de cette thèse consistait en la conception d'une architecture de gestion de la mobilité dans les réseaux denses 5G. Dans un contexte de densification du réseau dans le but d'en accroître sa capacité globale, nous avons mis en oeuvre le mécanisme d'ID/Locator Split pour la gestion de la mobilité dans les réseaux denses 5G en tenant compte de plusieurs métriques d'évaluation de performances telles que le coût de signalisation, le coût de livraison des données, la latence des relèves, le taux de perte de paquets, le débit, ainsi que la probabilité d'interruption des sessions. Les principales contributions de nos travaux se résument comme suit :

1. Nous avons proposé un modèle horizontal de densification du réseau et avons introduit le concept de zone de localisation de service dans un réseau dont le fonctionnement est basé sur le mécanisme d'ID/Locator Split. Dans cette configuration, une LSA est localement gérée par un LA. Ce modèle présente l'avantage de pouvoir être déployé dans différents contextes. La proximité du LA de la périphérie du réseau est bénéfique à une gestion locale de la mobilité. En outre, d'un point de vue économique, le raccordement des SCs au LA réduit le coût de déploiement global du réseau densifié en évitant des liaisons directes des SCs au réseau coeur. Le modèle de densification proposé réduit le trafic de signalisation dû à la mobilité, il favorise un allègement de la charge totale du GW, permet une évolutivité du réseau en plus de renforcer sa résilience.
2. Nous avons proposé un modèle mathématique de représentation du réseau densifié. A cet effet, nous avons représenté le LSA et l'ensemble des SCs par une chaîne de Markov (2D) à temps discret. A travers ce modèle, le déplacement et l'activité d'un MN à travers le réseau sont suivis. Le modèle proposé intègre plusieurs paramètres tels que la taille du LSA, le taux d'arrivée et la durée des sessions du MN, le temps de résidence du MN au sein d'une SC.

3. Nous avons proposé un protocole de gestion de la mobilité utilisant le mécanisme d'ID/Locator Split dans un contexte de densification de réseaux 5G. Mieux, le protocole utilise les informations des autres couches de la pile TCP/IP afin d'offrir des mécanismes de relève proactifs. Cette décision se base sur une approche collaborative de gestion des relèves contrôlées par le MN et assistées par le réseau afin d'intégrer les préférences utilisateur lors des décisions de relève. De plus, la gestion locale de la mobilité permet de réduire le coût de signalisation, la latence des relèves, le taux de perte de paquets, la probabilité d'interruption des relèves et améliore considérablement le débit du réseau. Ce protocole favorise des relèves transparentes et garantit une meilleure QoE.
4. Nous avons proposé plusieurs modèles analytiques détaillés des métriques d'évaluation de performance d'un protocole utilisant le mécanisme ID/Locator Split. Il s'agit du coût de signalisation, du coût de livraison des paquets, de la latence des relèves, de la perte de paquets, du débit, de la probabilité d'interruption des sessions durant une relève. Mieux, les modèles proposés tiennent compte de l'impact conjoint de plusieurs paramètres du réseau (durée des session, taux d'arrivée des paquets, taille du LSA, temps de résidence au sein des SCs) sur les métriques évaluées.
5. Nous avons proposé une architecture de gestion de la mobilité dans les réseaux denses 5G. Cette architecture réutilise le concept du LSA ainsi que le protocole LISP-LSA ci-dessus mentionnés. Dans cette architecture, le trafic de signalisation est séparé du trafic de données. La signalisation du réseau est gérée par un réseau overlay dont le fonctionnement est basé sur le protocole LISP. Le plan de données quant à lui est composé d'un ensemble de routeurs tunnels dont les fonctionnalités ont été entendues. Les composantes essentielles de l'architecture proposée ont été implémentées dans le simulateur de réseau ns-3.26 afin de valider les résultats numériques de notre étude par des résultats de simulation. Les métriques évaluées ont montré de meilleures performances en comparaison à l'existant dans la littérature.

8.2 Limitations de la solution proposée

Plusieurs défis restent encore à relever pour la mise en service effective des réseaux denses 5G. En dépit des contributions de nos travaux, nous avons tout de même identifié certaines limitations à nos contributions. Elles sont énoncées comme suit :

1. La mise en oeuvre du concept de zone de localisation de service supposait une phase de planification préalable du réseau afin de déterminer la position du point d'ancrage

local (LA). Ceci pourrait créer une réticence de la part des opérateurs de réseaux à mettre en oeuvre l'approche proposée dans un réseau existant. Cependant, la virtualisation des composantes fonctionnelles du réseau, facilitera l'adoption de la solution proposée dans cette thèse. Par contre, lors du déploiement d'un nouveau réseau, la phase de planification permettra de mieux intégrer l'architecture proposée aux besoins des opérateurs.

2. La taille, K , du LSA était fixée dans le modèle analytique. Cependant, l'évaluation de performance a révélé que le coût associé à certaines métriques, notamment le coût de signalisation, variait en fonction de K , et qu'au delà d'un certain seuil, le gain de performance était perdu. Nous avons remarqué que les relèves inter-LSA génèrent un coût de signalisation plus élevé. Il serait intéressant de trouver une approche dynamique de modélisation ou de configuration du LSA afin d'ajuster la taille du LSA, de manière à réduire la fréquence des relèves inter-LSA. Dans cette approche, plusieurs paramètres tels que le nombre de MNs, la vitesse et la fréquence des relèves seront considérés. Ainsi, le gain de performance réalisé par le modèle de densification pourra être conservé dans la mesure du possible.
3. La configuration du LSA supposait un déploiement de blocs concentriques de SCs autour du LA. Bien que cette approche soit réalisable, elle ne couvre pas tous les scénarios de déploiement possible de SCs dans un réseau 5G. Dans une configuration de déploiement irrégulier des SCs, d'autres outils de modélisation tels que la géométrie stochastique seraient probablement plus adaptés à la modélisation du LSA.
4. Les protocoles proposés ont été essentiellement limités à la couche réseau L3. L'hétérogénéité au niveau du réseau 5G n'a pas été considérée. L'hétérogénéité de la 5G devra autant tenir compte de la diversité des technologies d'accès, mais également de la grande disparité de la nature des SCs (femtocellules, picocellules, microcellules). La prise en compte de ce facteur permettra de concevoir des algorithmes de relèvement adaptés à l'environnement hétérogène des réseaux 5G.
5. Lors de l'évaluation de la probabilité d'interruption d'une session, nous avons supposé que les ressources nécessaires à l'exécution des relèves ne dépassaient pas les ressources disponibles au niveau des SCs en mettant plutôt l'accent sur le temps de résidence et le temps de signalisation des relèves. Cependant, dans un environnement hétérogène comme celui des réseaux 5G, des procédures de contrôle d'admission doivent être prises en compte afin de maintenir un niveau de QoS approprié à travers divers types de réseaux d'accès.
6. Le mécanisme d'ID/Locator Split offre un certain niveau de sécurité. En effet, la position d'un MN (définie par son localisateur) est inconnue du CN qui n'a connaissance

que de son identifiant. Cependant, il serait approprié de renforcer au niveau des routeurs tunnels la protection des données liées à la localisation des MNs, d'introduire des mécanismes d'identification plus sécurisés au réseau.

7. Notre travail ne considère aucun aspect lié à la consommation énergétique des protocoles et des mécanismes de gestion de la mobilité proposés. Cependant, avec l'avènement de l'IoT et l'omniprésence de périphériques dont les ressources énergétiques sont limitées, il est tout aussi indispensable de concevoir des protocoles sensibles aux contraintes énergétiques.

