
République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohammed Seddik Ben Yahia-Jijel



Faculté des Sciences Exactes et Informatique
Département Informatique

Thèse

présentée pour l'obtention du diplôme de
Doctorat L.M.D
Filière : Informatique
Spécialité : Réseaux et Sécurité

par : **Asma BECHKIT**

PLACEMENT OPTIMAL DU RÉSEAU CŒUR ET OPTIMISATION DU TRAFIC DANS LES RÉSEAUX AUTO-DÉPLOYABLES

Soutenue le 27/06/2024, à 9h00, à la salle de conférence de la Bibliothèque Centrale

Devant le jury composé de

Président	M. TAFFAR	MCA	Université de Jijel
Directeur de thèse	A. MELIT	Prof.	Université de Jijel
Examineurs	A. LEMOUARI	Prof.	Université de Tamanghasset
	E. KERKOUCHE	Prof.	Université de Jijel
	M. BOULAICHE	MCA	Université de Skikda
	H. BOUAZIZ	MCA	Université de Jijel

Remerciements

Je tiens à remercier en premier lieu Dieu tout puissant qui m'a accordé la volonté et le courage pour mener à bien cette thèse.

Je tiens à adresser mes vifs remerciements à mon directeur de thèse Pr. MELIT Ali, pour avoir encadré mes travaux de thèse, pour sa disponibilité, ses précieux conseils et son encouragement tout au long de ce parcours.

Je tiens à remercier vivement les membres de jury, Dr. TAFFAR Mokhtar, Pr. LEMOUARI Ali, Pr. KERKOUICHE Elhillali, Dr. BOULAICHE Mehdi et Dr. BOUAZIZ Hamida, d'avoir accepté de faire partie du jury et d'avoir pris le temps pour lire et évaluer mon travail.

Je tiens aussi à exprimer mes vifs remerciements à BECHKIT Walid pour ses précieux conseils et remarques.

Je remercie chaleureusement toute ma famille, mes parents, mes frères, mes sœurs et tous mes amis. Ainsi que tous les enseignants du département Informatique.

Merci à toute personne qui m'a encouragée afin de terminer ce travail.

Résumé

Au cours des dernières années, nous assistons à l'émergence d'architectures auto-déployables au sein des réseaux cellulaires, combinée à une forte expansion de la virtualisation du réseau cœur. Ces avancées offrent une grande flexibilité en permettant la virtualisation des différentes fonctions du réseau cœur et leur co-localisation avec une ou plusieurs stations de base. Cependant, cela engendre de nouveaux défis, notamment en ce qui concerne le placement optimal des fonctions du réseau cœur en fonction de l'architecture physique et du trafic à acheminer depuis chaque station de base vers le réseau cœur. Dans le cadre de cette thèse, nous abordons ce problème en particulier et nous proposons un modèle générique et efficace qui permet d'identifier la meilleure station de base qui doit héberger les fonctions du réseau cœur tout en optimisant une métrique de performance souhaitée (flux global de trafic, des métriques d'équité entre les stations de base, etc.). Notre solution repose sur une modélisation conjointe du problème en tant que programme non linéaire mixte en nombres entiers (MINLP) et sur une linéarisation efficace des contraintes non linéaires. Notre modèle est générique et permet d'ajuster facilement les métriques à optimiser et/ou d'ajouter ou de modifier les contraintes selon le besoin.

En considérant la maximisation du flux global de trafic, nous avons dans un premier temps évalué les performances de notre modèle en le comparant à des solutions similaires de la littérature. Les résultats montrent que notre approche conjointe réduit considérablement le temps d'exécution par rapport aux solutions existantes tout en fournissant les mêmes résultats optimaux. Grâce à notre formulation générique, nous avons considéré dans un deuxième temps le problème d'équité entre les stations de base. En effet, les résultats montrent que la maximisation du flux global de trafic cause un déséquilibre important en termes de satisfaction absolue et relative par rapport à la demande de trafic des différentes stations de base. Nous avons donc proposé des variantes efficaces de notre modèle qui permettent d'assurer le placement optimal du réseau cœur tout en garantissant une équité entre les stations de base. Nous avons mené différentes simulations pour évaluer les solutions obtenues et les performances de nos variantes visant à garantir l'équité de satisfaction entre les stations de base. Les résultats de simulations nous ont permis de discuter du compromis entre la maximisation du flux de trafic et l'équité entre les stations de base, tout en mettant en avant

les avantages de notre approche de modélisation conjointe. Dans un troisième temps, nous avons enfin étudié plus en détail le passage à l'échelle de nos modèles en fonction du nombre de stations de base, du nombre de réseaux cœur à déployer et selon les métriques d'optimisation à considérer. Les résultats montrent que nos modèles offrent des temps d'exécution très raisonnables même dans le cas de grands réseaux avec des centaines de stations de base.

Mots clés : réseaux auto-déployables, réseaux cellulaires, virtualisation du cœur de réseau, placement optimal, linéarisation des contraintes, maximisation du flux de trafic, équité, scalabilité.

Abstract

Over the last few years, we have witnessed the emergence of self-deployable architectures in cellular networks, combined with a strong expansion of core network virtualization. These advances offer great flexibility by enabling core network functions to be virtualized and co-located with one or more base stations. However, this also brings new challenges, particularly in terms of optimal placement of core network functions according to the physical architecture and the traffic to be sent from each base station to the core network. In this thesis, we address this problem in particular, and propose a generic and efficient model to identify the best base station to host the core network functions, while optimizing a desired performance metric (maximum traffic flow, fairness metrics between base stations, etc.). Our solution is based on a joint modeling of the problem as a Mixed Integer Nonlinear Program (MINLP) and on an efficient linearization of the nonlinear constraints. Our model is generic, making it easy to adjust the metrics to be optimized and/or add or modify constraints as required.

When considering the maximization of overall traffic flow, we first evaluated the performance of our model by comparing it to similar solutions in the literature. The results show that our one-shot joint approach significantly reduces execution time compared to existing solutions, while delivering the same optimal results. Building on our generic formulation, we then addressed the problem of fairness between base stations. Indeed, the results show that maximizing overall traffic flow causes a significant imbalance in terms of absolute and relative satisfaction with respect to the traffic demand of the different base stations. We therefore proposed efficient variants of our model that ensure optimal placement of the core network while guaranteeing fairness between base stations. We conducted various simulations to evaluate the obtained optimal solutions and the performance of our fairness-aware solutions. The simulation results enabled us to discuss the trade-off between traffic flow maximization and fairness among base stations, while highlighting the benefits of our joint one-shot modeling approach. Finally, we studied in more detail the scaling of our models as a function of the number of base stations, the number of core networks to be deployed and the optimization metrics to be considered. The results show that our models offer very reasonable execution times even in the case of large networks with hundreds of base stations.

Key words : self deployable networks, cellular networks, core network virtualization, optimal placement, constraint linearization, traffic flow maximization, fairness, scalability.