8.3 Améliorations futures

Chaque limitation identifiée à la section précédente pourrait servir de point de départ à l'amélioration de notre travail. Le focus de notre recherche ayant été mis sur l'aspect de la densification du réseau 5G, il est tout aussi nécessaire d'inclure aux mécanismes que nous avons proposés, une plus grande flexibilité dans leur gestion, ainsi qu'une extension de leur champ d'application. A cet effet, nous proposons les améliorations suivantes :

- L'intégration du protocole proposé à une architecture basée sur le SDN (Software-Defined Networking) afin d'apporter plus de flexibilité dans la gestion du réseau. Du fait de la séparation du plan de signalisation et du plan de données, l'architecture proposée est tout à fait compatible avec la philosophie du SDN. Cette intégration nécessitera probablement la conception de nouveaux protocoles combinant à la fois l'aspect de densification du réseau à sa gestion centralisée ;
- La virtualisation de toutes les composantes fonctionnelles du réseau réduira les coûts de déploiement et facilitera son extension. En plus, elle permettra une optimisation plus efficiente de l'utilisation des ressources du réseau. Il sera possible de procéder à un déploiement dynamique de nouvelles fonctionnalités en fonction du besoin. La reconfiguration automatique du LSA désormais virtualisé et le positionnement dynamique du LA pourront faciliter l'administration des ressources du réseau par un partitionnement dynamique (Network Slicing). Il sera donc nécessaire de concevoir de nouveaux algorithmes d'optimisation de ressources et de routage adaptés au contexte ;
- L'élaboration d'algorithmes de relèves dotés de mécanismes de contrôle d'admission appropriés à l'environnement hétérogène des réseaux d'accès de la 5G ;
- L'extension du protocole LISP-LSA afin d'introduire un support de la mobilité de groupe. Cette extension trouvera une application au sein des systèmes de transports publics intelligents. Des groupes d'usagers mobiles seront desservis au sein d'une même unité de transport (Bus, métro, tramway...etc) par des connexions sans fil

- d'une plus grande capacité de données ;
- L'intégration de mécanismes d'intelligence artificielle, et d'apprentissage machine dans la mise en oeuvre du protocole LISP-LSA. En effet, la collecte d'une grande quantité d'information d'aide à la décision permettra au réseau de s'auto-administrer, de procéder à un déploiement ou à une reconfiguration automatique de ressources en fonction de l'heure, du besoin, de la journée, et des services offerts aux usagers du réseau 5G.

RÉFÉRENCES

- [1] Cisco Visual Networking Index, “Visual Networking Index : Global Mobile Data Traffic, Forecast Update, 2017–2022,” Cisco, San Jose, Rapport technique, Fév. 2019.
- [2] Ericsson, “Ericsson Mobility Report, Special edition, World Economic Forum.” Rapport technique, Jan. 2019.
- [3] Ericsson, “5G Radio Access, Uen 284 23-3204 Rev B,” Ericsson, Rapport technique, Fév. 2015.
- [4] Cisco, “Cisco Ultra 5G Packet Core Solution,” Cisco, San Jose, Rapport technique, 2018.
- [5] N. Panwar, S. Sharma et A. K. Singh, “A Survey on 5G : The Next Generation of Mobile Communication,” *Physical Communication*, vol. 18, p. 64–84, Mar. 2016.
- [6] I. F. Akyildiz, S. Nie, S.-C. Lin et M. Chandrasekaran, “5G Roadmap : 10 Key Enabling Technologies,” *Computer Networks*, vol. 106, p. 17–48, Sep. 2016.
- [7] M. Kamel, W. Hamouda et A. Youssef, “Ultra-Dense Networks : A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, n°. 4, p. 2522–2545, Mai 2016.
- [8] X. Ge, S. Tu, G. Mao, C. X. Wang et T. Han, “5G Ultra-Dense Cellular Networks,” *IEEE Wireless Communications*, vol. 23, n°. 1, p. 72–79, Fév. 2016.
- [9] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella et D. Wübben, “Cloud Technologies for Flexible 5G Radio Access Networks,” *IEEE Communications Magazine*, 2014.
- [10] S. Pierre, *Réseaux et systèmes informatiques mobiles : fondements, architectures et applications*.
- [11] I. Al-Surmi, M. Othman et B. M. Ali, “Mobility Management for IP-based Next Generation Mobile Networks : Review, challenge and perspective,” *Journal of Network and computer Applications*, vol. 35, n°. 1, p. 295–315, Jan. 2012.
- [12] L. F. Akyildiz, J. McNair, J. Ho et H. U. and, “Mobility Management in Current and Future Communications Networks,” *IEEE Network*, vol. 12, n°. 4, p. 39–49, Juil. 1998.
- [13] W. Wang et I. F. Akyildiz, “A New Signaling Protocol for Intersystem Roaming in Next-Generation Wireless Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 19, n°. 10, p. 2040–2052, Oct. 2001.

- [14] Akyildiz, Ian F and McNair, Janise and Ho, Joseph SM and Uzunalioglu, Huseyin and Wang, Wenye, “Mobility Management in Next-Generation Wireless Systems,” *Proceedings of the IEEE*, vol. 87, n^o. 8, p. 1347–1384, Août. 1999.
- [15] W. Stallings, *Foundations of Modern Networking : SDN NFV QoE IoT and Cloud*.
- [16] W. Yu, H. Xu, H. Zhang, D. Griffith et N. Golmie, “Ultra-Dense Networks : Survey of State of the Art and Future Directions,” dans *2016 25th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2016, p. 1–10.
- [17] D. López-Pérez, M. Ding, H. Claussen et A. H. Jafari, “Towards 1 Gbps/UE in Cellular Systems : Understanding Ultra-Dense Small Cell Deployments,” *IEEE Communications Surveys & Tutorials*, vol. 17, n^o. 4, p. 2078–2101, 2015.
- [18] M. Ding, D. López-Pérez, G. Mao, P. Wang et Z. Lin, “Will the Area Spectral Efficiency Monotonically Grow as Small Cells go Dense?” dans *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, p. 1–7.
- [19] N. Bhushan, J. Junyi Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel et S. Geirhofer, “Network Densification : The Dominant Theme for Wireless Evolution Into 5G,” *IEEE Communications Magazine*, vol. 52, n^o. 2, p. 82–89, Fév. 2014.
- [20] V. P. Kafle, H. Otsuki et M. Inoue, “An ID/Locator Split Architecture for Future Networks,” *IEEE Communications Magazine*, vol. 48, n^o. 2, p. 138–144, Fév. 2010.
- [21] ITU-T, “General Requirements for ID/Locator Separation in NGN,” ITU-T, Rapport technique, Jan. 2009, Series Y.2015 : Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Networks.
- [22] M. M. Sajjad, D. Jayalath et C. J. Bernardos, “A Comprehensive Review of Enhancements and Prospects of Fast Handovers for Mobile IPv6 Protocol,” *IEEE Access*, vol. 7, p. 4948–4978, Jan. 2019.
- [23] A. Gupta et R. K. Jha, “A Survey of 5G Network : Architecture and Emerging Technologies,” *IEEE access*, vol. 3, p. 1206–1232, Juil. 2015.
- [24] Q. C. Li, H. Niu, A. T. Papathanassiou et G. Wu, “5G Network Capacity : Key Elements and Technologies,” *IEEE Vehicular Technology Magazine*, vol. 9, n^o. 1, p. 71–78, Jan. 2014.
- [25] L. Cominardi, F. Giust, C. J. Bernardos et A. De La Oliva, “Distributed Mobility Management Solutions for Next Mobile Network Architectures,” *Computer Networks*, vol. 121, p. 124–136, Juil. 2017.

- [26] J. Carmona-Murillo, V. Friderikos et J. González-Sánchez, “A Hybrid DMM Solution and Trade-Off Analysis for Future Wireless Networks,” *Computer Networks*, vol. 133, p. 17–32, Mar. 2018.
- [27] F. Giust, L. Cominardi et C. J. Bernardos, “Distributed Mobility Management for Future 5G networks : Overview and Analysis of Existing Approaches,” *IEEE Communications Magazine*, vol. 53, n^o. 1, p. 142–149, Jan. 2015.
- [28] D. Johnson, C. Perkins et J. Arkko, “Mobility Support in IPv6,” IETF, Rapport technique, 2004.
- [29] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury et B. Patil, “Proxy Mobile IPv6,” IETF, Rapport technique, 2008.
- [30] H. Chan, D. Liu, P. Seite, H. Yokota et J. Korhonen, “Requirements for Distributed Mobility Management,” Rapport technique, 2014.
- [31] S. Jeon, S. Figueiredo, R. L. Aguiar et H. Choo, “Distributed Mobility Management for the Future Mobile Networks : A Comprehensive Analysis of Key Design Options,” *IEEE Access*, vol. 5, p. 11 423–11 436, Jui. 2017.
- [32] C. Bernardos, A. De La Oliva, F. Giust, J. Zuniga et A. Mourad, “Proxy Mobile IPv6 Extensions for Distributed Mobility Management,” *IETF, ID draft-ietf-dmm-pmipv6-dlif-00*, Aoû. 2018.
- [33] J. Carmona-Murillo, I. Soto, F. J. Rodríguez-Pérez, D. Cortés-Polo et J. L. González-Sánchez, “Performance Evaluation of Distributed Mobility Management Protocols : Limitations and Solutions for Future Mobile Networks,” *Mobile Information Systems*, vol. 2017, Fév. 2017.
- [34] L. Yi, H. Zhou, D. Huang et H. Zhang, “An Analytical Study of Distributed Mobility Management Schemes with a Flow Duration Based Model,” *Journal of Network and Computer Applications*, vol. 41, p. 351–357, Mai 2014.
- [35] Muhammad Mohtasim Sajjad, “A Software Defined Networking based Adaptive Multimode Decentralized Mobility Architecture for 5G,” Mémoire de maîtrise, Queensland University of Technology, Dept. Elect. Eng. Comput. Sci., Queensland Univ. Technol., Brisbane, QLD, Australia., 2018.
- [36] Bangerter, Boyd and Talwar, Shilpa and Arefi, Reza and Stewart, Ken, “Networks and Devices for the 5G Erea,” *IEEE Communications Magazine*, vol. 52, n^o. 2, p. 90–96, Fév. 2014.
- [37] R. K. Medhi D, *Network Routing : algorithms, protocols, and architectures*. Morgan Kaufmann, 2017.

- [38] H. Zhang, H. Luo et H. C. Chao, “Dealing With Mobility-Caused Outdated Mappings in Networks With Identifier/Locator Separation,” *IEEE Transactions on Emerging Topics in Computing*, vol. 4, n^o. 2, p. 199–213, Avr. 2016.
- [39] C. E. Perkins, “Mobile IP,” *IEEE communications Magazine*, vol. 35, n^o. 5, p. 84–99, Mai. 1997.
- [40] M.-K. Yi, J.-W. Choi et Y.-K. Yang, “A Comparative Analysis on the Signaling load of Proxy Mobile IPv6 and Hierarchical mobile IPv6,” dans *2009 4th International Symposium on Wireless Pervasive Computing*. IEEE, Fév. 2009, p. 1–5.
- [41] R. Koodli, “Fast Handovers for Mobile IPv6,” IETF, Rapport technique, 2005.
- [42] M. Tao, H. Yuan, S. Dong et H. Yu, “Initiative Movement Prediction Assisted Adaptive Handover Trigger Scheme in Fast MIPv6,” *Computer Communications*, vol. 35, n^o. 10, p. 1272–1282, Jui. 2012.
- [43] H. Soliman, C. Castelluccia, K. El Malki et L. Bellier, “Hierarchical Mobile IPv6 Mobility Management (HMIPv6),” IETF, Rapport technique, 2005.
- [44] K.-S. Kong, W. Lee, Y.-H. Han, M.-K. Shin et H. You, “Mobility Management for all-IP Mobile Networks : Mobile IPv6 vs. Proxy Mobile IPv6,” *IEEE Wireless communications*, vol. 15, n^o. 2, p. 36–45, Avr. 2008.
- [45] M.-S. Kim, S. Lee, D. Cypher et N. Golmie, “Performance Analysis of Fast Handover for Proxy Mobile IPv6,” *Information Sciences*, vol. 219, p. 208–224, Jan. 2013.
- [46] Ahmed, Atiq and Boulahia, Leila Merghem and Gaiti, Dominique, “Enabling Vertical Handover Decisions in Heterogeneous Wireless Networks : A State-of-the-Art and A Classification,” *IEEE Communications Surveys & Tutorials*, vol. PP, n^o. 99, p. 1–36, 2013.
- [47] D. Charilas et A. Panagopoulous, “Multiaccess Radio Network Enviroments,” *IEEE Vehicular Technology Magazine*, vol. 5, n^o. 4, p. 40–49, Déc. 2010.
- [48] M. Kim, T.-w. Moon et S.-j. Cho, “A study on IEEE 802.21 MIH Frameworks in Heterogeneous Wireless Networks,” dans *2009 11th International Conference on Advanced Communication Technology*, vol. 1. IEEE, 2009, p. 242–246.
- [49] B. S. Ghahfarokhi et N. Movahhedinia, “A Survey on Applications of IEEE 802.21 Media Independent Handover framework in next generation wireless networks,” *Computer Communications*, vol. 36, n^o. 10, p. 1101–1119, Jui. 2013.
- [50] J. Fu, H. Ji et X. Li, “User-Adaptive Vertical Handover Scheme Based on MIH for Heterogeneous Wireless Networks,” dans *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*. IEEE, Sep. 2009, p. 1–4.

- [51] Y. Kim, S. Pack, C. G. Kang et S. Park, “An Enhanced Information Server for Seamless Vertical Handover in IEEE 802.21 MIH Networks,” *Computer Networks*, vol. 55, n^o. 1, p. 147–158, Jan. 2011.
- [52] Huang, Ze-qun and Bai, Song-nan and Jung, Jaeil, “A MIH Services Based Application-Driven Vertical Handoff Scheme for Wireless Networks,” dans *2009 Fifth International Joint Conference on INC, IMS and IDC*. IEEE, 2009, p. 1428–1431.
- [53] P. Neves, F. Fontes, S. Sargento, M. Melo et K. Pentikousis, “Enhanced Media Independent Handover Framework,” dans *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*. IEEE, apr 2009, p. 1–5.
- [54] Y. Wang, P. Zhang, Y. Zhou, J. Yuan, F. Liu et G. Li, “Handover Management in Enhanced MIH Framework for Heterogeneous Wireless Networks Environment,” *Wireless Personal Communications*, vol. 52, n^o. 3, p. 615–636, Nov. 2010.
- [55] M. Shadaram, D. Akopian, G. H. S. Carvalho, F. Shen, M. Yang, M. Palesi, C.-P. Lin, H.-L. Chen et J.-S. Leu, “A Predictive Handover Scheme to Improve Service Quality in the IEEE 802.21 network,” *Computers & Electrical Engineering*, vol. 38, n^o. 3, p. 681–693, Mai 2012.
- [56] Sriram, K. and Golmie, N., “Vertical Handoff Decision Algorithms for Providing Optimized Performance in Heterogeneous Wireless Networks,” *IEEE Transactions on Vehicular Technology*, vol. 58, n^o. 2, p. 865–881, Fév. 2009.
- [57] K. Sriram et N. Golmie, “Vertical Handoff Decision Algorithms for Providing Optimized Performance in Heterogeneous Wireless Networks,” *IEEE Transactions on Vehicular Technology*, vol. 58, n^o. 2, p. 865–881, Fév. 2009.
- [58] R. Ferrús, A. Brunstrom, K.-J. Grinnemo, R. Fracchia, G. Galante, F. Casadevall et al., “Towards Transport-Layer Mobility : Evolution of SCTP multihoming,” *Computer Communications*, vol. 31, n^o. 5, p. 980–998, Mar. 2008.
- [59] R. Stewart, “Stream control transmission protocol,” Rapport technique, 2007.
- [60] Ł. Budzisz, R. Ferrús, A. Brunstrom, K.-J. Grinnemo, R. Fracchia, G. Galante et F. Casadevall, “Towards Transport-Layer Mobility : Evolution of SCTP multihoming,” *Computer Communications*, vol. 31, n^o. 5, p. 980–998, Mar. 2008.
- [61] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley et E. Schooler, “SIP : Session Initiation Protocol,” Rapport technique, 2002.
- [62] Y.-S. Chen et K.-L. Wu, “A Cross-Layer Partner-Assisted Handoff Scheme for Hierarchical Mobile IPv6 in IEEE 802.16 e systems,” *Wireless Communications and Mobile Computing*, vol. 11, n^o. 4, p. 522–541, Avr. 2011.