الملخص

خلال السنوات الأخيرة، شهدنا ظهور البنى ذاتية الانتشار ضمن الشبكات الخلوية، إلى جانب التوسع القوي في المحاكاة الافتراضية للشبكة المركزية. سمحت هذه التطورات بتوفير مرونة كبيرة من خلال تمكين وظائف الشبكة المركزية من أن تكون افتراضية ووضعها مع محطة قاعدية واحدة أو أكثر. ومع ذلك، فإن هذا يجلب أيضاً تحديات جديدة، لا سيما فيما يتعلق بتحديد الموقع الأمثل لوظائف الشبكة المركزية وفقاً للهندسة المادية وكمية البيانات التي سيتم إرسالها من كل محطة قاعدية إلى الشبكة المركزية. في هذه الأطروحة، نتناول هذه المشكلة على وجه الخصوص، ونقترح نموذجاً عاماً وفعالاً يسمح بتحديد أفضل محطة قاعدية لوضع وظائف الشبكة المركزية، مع تحسين مقياس الأداء المطلوب (التدفق الإجمالي، مقاييس التكافؤ بين المحطات القاعدية، الخ). يعتمد حلنا على نمذجة مشتركة للشبكة كبرنامج غير خطي متعدد الأعداد الصحيحة المختلطة (MINLP) وعلى عملية التحديد الخطي الفعال للقيود غير الخطية. نموذجنا يعتمد على صيغة عامة، مما يسمح بسهولة تعديل وضبط المقاييس المراد تحسينها و/أو إضافة أو تعديل القيود حسب الحاجة.

باعتبار تعظيم التدفق الإجمالي، قمنا أولاً بتقييم أداء نموذجنا من خلال مقارنته بالحلول المماثلة المقترحة مسبقاً من طرف الباحثين. تُظهر النتائج أن نهجنا المشترك (على دفعة واحدة) يقلل بشكل كبير من وقت التنفيذ مقارنةً بالحلول الحالية، مع تقديم نفس النتائج المثلى. بناءً على صيغتنا العامة، تناولنا بعد ذلك مشكلة التكافؤ بين المحطات القاعدية. في الواقع، تُظهر النتائج أن تعظيم التدفق الإجمالي يسبب اختلالاً كبيراً من حيث الرضا المطلق والنسبي فيما يتعلق بطلب المحطات القاعدية المختلفة. لذلك اقترحنا متغيرات فعالة لنموذجنا تضمن إيجاد المكان الأمثل للشبكة المركزية مع ضمان التكافؤ بين المحطات القاعدية. أجرينا العديد من عمليات المحاكاة لتقييم الحلول المثلى التي تم الحصول عليها وأداء حلولنا التي تأخذ بعين الاعتبار التكافؤ بين المحطات القاعدية. مكنتنا نتائج المحاكاة من مناقشة المفاضلة بين تعظيم التدفق الإجمالي والتكافؤ بين المحطات القاعدية، مع تسليط الضوء على فوائد نهج النمذجة المشتركة (ذات الدفعة الواحدة). أخيراً، درسنا بمزيد من التفصيل توسيع نطاق نماذجنا فيما يخص عدد المحطات القاعدية وعدد الشبكات المركزية التي سيتم وضعها ومقاييس التحسين التي يجب أخذها بعين الاعتبار. تُظهر النتائج أن نماذجنا توفر أوقات تنفيذ معقولة جداً حتى في حالة الشبكات الكبيرة التي تضم مئات المحطات القاعدية.

الكلمات المفتاحية:

شبكات الانتشار الذاتي، الشبكات الخلوية، المحاكاة الافتراضية للشبكة المركزية، مشكلة الموضع الأمثل، برمجة خطية،

تعظيم التدفق الإجمالي، التكافؤ.

Table des matières

Introduction générale	15
1 Réseaux cellulaires auto-déployables et positionnement de nos travaux	17
1.1 Contexte général : réseaux cellulaires auto-déployables	18
1.1.1 Réseaux cellulaires et leur évolution des premiers systèmes ana- logiques aux futurs réseaux de sixième génération	18
1.1.2 Émergence des drones et des robots	22
1.1.3 Émergence de la virtualisation	24
1.1.4 Réseaux cellulaires auto-déployables	26
1.2 Objectifs de la thèse, verrous traités et aperçu des contributions	28
1.2.1 Verrous traités et objectifs de la thèse	28
1.2.2 Résumé des contributions	30
1.3 Structure détaillée du document	30
2 État de l’art sur les réseaux auto-déployables et le placement optimal	33
2.1 Conception et expérimentation des réseaux cellulaires auto-déployables	34
2.2 Placement optimal du réseau cœur dans les réseaux cellulaires auto- déployables	36
2.3 Placement optimal des stations de base aériennes	37
2.4 Placement optimal dans les réseaux conventionnels	38
2.5 Placement optimal dans les environnements NFV	41
2.6 Placement optimal dans le Mobile Edge Computing	43
2.7 Conclusions	46
3 De nouveaux modèles pour le placement optimal des réseaux cœur	47
3.1 Notations et énoncé du problème	48
3.2 Modèles itératifs de la littérature	50
3.3 Notre modèle optimisant le placement du réseau cœur en maximisant les flux	52
3.3.1 Premier modèle intuitif (non-linéaire)	53

3.3.2	Linéarisation des contraintes de notre modèle	54
3.4	Extensions du modèle	56
3.5	Conclusions	58
4	Évaluation de performances des modèles proposés	59
4.1	Configuration et paramètres des simulations	60
4.2	Validation et évaluation des performances dans le contexte homogène des flux initiaux des stations de base	61
4.3	Validation et évaluation des performances dans le contexte hétérogène des demandes de flux et des flux initiaux	63
4.3.1	Évaluation de performances en fonction de la taille du réseau	64
4.3.2	Évaluation de performances en fonction de la connectivité du ré- seau	66
4.4	Conclusions	67
5	Gestion de l'équité entre les stations de base	69
5.1	Motivations	70
5.2	Métriques prises en compte	71
5.3	Modèles optimaux d'équité	72
5.3.1	Maximiser la satisfaction proportionnelle moyenne	72
5.3.2	Minimiser la variance des satisfactions proportionnelles	72
5.3.3	Maximiser la satisfaction proportionnelle minimale	73
5.4	Évaluation des modèles optimaux d'équité	74
5.4.1	Évaluation des performances en termes de métriques de flux et d'équité	74
5.4.2	Évaluation du temps d'exécution	76
5.5	Conclusions	78
6	Étude de la scalabilité des modèles de placement optimal	79
6.1	Scalabilité des modèles maximisant le flux selon le nombre de stations de base	80
6.2	Scalabilité des modèles maximisant le flux selon le nombre de réseaux cœur	82
6.3	Étude de la scalabilité en fonction des métriques d'optimisation	86
6.4	Conclusions	88
	Conclusions et perspectives	89
	Résumé des contributions	89
	Perspectives et travaux futurs	91