- [63] H. Ahmed, S. Pierre et A. Quintero, "A Cooperative Road Topology Based Handoff Management Scheme," *IEEE Transactions on Vehicular Technology*, Oct. 2018.
- [64] L. J. Zhang et S. Pierre, "An Enhanced Fast Handover with Seamless Mobility Support for Next-Generation Wireless Networks," *Journal of Network and Computer Applications*, vol. 46, p. 322–335, Nov. 2014.
- [65] Abou-Khalil G, Krishnan S, Pierre S., "Seamless Handover for Multicast Mobile IPv6 Traffic." *International Journal of Wireless and Mobile Computing*, vol. 7, p. 308–317, Jan. 2014.
- [66] C. Makaya and S. Pierre, "An Architecture for Seamless Mobility Support in IP-Based Next-Generation Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, n^o. 2, p. 1209–1225, Mar. 2008.
- [67] C. Makaya C. and S. Pierre, "Enhanced Fast Handoff Scheme for Heterogeneous Wireless Networks," *Computer Communications*., vol. 31, n^o. 10, p. 2016–2029, Jui. 2008.
- [68] Ezzouhairi A, Quintero A., Pierre S., "Adaptive End-to-End Mobility Scheme for Seamless Horizontal and Vertical Handoffs," *Ubiquitous Computing and Communication Journal*, vol. 4, p. 28–41, 2010.
- [69] A. Quintero, S. Pierre et L. Alaoui, "A Mobility Management Model Based on Users' Mobility Profiles for IPv6 Networks," *Computer Communications*, vol. 30, n^o. 1, p. 66 – 80, 2006.
- [70] R. Beaubrun, S. Pierre et J. Conan, "An Approach for Managing Global Mobility and Roaming in the Next-Generation Wireless Systems", *Computer Communications*, vol. 28, n^o. 5, p. 571 – 581, 2005.
- [71] H. Modares, A. Moravejosharieh, J. Lloret et R. B. Salleh, "A Survey on Proxy Mobile IPv6 Handover," *IEEE Systems Journal*, vol. 10, n^o. 1, p. 208–217, Jan. 2014.
- [72] J. Montavont, E. Iovov et T. Noel, "Analysis of Mobile IPv6 Handover Optimizations and Their Impact on Real-Time Communication," dans *2007 IEEE Wireless Communications and Networking Conference*. IEEE, 2007, p. 3244–3249.
- [73] A. Papapostolou et H. Chaouchi, "Handoff With Energy Awareness for Future Pervasive Environments," *The Journal of Supercomputing*, vol. 64, n^o. 2, p. 357–382, May 2013.
- [74] M. Tao, F. Liu et C. Qu, "Predictive timely trigger and performance optimization for deterministic handovers," *International Journal of Wireless Information Networks*, vol. 22, n^o. 2, p. 97–104, Mar. 2015.
- [75] L. Zhang et Y.-C. Tian, "An Enhanced Fast Handover Triggering Mechanism for Fast Proxy Mobile IPv6," *Wireless Networks*, vol. 24, n^o. 2, p. 513–522, Fev. 2018.

- [76] M. Khan et K. Han, “An Optimized Network Selection and Handover Triggering Scheme for Heterogeneous Self-Organized Wireless Networks,” *Mathematical Problems in Engineering*, vol. 2014, Jun. 2014.
- [77] A. E. Bergh et N. Ventura, “PA-FMIP : A Mobility Prediction Assisted Fast Handover Protocol,” dans *MILCOM 2006-2006 IEEE Military Communications conference*. IEEE, 2006, p. 1–7.
- [78] L. J. Zhang, W. K. Chen, Y. Zhang, A. Quintero et S. Pierre, “Seamless Mobility Management Schemes for IPv6-based Wireless Networks,” *Mobile Networks and Applications*, vol. 19, n^o. 6, p. 745–757, Dec. 2014.
- [79] J. Espi, R. Atkinson, I. Andonovic et J. Dunlop, “Proactive route optimization for fast mobile ipv6,” dans *2009 IEEE 70th Vehicular Technology Conference Fall*. IEEE, 2009, p. 1–5.
- [80] K. El Malki, “Simultaneous Bindings for Mobile IPv6 Fast Handovers,” *draft-elmalki-mobileip-bicasting-v6-02.txt*, 2002.
- [81] N. V. Lopes, M. J. Nicolau et A. Santos, “A QoS-enable Solution for Mobile Environments,” dans *Actas da 10a Conferência sobre Redes de Computadores (CRC 2010)*. Universidade do Minho, 2010.
- [82] Lopes, Nuno Vasco and Nicolau, Maria João and Santos, Alexandre, “A QoS-enabled Resource Management Scheme for F-HMIPv6 Micro Mobility Approach,” *Telecommunication systems*, vol. 52, n^o. 1, p. 341–357, 2013.
- [83] R. Bless, J. Hillebrand, C. Prehofer et M. Zitterbart, “A Quality-of-Service Signaling Architecture for Seamless Handover Support in Next Generation, IP-based Mobile Networks,” *Wireless Personal Communications*, vol. 43, n^o. 3, p. 817–835, Nov. 2007.
- [84] J. Loughney, M. Nakhjiri, C. Perkins et R. Koodli, “Context Transfer Protocol (CXTP),” Rapport technique, 2005.
- [85] C. Liu, D. Qian, Y. Liu et K. Xiao, “A framework for end-to-end qos context transfer in mobile ipv6,” dans *IFIP International Conference on Personal Wireless Communications*. Springer, 2004, p. 466–475.
- [86] S. H. Hwang, J. H. Kim, C. S. Hong et J.-S. Sung, “Localized Management for Proxy Mobile IPv6,” dans *Int Conf on Information Networking, ICOIN*, 2010.
- [87] A. J. Jabir, S. K. Subramaniam, Z. Z. Ahmad et N. A. W. A. Hamid, “A Cluster-Based Proxy Mobile IPv6 for IP-WSNs,” *EURASIP Journal on Wireless Communications and networking*, vol. 2012, n^o. 1, p. 173, Mai 2012.
- [88] S. Park, “Fast localized proxy mobile IPv6 (FLPMIPv6),” Rapport technique, Fév. 2007.

- [89] J.-I. Kim, S.-J. Koh et N.-S. Ko, “B-PMIPv6 : PMIPv6 with Bicasting for Soft Handover,” dans *2009 11th International Conference on Advanced Communication Technology*, vol. 1. IEEE, 2009, p. 218–221.
- [90] C.-H. Shih, J.-L. Kuo, C.-H. Huang et Y.-C. Chen, “A proxy-based fast handover scheme for hierarchical mobile IPv6,” dans *Proceedings of the 5th international conference on ubiquitous information management and communication*. ACM, 2011, p. 21.
- [91] Jabir, Adnan J and Shamala, Subramaniam and Zuriati, Z and Hamid, Nawa, “A Comprehensive Survey of the Current Trends and Extensions for the Proxy Mobile IPv6 Protocol,” *IEEE Systems Journal*, vol. 12, n^o. 1, p. 1065–1081, Mar. 2018.
- [92] H. Ali-Ahmad, M. Ouzzif, P. Bertin et X. Lagrange, “Distributed Mobility Management : Approaches and Analysis,” dans *2013 IEEE international conference on Communications workshops (ICC)*. IEEE, Jui. 2013, p. 1297–1302.
- [93] Ali-Ahmad, Hassan and Ouzzif, Meryem and Bertin, Philippe and Lagrange, Xavier, “Distributed Dynamic Mobile IPv6 : Design and evaluation,” dans *2013 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, Avr. 2013, p. 2166–2171.
- [94] J.-H. Lee, J.-M. Bonnin, P. Seite et H. A. Chan, “Distributed IP Mobility Management from the Perspective of the IETF : Motivations, Requirements, Approaches, Comparison, and Challenges,” *IEEE Wireless Communications*, vol. 20, n^o. 5, p. 159–168, Oct. 2013.
- [95] J. C. Zuniga, C. J. Bernardos, A. De La Oliva, T. Melia, R. Costa et A. Reznik, “Distributed Mobility Management : a standards landscape,” *IEEE Communications Magazine*, vol. 51, n^o. 3, p. 80–87, Mar. 2013.
- [96] H. Ali-Ahmad, M. Ouzzif, P. Bertin et X. Lagrange, “Performance Analysis on Network-based Distributed Mobility Management,” *Wireless personal communications*, vol. 74, n^o. 4, p. 1245–1263, Fév. 2014.
- [97] C. J. Bernardos, A. De la Oliva et F. Giust, “A PMIPv6-based Solution for Distributed Mobility Management,” Jui. 2017.
- [98] F. Giust, C. J. Bernardos et A. de la Oliva, “Analytic Evaluation and Experimental Validation of a Network-Based IPv6 Distributed Mobility Management Solution,” *IEEE Transactions on Mobile Computing*, vol. 13, n^o. 11, p. 2484–2497, Nov. 2014.
- [99] F. Giust, L. Cominardi et C. Bernardos, “Distributed Mobility Management for Future 5G Networks : overview and analysis of existing approaches,” *IEEE Communications Magazine*, vol. 53, n^o. 1, p. 142–149, Jan. 2015.