Liste des tableaux

2.1	Classification représentative des travaux étudiés	46
3.1	Principales notations utilisées	49
4.1	Paramètres de simulations par défaut	61
4.2	Résultats de la comparaison entre le modèle itératif basé sur LP et notre modèle, en termes de flux maximal, de temps d'exécution et du meilleur emplacement du réseau cœur (moyennes sur 30 itérations - cas des flux initiaux homogènes)	62
4.3	Résultats de la comparaison du modèle itératif basé sur LP, du modèle itératif basé sur Min-Cut et de notre modèle, en termes de flux maximal et de temps d'exécution en fonction de n (moyennes sur 30 itérations - cas hétérogène)	64
4.4	Résultats de la comparaison du modèle itératif basé sur LP, du modèle itératif basé sur Min-Cut et de notre modèle en termes de flux maximal et de temps d'exécution en fonction de ρ (moyennes sur 30 itérations - cas hétérogène)	67
5.1	Exemple de trafic optimal avec maximisation du flux global dans le réseau ($\rho = 0.3$ et $n = 10$). La position optimale du réseau cœur est la station 5 dans cet exemple	70
5.2	Résultats de la comparaison du temps d'exécution des quatre modèles en fonction de la taille du réseau n (moyennes sur 100 itérations)	77
6.1	Étude de la scalabilité en fonction de la taille du réseau : comparaison entre le modèle itératif basé sur LP, le modèle itératif basé sur Min-Cut et notre modèle (moyennes sur 30 itérations - cas hétérogène)	81
6.2	Étude de scalabilité en fonction de ρ ($n = 500$ - moyennes sur 50 itérations)	81
6.3	Étude de la scalabilité en fonction du nombre de réseaux cœur et du nombre de nœuds ($\rho = 0.2$ - moyennes sur 50 itérations)	83
6.4	Flux maximal en fonction du nombre de nœuds et de réseaux cœur	85

6.5	Temps d'exécution moyen des quatre modèles (Max Flow, Max Avg. Satisf., Min Norm. Var., Max-Min Satisf.) en fonction du nombre de nœuds	86
-----	---	----

Table des figures

1.1	Architecture simplifiée d'un réseau cellulaire 4G	20
1.2	Trois piliers de l'émergence des réseaux cellulaires auto-déployables . . .	26
3.1	Exemple d'un réseau auto-déployable avec 10 stations de base interconnectées et un réseau cœur local co-localisé avec une des stations de base	48
4.1	Temps moyen d'exécution (en secondes) en fonction du nombre de nœuds n (moyennes sur 30 itérations - cas des flux initiaux homogènes)	63
4.2	Temps moyen d'exécution (en secondes) en fonction du nombre de nœuds n (moyennes sur 30 itérations - cas hétérogène)	65
4.3	Temps moyen d'exécution (en secondes) en fonction du rayon de connectivité ρ ($n = 50$ - cas hétérogène)	66
5.1	Comparaison des quatre modèles (Max Flow, Max Avg. Satisf., Min Norm. Var., Max-Min Satisf.) en fonction des quatre métriques correspondantes	75
5.2	Temps moyen d'exécution (en secondes) des quatre modèles en fonction de la taille du réseau (moyennes sur 100 itérations)	77
6.1	Temps moyen d'exécution (en secondes) pour un nombre de nœuds, n , allant de 50 à 400 ($\rho = 0.2$ - moyennes sur 30 itérations)	80
6.2	Temps moyen d'exécution (en secondes) en fonction du nombre de nœuds n et du rayon de connectivité ρ	82
6.3	Temps moyen d'exécution (secondes) en fonction du nombre de nœuds n pour différents nombres de réseaux cœur ($\rho = 0.2$ - moyennes sur 50 itérations)	83
6.4	Temps moyen d'exécution (en secondes) en fonction du nombre de nœuds n et du nombre de réseaux cœur ($\rho = 0.2$ - moyennes sur 50 itérations) .	84
6.5	Flux maximal en fonction du nombre de nœuds et de réseaux cœur . . .	85
6.6	Temps moyen d'exécution (en secondes) des quatre modèles en fonction de la taille du réseau (nombre de nœuds)	87

Liste des Abréviations

1G	Première génération de réseaux cellulaires
2G	Deuxième génération de réseaux cellulaires
3G	Troisième génération de réseaux cellulaires
3GPP	3rd Generation Partnership Project
4G	Quatrième génération de réseaux cellulaires
5G	Cinquième génération de réseaux cellulaires
6G	Sixième génération de réseaux cellulaires
ABSs	Aerial Base Stations - Stations de base aériennes
BS	Base Station - Station de base
BSC	Base Station Controller - Contrôleur de station de base
BTS	Base Transceiver Station - Station de base dans les réseaux 2G
CN	Core Network - Réseau cœur
CNF	Containers Network Functions
CPU	Central Processing Unit - Unité centrale de traitement
DSs	Dominating Sets - Ensembles dominants
EPC	Evolved Packet Core - réseau cœur dans les réseaux 4G
eNodeB	evolved Node B - Nom donné aux stations de base intelligentes dans les réseaux 4G
GSM	Global System for Mobile Communications
ILP	Integer Linear Programming - Programmation linéaire en nombres entiers
IoT	Internet of Things - Internet des objets
LP	Linear Programming - Programmation Linéaire
LTE	Long Term Evolution
MEC	Mobile Edge Computing
MME	Mobility Management Entity - Entité de gestion de mobilité
MINLP	Mixed Integer NonLinear Program - Programme non linéaire mixte en nombres entiers
MILP	Mixed Integer Linear Program - Programme linéaire mixte en nombres entiers

MIQP	Mixed-Integer Quadratic Programming
MIMO	Multiple-Input Multiple-Output
NFV	Network Functions Virtualization - Virtualisation des fonctions réseau
NS	Network Services - Services réseau
P-GW	Packet data network GateWay
Qos	Quality of Service - Qualité de service
RAN	Radio Access Network - Réseau d'accès radio
RNC	Radio Network Controller
SDN	Software Defined Networking - Réseau définie par logiciel
S-GW	Serving GateWay
SMS	Short Message Service - Service de messagerie courte
UAV	Unmanned Aerial Vehicle - Véhicule aérien sans pilote
UE	User Equipment - Équipement utilisateur
UMTS	Universal Mobile Telecommunications System
UPF	User Plane Function - Fonction du plan utilisateur
WMN	Wireless Mesh Networks - Réseaux maillés sans fil
WSN	Wireless Sensor Networks - Réseaux de capteurs sans fil

Introduction générale

Le domaine des réseaux cellulaires a connu des évolutions majeures ces dernières années. D'une part, le développement des systèmes mobiles tels que les robots et les drones a ouvert la voie à l'intégration de stations de base voire du réseau cœur sur ces unités mobiles [1, 2, 3, 4, 5, 6]. D'autre part, l'émergence de la virtualisation des fonctions de réseau (NFV, Network Function Virtualization) et la mise en réseau définie par logiciel (SDN, Software-Defined Networking) sont exploitées de plus en plus depuis les réseaux 4G offrant une plus grande flexibilité dans le réseau cœur [7, 8]. Ces avancées, combinées aux capacités d'auto-organisation de normes récentes, ont donné naissance au concept de réseau cellulaire auto-déployable, permettant un déploiement plus souple et plus rapide des réseaux cellulaires [9, 10, 11]. Ce concept a une importance particulière notamment avec l'émergence des réseaux privés et le rôle prévu des drones dans la future normalisation des réseaux cellulaires 6G [12, 13].

Le concept de réseau cellulaire auto-déployable [9, 10] trouve des applications très variées comme l'extension de la couverture cellulaire, la gestion des événements très fréquentés et la restauration des communications lors de situations d'urgence. Dans ce dernier cas à titre d'exemple, les réseaux traditionnels sont souvent hors service après une catastrophe naturelle ou humaine. Les réseaux auto-déployables peuvent offrir une solution pratique pour rétablir rapidement les communications entre les individus et les premiers intervenants (secours) [9, 14, 15].

Les réseaux auto-déployables permettent un déploiement rapide et facile des stations de base, qu'elles soient connectées à un réseau traditionnel pour étendre la couverture ou déployées de manière autonome [4, 16]. Dans cette thèse, nous nous concentrons sur le déploiement autonome où aucune liaison vers un réseau traditionnel n'est nécessaire, et où toutes les fonctionnalités du réseau cœur peuvent être regroupées et co-localisées avec une ou plusieurs stations de base. Cela forme ce qu'on appelle un réseau cœur (CN, Core Network) local [9, 10].

Malgré les différents avantages des réseaux auto-déployables, ces derniers amènent de nouveaux défis scientifiques. L'un de ces défis concerne le placement optimal des réseaux cœur locaux partagés. En effet, cette tâche doit être calculée rapidement tout en essayant d'atteindre différents objectifs tels que la maximisation du flux global de trafic ou la garantie d'une équité entre les stations de base par exemple.

C'est précisément le cœur de cette thèse : l'étude du placement efficace des réseaux cœur locaux dans un réseau cellulaire auto-déployable. Pour cela, nous avons d'abord proposé un modèle générique qui permet de localiser les réseaux cœur tout en optimisant une métrique donnée de performance. Nous avons modélisé le problème sous forme d'un programme non linéaire mixte en nombres entiers, puis linéarisé les contraintes pour améliorer l'efficacité en termes de temps d'exécution. Nous avons validé et évalué notre modèle et comparé ses performances à celles des solutions de la littérature. Nous avons ensuite considéré un autre défi important concernant l'équité entre les stations de base. Pour cela, nous avons proposé plusieurs variantes de notre modèle et évalué et discuté les résultats en termes de métriques de flux de trafic et d'équité et en termes de temps de calcul. Nous avons enfin étudié la scalabilité (passage à l'échelle) de nos modèles en fonction de plusieurs paramètres.

Nous commençons ce document par discuter dans le chapitre 1 l'évolution des réseaux cellulaires durant les quarante dernières années; l'émergence des robots, des drones et de la virtualisation; le concept des réseaux cellulaires auto-déployables et enfin le positionnement de nos travaux. Nous procédons ensuite dans le chapitre 2, à une revue de la littérature, en discutant d'abord plus en détails les réseaux auto-déployables et les travaux expérimentant ces réseaux émergents. Ensuite, nous nous concentrons sur les travaux traitant le problème du placement optimal, en se focalisant sur le placement optimal du réseau cœur. Nous examinons également quelques recherches concernant le placement optimal dans d'autres types de réseaux conventionnels, dans le domaine de virtualisation des fonctions réseaux (NFV) et dans le domaine de l'informatique mobile en périphérie (MEC, Mobile Edge Computing).

Dans le chapitre 3, nous présentons en détail notre modèle de placement en exposant la formulation intuitive et non linéaire du problème et en discutant en détail la linéarisation des contraintes non linéaires et les avantages du modèle résultant. Nous dédions ensuite le chapitre 4 à l'évaluation des performances du modèle proposé et à la comparaison par rapport aux solutions de la littérature. Nous discutons les résultats en fonction de la taille du réseau et du degré de connectivité, tout en considérant différents scénarios et contextes.

Dans le chapitre 5, nous traitons le problème d'équité entre les stations de base et présentons plusieurs variantes de nos modèles visant à assurer une équité dans le réseau. Nous discutons sur la base des résultats le compromis entre la maximisation du flux de trafic et l'équité, tout en mettant en avant les avantages de nos modèles. Dans le chapitre 6, nous examinons la scalabilité de nos modèles en fonction du nombre de stations de base, du nombre de réseaux cœur et selon les métriques d'optimisation. Enfin, nous dressons, dans les conclusions générales, un bilan des contributions et des résultats et discutons quelques travaux futurs que nous souhaitons explorer.

Chapitre 1

Réseaux cellulaires auto-déployables et positionnement de nos travaux

Contenu du chapitre

1.1	Contexte général : réseaux cellulaires auto-déployables	18
1.1.1	Réseaux cellulaires et leur évolution des premiers systèmes analogiques aux futurs réseaux de sixième génération	18
1.1.2	Émergence des drones et des robots	22
1.1.3	Émergence de la virtualisation	24
1.1.4	Réseaux cellulaires auto-déployables	26
1.2	Objectifs de la thèse, verrous traités et aperçu des contributions . . .	28
1.2.1	Verrous traités et objectifs de la thèse	28
1.2.2	Résumé des contributions	30
1.3	Structure détaillée du document	30

Dans ce chapitre introductif, nous allons tout d’abord présenter le contexte général de notre thèse en discutant brièvement l’évolution des réseaux cellulaires, l’émergence des robots et des drones et leur place dans le monde des réseaux cellulaires ainsi que l’émergence de la virtualisation. Cela nous permettra de bien positionner et introduire le contexte plus spécifique de notre thèse lié aux réseaux cellulaires auto-déployables. Nous présenterons par la suite les objectifs de la thèse tout en discutant principalement le verrou du placement optimal du réseau cœur dans un réseau auto-déployable. Nous finissons ce chapitre introductif par un résumé de nos contributions qui seront détaillées dans les chapitres suivants.

1.1 Contexte général : réseaux cellulaires auto-déployables

Nous traitons dans cette thèse des problématiques liées aux réseaux cellulaires auto-déployables. Avant d'introduire le concept de ces réseaux, nous commençons tout d'abord par présenter les réseaux cellulaires classiques et leur évolution depuis les réseaux de première génération. Nous discutons aussi la place que les réseaux de drones et de robots ainsi que la virtualisation prennent au sein des réseaux cellulaires au cours des dernières années.