- [100] T. Nguyen, C. Bonnet et J. Harri, “SDN-based Distributed Mobility Management for 5G Networks,” dans *2016 IEEE Wireless Communications and Networking Conference*, Avr. 2016, p. 1–7.
- [101] Tomovic, Slavica and Pejvanovic-Djurisic, Milica and Radusinovic, Igor, “SDN-based Mobile Networks : Concepts and Benefits,” *Wireless Personal Communications*, vol. 78, n^o. 3, p. 1629–1644, Oct. 2014.
- [102] Bernardos, Carlos Jesús and De La Oliva, Antonio and Serrano, Pablo and Banchs, Albert and Contreras, Luis M and Jin, Hao and Zúñiga, Juan Carlos, “An Architecture for Software-Defined Wireless Networking,” *IEEE Wireless Communications Magazine*, vol. 21, n^o. 3, p. 52–61, Mar. 2014.
- [103] Bilen, Tugce and Canberk, Berk and Chowdhury, Kaushik R, “Handover Management in Software-Defined Ultra-Dense 5G Networks,” *IEEE Network*, vol. 31, n^o. 4, p. 49–55, Juil. 2017.
- [104] Prados-Garzon, Jonathan and Adamuz-Hinojosa, Oscar and Ameigeiras, Pablo and Ramos-Munoz, Juan J and Andres-Maldonado, Pilar and Lopez-Soler, Juan M, “Handover Implementation in a 5G SDN-Based Mobile Network Architecture,” dans *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2016, p. 1–6.
- [105] Kim, Yong-hwan and Lim, Hyun-kyo and Kim, Kyoung-han and Han, Youn-Hee, “A SDN-Based Distributed Mobility Management in LTE/EPC Network,” *The Journal of Supercomputing*, vol. 73, n^o. 7, p. 2919–2933, Juil. 2017.
- [106] Seo, Eunil and Wi, Sarang Sarang and Zalyubovskiy, Vyacheslav and Chung, Tai-Myounga, “The Scalable LISP-Deployed Software-Defined Wireless Network (LISP-SDWN) for a Next Generation Wireless Network,” *IEEE Access*, vol. 6, p. 66 305–66 321, Nov. 2018.
- [107] Nguyen, Tien-Thinh and Bonnet, Christian and Harri, Jérôme, “SDN-based Distributed Mobility Management for 5G Networks,” dans *2016 IEEE Wireless Communications and Networking Conference*. IEEE, 2016, p. 1–7.
- [108] H. Ko, I. Jang, J. Lee, S. Pack et G. Lee, “SDN-based Distributed Mobility Management for 5G,” dans *2017 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2017, p. 116–117.
- [109] Zhang, Haijun and Liu, Na and Chu, Xiaoli and Long, Keping and Aghvami, Abdol-Hamid and Leung, Victor CM, “Network Slicing Based 5G and Future Mobile Networks : Mobility, Resource Management, and Challenges,” *IEEE Communications Magazine*, vol. 55, n^o. 8, p. 138–145, Aoû. 2017.

- [110] W. Shang, C. Bao et X. Li, “Research and comparison on the ID/Locator Separation Network Architecture,” dans *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*. IEEE, Apr. 2012, p. 1198–1203.
- [111] “Host Identity Protocol (HIP) - RFC 5201,” Rapport technique.
- [112] M. E. M. Campista, M. G. Rubinstein, I. M. Moraes, L. H. M. K. Costa et O. C. M. B. Duarte, “Challenges and Research Directions for the Future Internetworking,” *IEEE Communications Surveys & Tutorials*, vol. PP, n^o. 99, p. 1–30, 2013.
- [113] A. Gurtov, *Host Identity Protocol (HIP) : Towards the Secure Mobile Internet*, 2008.
- [114] M. B. E. Nordmark, “RFC 5533 : Shim6 : Level 3 Multihoming Shim Protocol for IPv6,” Rapport technique, 2009.
- [115] A. Garcia-Martinez, M. Bagnulo et I. Beijnum, “The Shim6 Architecture for IPv6 Multihoming,” *IEEE Communications Magazine*, vol. 48, n^o. 9, p. 152–157, Sep. 2010.
- [116] Chakchai So-In, Raj Jain, Subharthi Paul, Jianli Pan, “Future Wireless Networks : key issues and a survey (ID/Locator Split Perspective),” *International Journal of Communication Networks and Distributed Systems*, vol. 8, n^o. 1, p. 24–52, 2012.
- [117] D. Farinacci, D. Lewis, D. Meyer et V. Fuller, “The Locator/ID Separation Protocol (LISP),” IETF RFC 6830, Rapport technique, Jan. 2013.
- [118] S. Paul, J. Pan et R. Jain, “Architectures for the Future Networks and the Next Generation Internet : A survey,” *Computer Communications*, vol. 34, n^o. 1, p. 2–42, Jan. 2011.
- [119] T. Tronco, édit., *New Network Architectures : The Path to the Future Internet (Livre numérique Google)*, 2010.
- [120] EU FP7 ICT-317669-METIS/D6.5 METIS, “Deliverable D6.5 Version : 1, Report on simulation results and evaluations,” Rapport technique, Fev. 2015. [En ligne]. Disponible : <https://www.metis2020.com/wp-content/uploads/deliverables/index.html>
- [121] EU FP7 ICT-317669-METIS/D6.1 METIS, “Deliverable D6.1 Version : 1, Simulation guidelines,” Rapport technique, Oct. 2013. [En ligne]. Disponible : <https://www.metis2020.com/wp-content/uploads/deliverables/index.html>
- [122] K. H. Chiang et N. Shenoy, “A 2-D Random-Walk Mobility Model for Location-Management Studies in Wireless Networks,” *IEEE Transactions on Vehicular Technology*, vol. 53, n^o. 2, p. 413–424, Mar. 2004.
- [123] I. F. Akyildiz, Y.-B. Lin, W.-R. Lai, R.-J. Chen *et al.*, “A New Random Walk Model for PCS Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 18, n^o. 7, p. 1254–1260, Juil. 2000.

- [124] Cisco Visual Networking Index, “Visual Networking Index : Forecast and Methodology, 2015 - 2020,” Cisco, San Jose, Rapport technique, Jui. 2016.
- [125] D. Xenakis, N. Passas, L. Merakos et C. Verikoukis, “Mobility Management for Femtocells in LTE-Advanced : Key Aspects and Survey of Handover Decision Algorithms,” *IEEE Communications Surveys & Tutorials*, vol. 16, n^o. 1, p. 64–91, Jan. 2014.
- [126] C. White, D. Lewis, D. Meyer et D. Farinacci, “LISP Mobile Node,” IETF Internet draft, Oct. 2011.[Online]. Available : <http://tools.ietf.org/html/draft-meyer-lisp-mn-06>, Rapport technique, 2014.
- [127] D. Pacheco-Paramo, I. F. Akyildiz et V. Casares-Giner, “Local Anchor Based Location Management Schemes for Small Cells in HetNets,” *IEEE Transactions on Mobile Computing*, vol. 15, n^o. 4, p. 883–894, Avr. 2016.
- [128] R. Balakrishnan et I. Akyildiz, “Local Anchor Schemes for Seamless and Low-Cost Handover in Coordinated Small Cells,” *IEEE Transactions on Mobile Computing*, vol. 15, n^o. 5, p. 1182–1196, Mai. 2016.
- [129] T. Guo, A. ul Quddus, N. Wang et R. Tafazolli, “Local Mobility Management for Networked Femtocells Based on X2 Traffic Forwarding,” *IEEE Transactions on Vehicular Technology*, vol. 62, n^o. 1, p. 326–340, Jan. 2013.
- [130] C. Makaya et S. Pierre, “An Analytical Framework for Performance Evaluation of IPv6-based Mobility Management Protocols,” *IEEE Transactions on Wireless Communications*, vol. 7, n^o. 3, p. 972–983, Mar. 2008.
- [131] H. Saleet, O. Basir, R. Langar et R. Boutaba, “Region-Based Location-Service-Management Protocol for VANETs,” *IEEE Transactions on Vehicular Technology*, vol. 59, n^o. 2, p. 917–931, Fév. 2010.
- [132] Y. Xiao, Y. Pan et J. Li, “Design and Analysis of Location Management for 3G Cellular Networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, n^o. 4, p. 339–349, Avr. 2004.
- [133] C. VNI, “Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update The Cisco ® Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update,” *Cisco Public Information*, Jui. 2016.
- [134] T. Condeixa et S. Sargento, “Studying the Integration of Distributed and Dynamic Schemes in the Mobility Management,” *Computer Networks*, vol. 60, p. 46–59, Fév. 2014.
- [135] S. Paul, R. Jain et J. Pan, “An Identifier/Locator Split Architecture for Exploring Path Diversity through Site Multi-Homing - A Hybrid Host-Network Cooperative Ap-

- proach,” dans *2010 IEEE International Conference on Communications*, Mai. 2010, p. 1–5.
- [136] G. Pujolle, *Software Networks : Virtualization, SDN, 5G and Security*. John Wiley & Sons, 2015.
 - [137] E. M. O. Fafolahan et S. Pierre, “Local Anchor-Based Architecture for Mobility Support in Identifier Locator Split Networks,” dans *Proceedings of the Second International Conference on Internet of things and Cloud Computing*. ACM, Mar. 2017, p. 59.
 - [138] 3GPP, “LTE ; Evolved Universal Terrestrial Radio Access (E-UTRA) ; TDD Home eNodeB(HeNB) ; Radio Frequency(RF) requirements analysis,” Rapport technique, Juil. 2018.
 - [139] A. Rodríguez Natal, L. Jakab, M. Portolés, V. Ermagan, P. Natarajan, F. Maino, D. Meyer et A. Cabellos Aparicio, “LISP-MN : Mobile Networking Through LISP,” *Wireless Personal Communications*, vol. 70, n^o. 1, p. 253–266, Jui. 2012.
 - [140] M. Menth, D. Klein et M. Hartmann, “Improvements to LISP Mobile Node,” dans *2010 22nd International Teletraffic Congress (ITC 22)*, Sep. 2010, p. 1–8.
 - [141] M. Gohar et S.-J. Koh, “A Distributed Mobility Control Scheme in LISP Networks,” *Wireless Networks*, vol. 20, n^o. 2, p. 245–259, Fév. 2014.
 - [142] M. Gohar et S. J. Koh, “Network-Based Distributed Mobility Control in Localized Mobile LISP Networks,” *IEEE Communications Letters*, vol. 16, n^o. 1, p. 104–107, Jan. 2012.
 - [143] O. Semiari, W. Saad, M. Bennis et B. Maham, “Caching Meets Millimeter Wave Communications for Enhanced Mobility Management in 5G Networks,” *IEEE Transactions on Wireless Communications*, vol. 17, n^o. 2, p. 779–793, Fév. 2018.
 - [144] A. D. Assouma, R. Beaubrun et S. Pierre, “Mobility Management in Heterogeneous Wireless Networks,” *IEEE journal on selected areas in communications*, vol. 24, n^o. 3, p. 638–648, Mar. 2006.
 - [145] X. Ge, J. Ye, Y. Yang et Q. Li, “User Mobility Evaluation for 5G Small Cell Networks Based on Individual Mobility Model,” *IEEE Journal on Selected Areas in Communications*, vol. 34, n^o. 3, p. 528–541, Mar. 2016.
 - [146] H. A. Chan, “Distributed Mobility Management with Mobile IP,” dans *2012 IEEE International Conference on Communications (ICC)*. IEEE, Jui. 2012, p. 6850–6854.
 - [147] T.-T. Nguyen et C. Bonnet, “DMMS : A Flexible Architecture for Multicast Listener Support in a Distributed Mobility Management Environment,” *Computer Networks*, vol. 94, p. 129–144, Jan. 2016.

- [148] F. Giust, A. D. L. Oliva, C. J. Bernardos et R. P. F. D. Costa, “A Network-Based Localized Mobility Solution for Distributed Mobility Management,” dans *2011 The 14th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Oct. 2011, p. 1–5.
- [149] J. H. Lee, J. M. Bonnin et X. Lagrange, “Host-based Distributed Mobility Management Support Protocol for IPv6 Mobile Networks,” dans *2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2012, p. 61–68.
- [150] Yuguang Fang et I. Chlamtac, “Teletraffic Analysis and Mobility Modeling of PCS Networks,” *IEEE Transactions on Communications*, vol. 47, n^o. 7, p. 1062–1072, Juil. 1999.
- [151] Y. Fang, I. Chlamtac et Yi-Bing Lin, “Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks,” *IEEE Transactions on Computers*, vol. 47, n^o. 6, p. 679–692, Jui. 1998.
- [152] C. Makaya et S. Pierre, “An Architecture for Seamless Mobility Support in IP-Based Next-Generation Wireless Networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, n^o. 2, p. 1209–1225, Mar. 2008.
- [153] J. W. Mark et W. Zhuang, *Wireless communications and networking*. Prentice Hall, 2003.
- [154] Nsnam, “Ns-3 : Network simulator 3, ver 3.26,” Rapport technique, Oct. 2016.
- [155] G. T. 25.912, “Feasibility Study for Evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN),” 2012.
- [156] B. Liang et Z. J. Haas, “Predictive Distance-based Mobility Management for PCS Networks,” dans *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, Mar. 1999, p. 1377–1384.
- [157] A. H. Zahran, B. Liang et A. Saleh, “Mobility Modeling and Performance Evaluation of Heterogeneous Wireless Networks,” *IEEE Transactions on Mobile Computing*, vol. 7, n^o. 8, p. 1041–1056, Aoû. 2008.
- [158] 5GPP Architecture Working Group, “5G Architecture White paper, Version 2.0,” 5GPP, Rapport technique, Jan. 2018.
- [159] T. Condeixa et S. Sargento, “Context-Aware Adaptive IP Mobility Anchoring,” *Computer Networks*, vol. 71, p. 84–99, Oct. 2014.

- [160] H. Wang, S. Chen, M. Ai et H. Xu, “Localized Mobility Management for 5G Ultra Dense Network,” *IEEE Transactions on Vehicular Technology*, vol. 66, n°. 9, p. 8535–8552, Sep. 2017.
- [161] A. S. Chekkouri, A. Ezzouhairi et S. Pierre, “A New Integrated VANET-LTE-A Architecture for Enhanced Mobility in Small Cells HetNet using Dynamic Gateway and Traffic Forwarding,” *Computer Networks*, vol. 140, p. 15–27, Juil. 2018.
- [162] E. Demarchou, C. Psomas et I. Krikidis, “Mobility Management in Ultra-Dense Networks : Handover Skipping Techniques,” *IEEE Access*, vol. 6, p. 11 921–11 930, 2018.
- [163] M. Farokhi, A. Zolghadrasli et N. M. Yamchi, “Mobility-Based Cell and Resource Allocation for Heterogeneous Ultra-Dense Cellular Networks,” *IEEE Access*, vol. 6, p. 66 940–66 953, Nov. 2018.
- [164] H. Luo, H. Zhang et C. Qiao, “Efficient Mobility Support by Indirect Mapping in Networks With Locator/Identifier Separation,” *IEEE Transactions on Vehicular Technology*, vol. 60, n°. 5, p. 2265–2279, Jui. 2011.
- [165] F. Qiu, H. Zhou, X. Li, G. Wang et H. Zhang, “A Distributed Mobility Management Scheme in Networks with the Locator/Identifier Separation,” *International Journal of Communication Systems*, vol. 27, n°. 10, p. 1874–1893, Oct. 2014.
- [166] V. M. Dino Farinacci, *The LISP Network : Evolution to the Next-Generation of Data Networks*, 1^{er} éd., ser. The Networking Technology. Cisco Press, Jan. 2019, vol. 1.
- [167] Eric M.O. Fafolahan et Samuel Pierre, “A Seamless Mobility Management Protocol in 5G Locator Identifier Split Dense Small Cells,” *IEEE Transactions on Mobile Computing*, Mai 2019.