1.1.1 Réseaux cellulaires et leur évolution des premiers systèmes analogiques aux futurs réseaux de sixième génération

Un réseau cellulaire est un type de réseau de communication sans fil dans lequel des appareils mobiles communiquent entre eux. Initialement, ces appareils mobiles se limitaient aux téléphones portables et le principal service assuré était la transmission de la voix. Les réseaux cellulaires ont depuis beaucoup évolué à la fois en termes de types de dispositifs à connecter (tablettes, objets connectés, véhicules connectés, etc.) mais aussi en termes de services offerts tels que l'accès à Internet, la messagerie multi-média ou le streaming vidéo, à titre d'exemples. Dans un réseau cellulaire, le premier équipement permettant de relier les dispositifs mobiles à l'infrastructure centrale est appelé une station de base. Les stations de base sont réparties dans des zones géographiques appelés "cellules" d'où le terme "réseau cellulaire". Chaque cellule est ainsi desservie par une station de base. Cette organisation permet la réutilisation des ressources radio sur plusieurs zones géographiques ce qui offre une couverture étendue et une bonne répartition de trafic [17, 18, 19, 20, 21, 22].

L'évolution des réseaux cellulaires a connu plusieurs générations, depuis les premiers systèmes analogiques jusqu'aux technologies plus avancées comme les réseaux de cinquième génération (5G) et prochainement les réseaux de sixième génération (6G). Chaque génération apporte des améliorations significatives en termes de vitesse de transmission des données, de capacité du réseau, de latence et de nombre d'appareils qui peuvent être connectés simultanément. Cela a aussi progressivement ouvert la voie à de nouvelles applications et services mobiles innovants [17, 18, 19, 20, 21, 22]. Les normes implémentées dans ces différentes générations sont élaborées et mises à jour par une organisation appelée 3GPP (3rd Generation Partnership Project)¹. Cette dernière est une coopération entre plusieurs organismes de normalisation. Les différentes normes sont ensuite mises en oeuvre par les différents opérateurs de réseau mobile dans les différents pays.

1. <https://www.3gpp.org/>

Réseaux cellulaires 1G : la première génération des réseaux cellulaires a connu ses premiers grands déploiements dans les années 1980 pour la transmission de la voix. Cette génération se basait sur des communications analogiques pour la transmission des signaux vocaux sans aucune numérisation. De ce fait, les technologies utilisées offraient une faible capacité et une faible qualité de la voix. De même, ces technologies ne géraient pas l'itinérance entre différentes régions ou pays. Il est à noter que ces technologies se basaient exclusivement sur la commutation de circuit où une connexion physique est établie entre les deux parties communicantes et maintenue tout au long de l'appel [17].

Réseaux cellulaires 2G : la fin des années 1980 et le début des années 1990 ont connu l'émergence des réseaux de deuxième génération, avec une avancée technologique importante par rapport à la première génération. En effet, les réseaux de deuxième génération ont introduit la communication numérique pour transmettre la voix, ce qui a permis à la fois d'améliorer la qualité de la voix échangée ainsi que la capacité de ces réseaux cellulaires. En plus de cela, d'autres services ont été introduit comme le service de messagerie par texte (SMS). Il est à noter que les premières versions des technologies de deuxième génération comme le GSM (Global System for Mobile Communications) se basaient toujours sur la commutation de circuit [18].

Réseaux cellulaires 3G : le début des années 2000 a connu le développement des réseaux de troisième génération. Ces réseaux offrent des débits plus importants que les réseaux 2G et ont principalement permis de nouveaux services liés à la transmission des données comme l'accès à Internet, la téléphonie IP, le streaming des vidéos, la vidéo-conférence, etc. Les débits offerts par les réseaux 3G atteignent en moyenne quelques mégabits par seconde. Afin d'atteindre ces débits, les technologies de troisième génération comme UMTS (Universal Mobile Telecommunications System) utilisent des technologies de communication offrant une capacité plus importante. Par ailleurs, des communications par commutation de paquet ont été introduites dans cette génération, ce qui a permis une gestion plus efficace des ressources. Ainsi, les réseaux 3G incorporent deux domaines : le domaine dit "circuit" utilisé pour la transmission de la voix et le domaine dit "paquet" pour la transmission des données. Il est à noter que le développement des réseaux 3G a coïncidé avec l'apparition et l'émergence des "smartphones" permettant une navigation plus simple sur Internet, ce qui a encouragé le déploiement massif de ces réseaux tout en contribuant à leur succès [19].

Réseaux cellulaires 4G : les évolutions mentionnées ci-dessous ont ouvert la voie au développement des réseaux de quatrième génération par la suite. Après plusieurs années de normalisation, ces derniers ont été déployés à partir des années 2010 avec un

Chapitre 1. Réseaux cellulaires auto-déployables et positionnement de nos travaux

double objectif : i) l'augmentation des capacités grâce à l'utilisation de nouvelles technologies à très haute efficacité spectrale, et ii) la réduction de la latence (temps de bout en bout d'une transmission) ce qui est important pour les applications "temps réel". Les réseaux 4G permettent aussi de gérer un plus grand nombre d'appareils connectés simultanément avec des débits qui peuvent atteindre plusieurs dizaines de mégabits par seconde. Cela a permis d'améliorer l'expérience utilisateur et a surtout ouvert la voie à de nouvelles applications, comme les jeux en ligne, le streaming haute qualité, etc. [20]. La principale technologie déployée dans les réseaux de quatrième génération est la norme LTE (Long-Term Evolution). Il est important à noter que les standards de quatrième génération proposent une communication par commutation de paquet pour tout le trafic y compris la voix. Cela a permis entre autres de simplifier l'architecture du réseau cœur, comme nous l'aborderons par la suite [20]. Un exemple illustratif d'une architecture simplifiée d'un réseau cellulaire 4G est présenté dans la figure 1.1 et sera discuté par la suite.