ANNEXE A Supplemental Material of the paper *A Seamless Mobility Management Protocol in 5G Locator Identifier Split Dense Small Cells*

A.1 Overview of the proposed Local anchor based network architecture

A.1.1 Registration procedure

The L3 signaling procedure is presented in Fig. A.1. After the L2 attachment, the MN is assigned a new locator by the DHCP system and sends a Map Register message, containing its identifier and locator to its attached xTR. The xTR updates its local cache and transfers the registration message to its LA along with the MN's parameters. The LA updates its local cache and transfers the request to the GW. Thereafter, the GW updates its cache with the provided information from the LA. Then, it sends a Map Reply message back to the LA. This message is further delivered to the xTR and to the MN. Following its first attachment, the MN sends a Map Discovery Request message to the LA to retrieve detailed informations (prefix address advertised, ingress and egress address, QoS parameters...) of xTRs in the LSA. The LA, which has a global view of the LSA, responds back to the MN with a Map Discovery Response message containing the list of xTRs along with networks information (ingress, egress IP addresses, network type, supported QoS levels...etc). The provided information is further used by the MN for fast handovers.

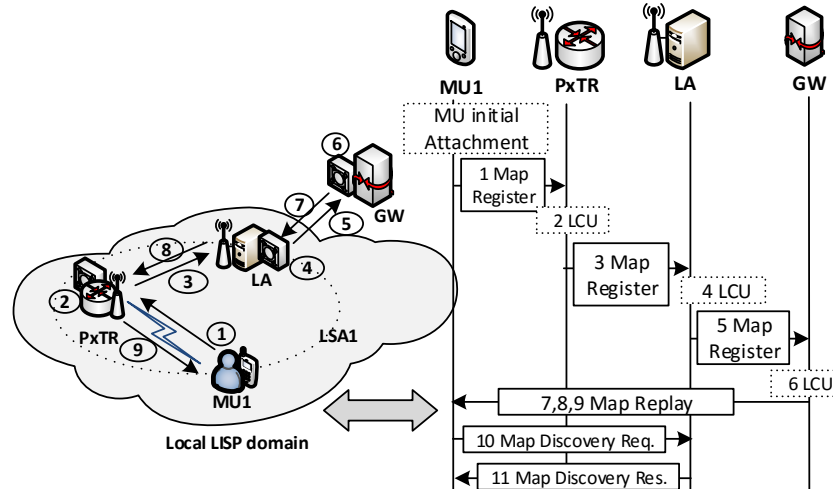


Figure A.1 Registration procedure

A.2 Link configuration and Simulation parameters

A.2.1 Example of link configuration in the LSA

Fig. A.2 shows a configuration of 3 neighboring SCs, located in the same LSA, which are connected to the LA along with the GW. For instance, the SC denoted by xTR1 is connected to the LA using a point-to-point link. Its interface's IP address is 10.0.0.1/30 whereas the LA's IP address is 10.0.0.2/30. Neighboring cells also have a point-to-point link connection for sending signaling and data packets. In Fig. A.2, xTR2 is connected to xTR3 using its interface with IP address set to 12.0.0.5/30 whereas xTR3 sends packets towards xTR2 with its interface configured with the IP address 12.0.0.6/30. Following the same approach, we configure the grid network of 50 SCs to run the simulations.

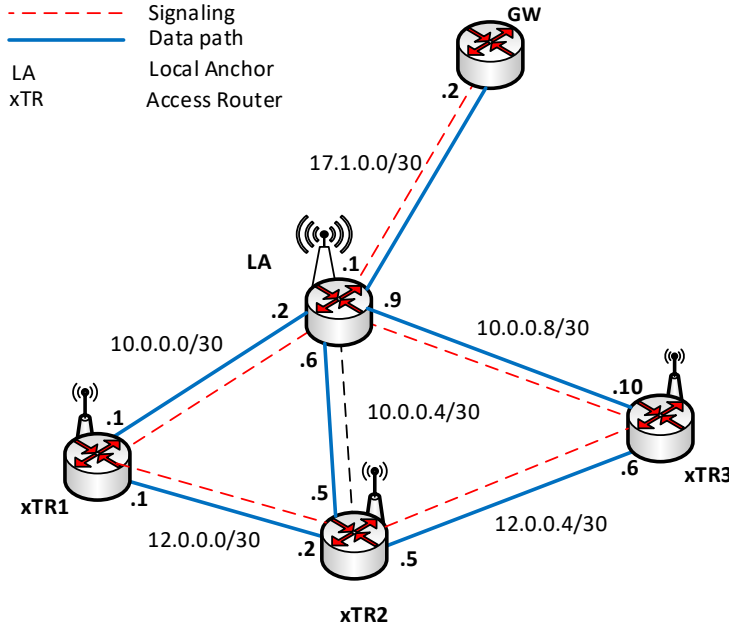


Figure A.2 Example of link configuration in LSA

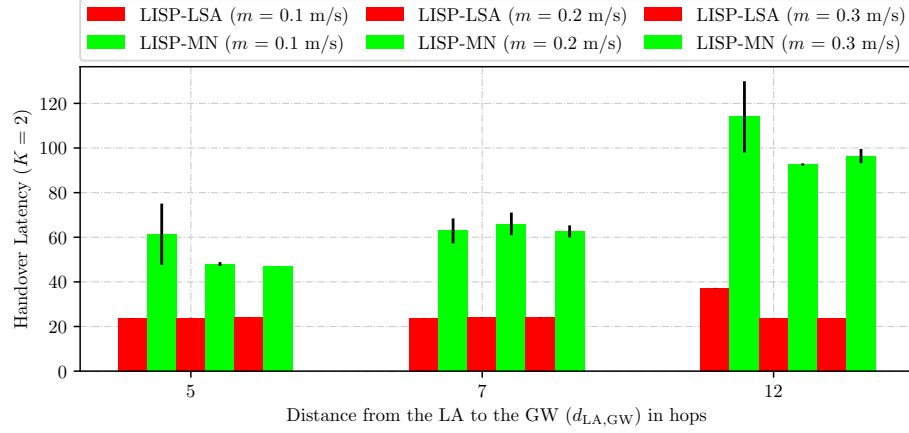
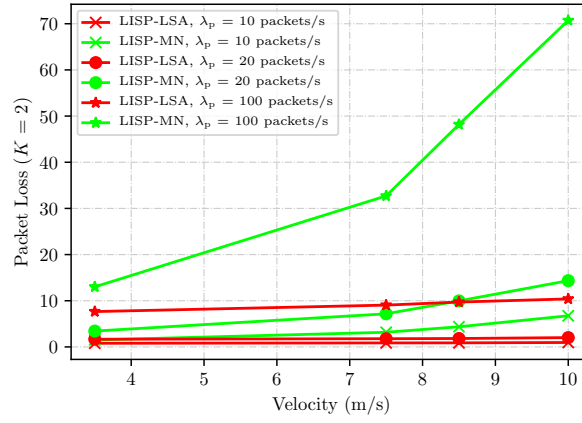
The simulation parameters are presented in Table B.2.

Fig. A.3 plots the handover latency against $d_{LA,GW}$ and m . It is shown that the handover latency of the LISP-MN protocol increases with $d_{LA,GW}$. The LISP-LSA scheme uses the proximity of the LA to the edge of the network to reduce the location update latency. The proposed protocol reduces the handover latency up to 80%.