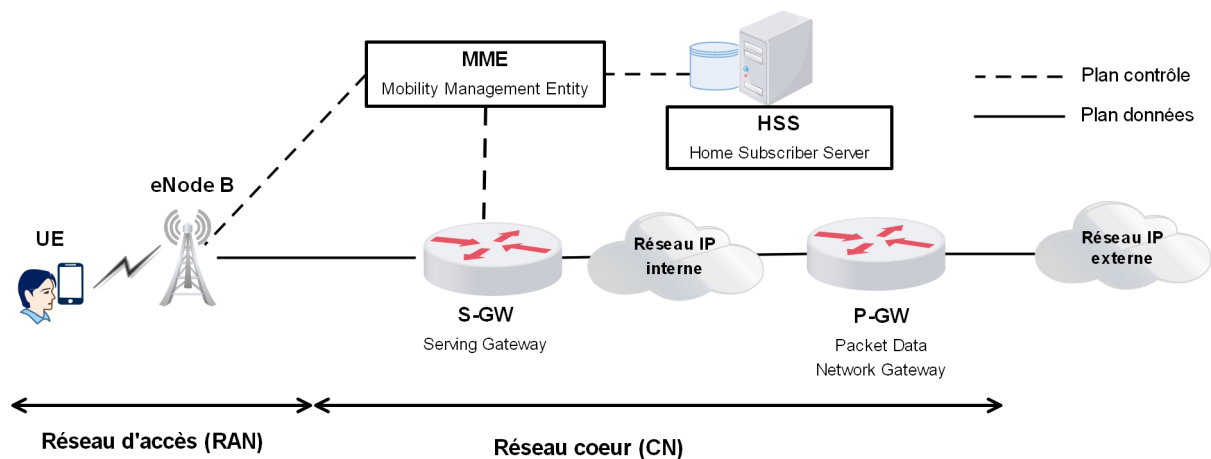


FIGURE 1.1 – Architecture simplifiée d'un réseau cellulaire 4G

Réseaux cellulaires 5G : même si les réflexions et les recherches ont débuté avant, les premiers déploiements des réseaux 5G ont commencé à travers le monde au début des années 2020. Ces réseaux offrent des débits extrêmement élevés atteignant des gigabits par seconde. Ces réseaux offrent aussi une latence extrêmement réduite. Les débits offerts par cette génération ont été obtenus grâce à différentes avancées et notamment l'utilisation du "Massive MIMO (Multiple-Input Multiple-Output)", une technologie qui consiste à utiliser un grand nombre d'antennes pour augmenter à la fois la capacité du réseau et la couverture. Ces réseaux ont aussi bénéficié du slicing, un concept qui permet de diviser virtuellement une infrastructure réseau en plusieurs réseaux logiques, offrant ainsi une meilleure utilisation personnalisée et plus efficace

des ressources. Toutes ces avancées ont permis de pénétrer encore de nouveaux domaines nécessitant des délais très réduits et/ou un nombre très important de dispositifs comme la télé-médecine, la réalité virtuelle, la réalité augmentée, les véhicules autonomes, et différents applications de l'Internet des objets (IoT) [21, 23, 24, 25]. Même si la virtualisation des fonctions réseau a commencé dans les réseaux 4G, elle s'est largement répandue dans les standards des réseaux 5G. Nous reviendrons plus tard avec un peu plus de détails à cette notion de virtualisation des fonctions réseau.

Vers des réseaux cellulaires 6G : les recherches se poursuivent aujourd'hui dans un objectif de développer une nouvelle génération de réseau (6G). Même si les standards ne sont pas encore établis, les travaux en cours montrent que ces réseaux offriront encore une meilleure capacité, une latence plus réduite avec une connectivité omniprésente (une plus large couverture). Les recherches actuelles montrent aussi que ces réseaux bénéficieront entre autres des avancées réalisées dans le domaine de l'intelligence artificielle à différents niveaux [22, 26, 27, 28]. A titre d'exemple, Saad et al. présentent dans [26] une vision des systèmes 6G tout en identifiant les applications potentiels de ces systèmes. Ils proposent un nouvel ensemble de classes de services pour la 6G. Chafii et al. identifient plus récemment dans [22] 12 verrous technologiques que les réseaux 6G doivent adresser.

Les réseaux cellulaires classiques sont souvent organisés en deux parties : le **réseau d'accès** (Radio Access Network) et le **réseau cœur** (Core Network)² (voir par exemple l'architecture simplifiée d'un réseau 4G présentée en figure 1.1). Même si les noms des entités et des fonctions changent régulièrement d'une technologie à une autre, les principales fonctions du réseau d'accès et du réseau cœur restent similaires. En effet, le **réseau d'accès**, fournit une interface entre les dispositifs mobiles (terminaux) et le réseau cœur. Pour cela, il gère l'établissement et le maintien de la connexion, l'attribution et l'utilisation des ressources radio, la transmission et la réception des signaux radio entre les dispositifs mobiles et les stations de base et la mise en place des fonctions de sécurité (comme le chiffrement à partir des réseaux 2G et l'intégrité à partir des réseaux 3G) sur le lien sans fil. Il gère également la mobilité des dispositifs mobiles lorsqu'ils se déplacent dans une zone de plusieurs cellules appelée "zone de localisation". De son côté, le **réseau cœur** gère plusieurs fonctions indispensables au bon fonctionnement du réseau comme la gestion des appels et des sessions de données, la gestion de la mobilité au delà des zones de localisation, le routage et le transfert des paquets, l'authentification des utilisateurs et la facturation. Ce réseau cœur assure également une fonction très importante, à savoir l'interconnexion avec les réseaux externes comme les réseaux téléphoniques fixes et Internet [9, 18, 19, 20, 29].

2. L'équipement terminal est parfois inclus dans le réseau d'accès ou exclu comme une entité séparée

D'un point de vue architecturale, le **réseau d'accès** inclut dans les réseaux 2G et 3G i) les stations de base (appelées respectivement BTS et NodeB) qui sont les équipements passifs du réseau d'accès et ii) des contrôleurs de stations de base (appelés respectivement BSC pour Base Station Controller et RNC pour Radio Network Controller) et qui forment la partie intelligente du réseau d'accès. Depuis les réseaux 4G, cette architecture est simplifiée en proposant des stations de base intelligentes appelées eNodeB (evolved Node B) [18, 19, 20].

De l'autre côté, le **réseau cœur** comprend dans les réseaux de deuxième génération des commutateurs et différentes bases de données appelées registres pour gérer l'authentification, la mobilité, la facturation, etc. Les réseaux 3G ont introduit la commutation de paquet ce qui a impliqué la duplication des équipements. Le passage à une commutation exclusive de paquets à partir des réseaux 4G a permis de simplifier l'architecture du réseau cœur. Dans l'architecture d'un réseau cœur 4G, par exemple, un équipement appelé MME (Mobility Management Entity) est responsable de la signalisation pour la mobilité des utilisateurs et de la gestion des sessions (voir la figure 1.1). Des routeurs appelés S-GW (Serving Gateway) acheminent les paquets dans le réseau, tandis que d'autres routeurs, appelés P-GW (Packet Data Network Gateway), assurent le lien avec les réseaux externes [18, 19, 20] (voir la figure 1.1).

1.1.2 Émergence des drones et des robots

En parallèle du développement des réseaux cellulaires, le domaine de la robotique a connu ces dernières années des progrès remarquables grâce aux avancées technologiques dans différents axes. Parmi ceux-ci, on peut citer la mobilité autonome, la collaboration accrue, l'intelligence artificielle, l'automatique, etc. Nous abordons dans ce qui suit les robots de manière générale et les drones en particulier et leur place dans le domaine des réseaux cellulaires.