Fig. A.4 plots the packet loss against the MNs's velocity and λ_p . It is shown that the pa-

Table A.1 Simulation Parameters

Parameters	Value
Total number of SCs	50
Transmission power (T_x)	16 dBm
Noise figure	5 dB
Pathloss model	Friis Spectrum Propagation Loss Model
Handover algorithm	A2-A4-RSRQ, Triggering Event A2
Densification parameter (λ_d)	15 m, 20 m (default), 50 m
Simulation time	1500 s
Mobility model	Gauss-Markov Mobility Model $\alpha = 0.90$, $\bar{v} = 5.0$ m/s (default), $\sigma_v = 0.2$, $\sigma_d = 0.2$
Number of MNs	60 (default)
LA to GW link parameters	Data Rate : 100 Gb/s Link type : Point-to-Point MTU : 2000 Delay : 20 ms Network IP : 17.0.0.0/30
LA to xTRs link parameters	Data Rate : 100 Gb/s Link type : Point-to-Point MTU : 2000 Delay : 10 ms Network IP : 10.0.0.0/30
xTR to xTR	Data Rate : 100 Gb/s Link type : Point-to-Point MTU : 3000 Delay : 10 ms Network IP : 12.0.0.0/30
GW to Intermediate Router	Data Rate : 100 Gb/s Link type : Point-to-Point MTU : 2000 Delay : 20 ms Network IP : 17.1.0.4/30
Intermediate Router to AR _{cn}	Data Rate : 100 Gb/s Link type : Point-to-Point MTU : 2000 Delay : 20 ms Network IP : 17.1.0.8/30

Figure A.3 Handover latency versus m and $d_{LA,GW}$ Figure A.4 Packet Loss versus λ_p and MN's velocity

cket loss increases with the MN's velocity because of frequent handovers. The LISP-LSA scheme outperforms the LISP-MN scheme. As the handover latency of the LISP-MN scheme increases, the packet loss also increases with λ_p . It is observed that the proposed scheme achieves up to 90% packet loss saving at a high data packet rate with a high speed.

ANNEXE B Supplemental Material of the paper *An Architecture for Mobility Management in 5G Identifier Locator Split Dense Small Cells*

B.1 Analytical Models

B.1.1 Definitions

We express the throughput and the handover blocking probability as a combination of several factors that we briefly define below :

- The data transmission time : it is the total time required to the CN to send the data packets on the transmission link ;
- The end-to-end data delivery time : it is the total time elapsed when a data packet is sent from the CN to the MN ;
- The session disruption time : it is the total time required to update the remote nodes whether it is the LA, the GW or the remote MS with the MN's new locator ;
- The Layer-2 handover latency : it is defined as the time required to the L2 technology to perform a handover ;
- The movement dection time : It is the time required by the MN to detect that it has moved to a new xTR so that it can configure a new locator ;
- The locator configuration time : it is the time required by the IP stack on any node to configure a new locator and to update its local caches.

B.2 Analytical Models

Fig. B.1 presents the movement direction probabilities. The MN moves from the SC to its neighbors SCs with an equal probability.

B.3 Numerical and Simulation Parameters

The parameters used for the numerical results are presented in Table B.1. The simulation parameters are presented in Table B.2.

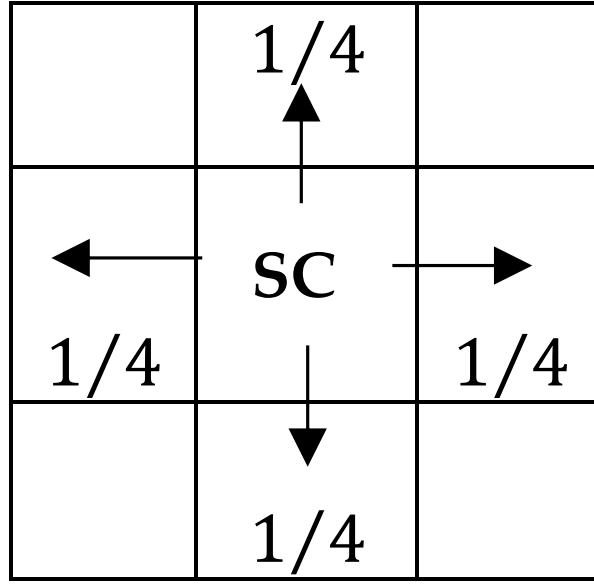


Figure B.1 The movement direction probabilities

B.4 Complementary simulation results

Fig. B.2 plots the handover blocking probability against v and the δ_{core} delay. As the core delay increases, we observed that the blocking rate significantly increases in the case of the LISP-MN protocol in comparison to the LISP-LSA protocol. It is observed that when $v = 3.5$ m/sec, the LISP-LSA protocol reduces the handover blocking probability by 34% independently of the core network delay. When the MN's velocity increases, LISP-LSA noticeably

Table B.1 Default values

Parameters	Value	Parameters	Value	Parameters	Value
λ	0.001/sec	t_{L2}	50 msec	λ_p	40,50,100,150 packets/sec
μ	0.01/sec	B_w	100 Mbps	L_w	2 msec
B_{wl}	11 Mbps	Δt	0.01 s	L_{wl}	10 msec
q	0.5	s_{RS}	52 bytes	s_d	1024 bytes
m	0.1/sec	τ	1	$\bar{\omega}_q$	$50 \cdot 10^{-6}$ sec
η	3	PC_{xTR}	8	PC_{LA}	12
$d_{MN,xTR}$	1	$d_{xTR,CN}$	5	$d_{LA,GW}$	7
T_{Proc}^{xTR}	10 msec	T_{Proc}^{LA}	10 msec	T_{Proc}^{GW}	20 msec
K	2,3,4	s_c	80 bytes	s_{RA}	80 bytes
PC_{GW}	24	$d_{xTR,CN}$	7	P_s	0.98

Table B.2 Simulation Parameters

Parameters	Value
Total number of SCs	50
Transmission power (T_x)	16 dBm
Noise figure	5 dB
Pathloss model	Friis Spectrum Propagation Loss Model
Handover algorithm	A2-A4-RSRQ, Triggering Event A2
Densification parameter (λ_d)	5 m, 10 m (default), 15 m
Simulation time	1500 sec
Mobility model	Gauss-Markov Mobility Model $\alpha = 0.90$, $\bar{v} = 5.0$ m/sec (default), $\sigma_v = 0.2$, $\sigma_d = 0.2$
Number of MNs	60 (default)
LA to GW link parameters	Data Rate : 100 Gb/sec Link type : Point-to-Point MTU : 2000 Delay : 10,50,100,150 msec Network IP : 17.0.0.0/30
LA to xTRs link parameters	Data Rate : 100 Gb/sec Link type : Point-to-Point MTU : 2000 Delay : 10 msec Network IP : 10.0.0.0/30
xTR to xTR	Data Rate : 100 Gb/sec Link type : Point-to-Point MTU : 3000 Delay : 10 msec Network IP : 12.0.0.0/30
GW to Intermediate Router	Data Rate : 100 Gb/sec Link type : Point-to-Point MTU : 2000 Delay : 25 (default),30,50 msec Network IP : 17.1.0.4/30
Intermediate Router to AR _{cn}	Data Rate : 100 Gb/sec Link type : Point-to-Point MTU : 2000 Delay : 25 (default),30,50 msec Network IP : 17.1.0.8/30

reduces the handover blocking rate. The reduction gain is about 37% when $v = 5.5$ m/sec and $\delta_{\text{core}} = 25$ msec. This gain appreciably increases up to 80% when v is greater than 7.5 m/sec and $\delta_{\text{core}} = 50$ msec. With the LISP-MN scheme, the MN performs a solicit-map-request with all its remote peers. This procedure is greatly affected by the core network delay, and the situation is further exacerbated as the MN's velocity increases. However, the LISP-LSA protocol is less sensitive to the core network delay, since it uses the fast handover schemes combined to the proximity of the LA to the network edge.

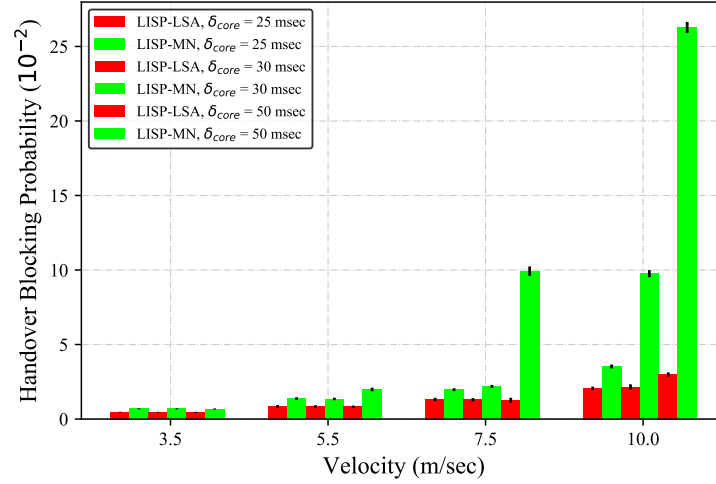


Figure B.2 Handover Blocking Probability versus v and σ_{core}