Les robots sont des machines programmables conçues pour effectuer des tâches précises de manière autonome ou semi-autonome. Les robots peuvent aussi interagir avec l'environnement en utilisant des éléments qu'ils intègrent tels que les capteurs, les actionnaires et des systèmes de contrôle. La communication et la coordination entre robots est un sujet d'actualité et d'intérêt majeur [30]. Les robots sont de plus en plus utilisés dans différents domaines. Dans le cas des réseaux cellulaires, ils peuvent être utilisés pour différentes tâches comme la maintenance par exemple. En effet, ils peuvent permettre l'inspection et l'entretien des équipements de télécommunication grâce à des caméras embarquées par exemple. Néanmoins l'application qui nous intéresse le plus dans le cadre de notre thèse concerne l'utilisation des robots qui embarquent ou qui transportent des équipements de réseaux cellulaires, comme les antennes, des stations

de base ou le réseau cœur. Comme nous le discuterons plus tard, cela pourra avoir un intérêt indéniable dans plusieurs situations comme les cas d'urgence suivant des accidents majeurs ou des catastrophes naturelles, où les infrastructures de télécommunication sont endommagées.

De leur côté, les véhicules aériens sans pilote (UAV) [31, 32, 33], le plus souvent connus par le nom des "drones", sont des aéronefs sans pilote qui peuvent voler de façon autonome ou être pilotés à distance. Ils sont souvent considérés comme une sous catégorie des robots et ils sont principalement utiles pour accomplir des tâches difficiles à effectuer par l'être humain et/ou dans des endroits inaccessibles par l'être humain [31, 32, 33]. De nos jours, les drones ont connu des avancées remarquables en ce qui concerne leur miniaturisation, la charge à porter, l'autonomie de vol, etc. Il est néanmoins important à noter que l'utilisation des drones est très réglementée et reste relativement contraignante dans différents pays du monde. En effet, beaucoup de pays mettent en place des réglementations et des restrictions en ce qui concerne l'utilisation des drones tel que la limitation de vol des drones sur certaines zones, les hauteurs autorisées, l'obligation de permis de pilotage, etc.

Les drones trouvent aujourd'hui énormément d'applications dans des domaines très variés [34, 35, 36]. Parmi ces derniers, nous pouvons citer par exemple la gestion des infrastructures civiles [37, 38], les opérations de recherche et de sauvetage [39, 40], le suivi environnemental [41], la livraison de marchandises [42], la sécurité et la surveillance [43], l'agriculture de précision [44], la collecte d'image [45], la collecte de données de capteurs [46], etc.

D'un point de vue réseau, plusieurs travaux ont traité les verrous de communications dans les réseaux de robots et les réseaux de drones i) qu'ils soient utilisés pour d'autres applications nécessitant une communication des données ou une coordination ou ii) qu'il soient utilisés spécifiquement pour étendre ou pour déployer un réseau sans fil [1, 47, 48]. Plus spécifiquement, les drones ouvrent de nouvelles perspectives dans le domaine des réseaux cellulaires. Ces derniers peuvent être utilisés soit pour étendre la couverture d'un réseau existant, soit pour déployer un réseau cellulaire autonome [4, 5, 6, 49, 50, 51, 52]. En effet, les drones peuvent permettre l'amélioration de la couverture dans les zones isolées où l'installation d'une infrastructure terrestre est coûteuse ou difficile. L'augmentation de la capacité dans des zones encombrées lors des événements à forte fréquentation constitue aussi un cas d'usage intéressant où cette technologie offre une flexibilité importante et un intérêt indéniable. Enfin, les drones permettent aussi le déploiement rapide visant à remplacer temporairement un réseau classique endommagé lors des catastrophes naturels ou humaines [1, 9, 47, 49, 50, 51, 52].

En ce qui concerne les équipements et les architectures des réseaux cellulaires, l'utilisation la plus répandue des drones consiste à embarquer des stations de base cellulaires [50, 51, 52] (eNodeB 4G par exemple), ce qui est envisageable dans la majorité des applications que nous venons de citer. Au delà d'embarquer les stations de base ou en complément, les drones peuvent être aussi utilisés en tant que équipement utilisateur (UE 4G par exemple) pour accomplir différentes tâches tel que, la surveillance et la photographie dans les situations de catastrophes, dans les événements sportifs, etc. Ils peuvent aussi être utilisés en tant que nœuds relais déployés à distance afin de recueillir les informations et de les transmettre à une station de base terrestre proche.

Au delà de ces utilisations ou en complément, les drones peuvent aussi embarquer une partie ou l'entièreté du réseau cœur. Cela est de plus en plus envisageable grâce à la virtualisation et à la flexibilité offerte par les derniers standards des réseaux cellulaires comme nous le discuterons plus tard dans cette thèse. Dans la littérature, plusieurs travaux ont récemment considéré l'intégration des drones dans les réseaux cellulaires sous différentes formes [1, 2, 3, 4, 5, 6, 9, 10, 49, 50, 51, 52, 53]. Dans notre cas d'étude, nous nous intéressons plus particulièrement au déploiement autonome de réseaux cellulaires où les drones ou les robots sont utilisés pour embarquer des stations de base [1, 2, 50, 51, 52] et/ou le réseau cœur [3, 4, 5, 6, 9, 10]. Nous discuterons quelques travaux représentatifs de cet axe plus en détail dans le prochain chapitre.

1.1.3 Émergence de la virtualisation

De manière générale, la virtualisation est une technologie qui permet à plusieurs instances virtuelles de ressources informatiques de fonctionner sur un seul système matériel physique. Cela permet donc une économie de l'espace, de l'énergie et des coûts de maintenance ainsi qu'une gestion plus rapide. Cela permet aussi l'ajustement rapide et facile des fonctions et des ressources avec une économie des coûts. Les ressources en question peuvent être des serveurs, des dispositifs de stockage, des dispositifs réseau ou même des systèmes d'exploitation entiers.

Dans le domaine des réseaux, en particulier, les fonctionnalités "réseau" sont traditionnellement assurées par des boîtiers matériels dédiés (commutateurs, routeurs, etc.) qui sont intégrés statiquement au réseau pour fournir des services. Mais avec l'augmentation et la variété des boîtiers ainsi que l'augmentation des utilisateurs réseau, la gestion du réseau est devenue difficile. Pour résoudre ce problème, la virtualisation des fonctions réseau (VNF) a émergé en permettant l'implémentation des fonctions réseau du matériel dédié sous forme de logiciel afin de les exécuter sur des serveurs standard tout en répondant aux besoins spécifiques d'un réseau donné. La virtualisation permet ainsi une bonne gestion plus efficace des ressources informatiques, un déploiement et

une gestion plus facile des fonctions réseau et une automatisation de la mise en place des services réseau à la demande. La virtualisation permet au final d'avoir des infrastructures optimisées avec une plus grande efficacité et flexibilité ce qui constitue un atout important dans les infrastructures modernes [54, 55, 56, 57].

Dans le même objectif commun de transformer et de moderniser les infrastructures réseau pour les rendre plus agiles, flexibles et rentables, nous assistons depuis quelques années à l'émergence de la mise en réseau définie par logiciel, plus connue sous son dénomination en anglais Software Defined Networking (SDN). Le SDN est une approche de gestion des réseaux qui permet de les contrôler et de les gérer à partir d'un logiciel centralisé, plutôt que de configurer chaque équipement individuellement. Cela simplifie la gestion du réseau et permet une plus grande flexibilité. Cela permet aussi de séparer le plan de contrôle et le plan de données en orientant le plan de contrôle vers le logiciel et en laissant le plan de données dans le matériel. Ainsi, contrairement aux réseaux traditionnels, où les plans de contrôle et de données sont généralement mis en oeuvre au sein des dispositifs de réseau eux-mêmes (commutateurs, routeurs, etc.), le plan de contrôle dans un SDN est découplé et déplacé vers un contrôleur centralisé, qui a une vue globale de la topologie du réseau et des flux de trafic. Le SDN offre aux administrateurs la possibilité de superviser le réseau, de modifier ses paramètres de configuration et d'accroître sa capacité, le tout à partir d'une interface utilisateur centralisée [54, 58, 59, 60].

En particulier, les réseaux cellulaires ont bénéficié de ces progrès dans les domaines de la virtualisation des fonctions réseau (NFV) et de la mise en réseau définie par logiciel (SDN)[61, 62, 63, 64, 65]. Dans les réseaux LTE (4G) par exemple, on trouve que le réseau cœur, appelé EPC, a été traditionnellement basé sur des équipements matériels propriétaires. Cela présente des limitations en termes de flexibilité et de coûts, car la moindre modification ou adaptation du réseau nécessite souvent le remplacement de ces équipements. La virtualisation des fonctions réseau (NFV), permet de transformer l'EPC en remplaçant les équipements matériels par des solutions logicielles flexibles et évolutives, ce qui rend le réseau plus flexible, et permet de faciliter les mises à jour logicielles pour le lancement rapide de nouveaux services [56, 66].

De manière générale, les technologies NFV et SDN sont largement considérées comme des principaux facilitateurs dans l'architecture, la conception, l'exploitation et la gestion des réseaux cellulaires depuis les réseaux 4G et de manière plus prononcée dans les réseaux 5G et dans les futures standardisations à venir des réseaux 6G [59, 63, 67, 68, 69]. Cependant leurs utilisations amènent de nouveaux défis scientifiques et techniques dont certains sont liés au placement optimal des fonctions réseau et des contrôleurs et qui seront discutés dans le chapitre suivant.

Il est enfin important de mentionner que peu de travaux dans la littérature adressent l'utilisation de SDN et NFV dans le contexte de robots ou de drones de manière générale et en particulier dans le cas des réseaux cellulaires [70]. Motivé par ce constat, les auteurs de [70], proposent une discussion sur les technologies SDN et NFV et leurs utilisations dans le cadre de réseaux de drones. Une partie est dédiée au cas particulier des réseaux de drones pour les communications cellulaires, avec les défis de recherche ouverts dans ce cadre (énergie, durée de vol, couverture 3D, déploiement des UAV dans des environnements 3D, etc.).

1.1.4 Réseaux cellulaires auto-déployables

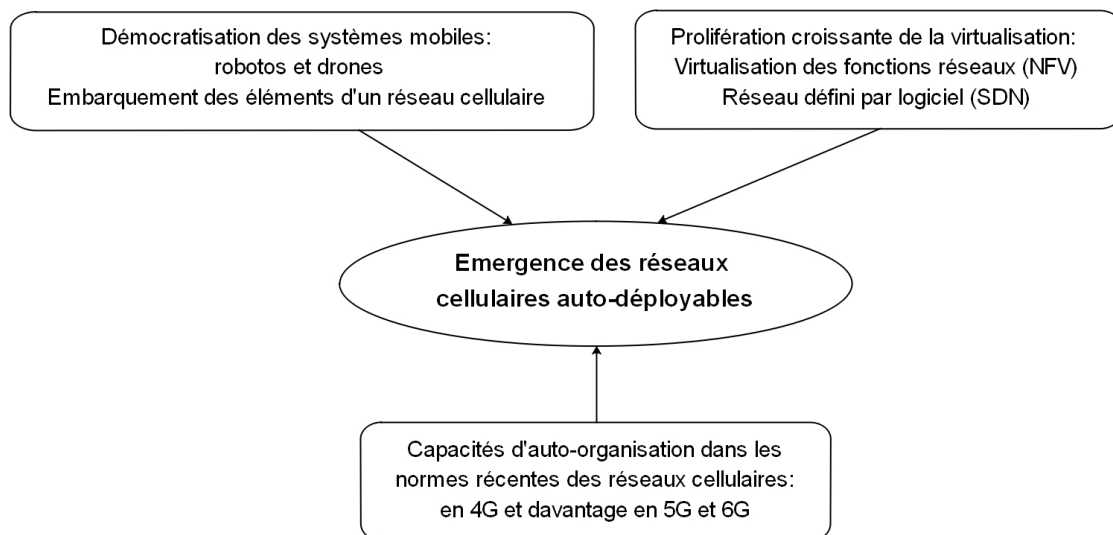


FIGURE 1.2 – Trois piliers de l'émergence des réseaux cellulaires auto-déployables

Comme discuté dans les deux sous sections précédentes, le domaine des réseaux cellulaires a largement bénéficié des évolutions majeures que nous venons de discuter. D'une part, la démocratisation des systèmes mobiles tels que les robots et les véhicules aériens sans pilote (UAV) a offert la possibilité d'embarquer des stations de base [1, 2, 50, 51, 52] ou même le réseau cœur [3, 4, 5, 9, 10] sur ces unités mobiles. D'autre part, la prolifération croissante de la virtualisation des fonctions de réseaux (NFV) et du réseau définie par logiciel (SDN) a offert plus de flexibilité dans le réseau cœur [7, 8, 16]. Comme illustré dans la figure 1.2, les avancées dans ces deux domaines, combinées aux capacités d'auto-organisation offertes par les normes récentes des réseaux cellulaires (5G en particulier et la futur norme 6G), ont donné naissance au concept de réseau cellulaire auto-déployable [9, 10], qui permet un déploiement plus souple et plus rapide des réseaux cellulaires [9, 10, 11].

Un réseau cellulaire auto-déployable [9, 10] peut prendre plusieurs formes mais reste un réseau cellulaire qui peut se déployer rapidement, sans nécessité d'une lourde infrastructure centralisée ni d'une configuration statique préalable. Cette caractéristique les rend particulièrement adaptés aux environnements dynamiques où les besoins de connectivité peuvent surgir soudainement et de manière imprévisible.

Le concept de réseaux auto-déployables est d'autant plus opportun et stimulant que les réseaux privés sont en pleine émergence [12] et que les drones prennent une bonne place dans la future normalisation 6G [13]. En effet, d'un côté, les réseaux cellulaires privés sont des réseaux cellulaires conçus pour répondre aux besoins spécifiques d'une organisation ou d'une entreprise. Contrairement aux réseaux cellulaires publics, qui sont gérés par des opérateurs de télécommunications et accessibles au grand public, les réseaux cellulaires privés sont déployés et gérés par l'organisation elle-même [12]. De l'autre côté, même si le standard 6G n'est pas encore officiellement défini, les diverses recherches et discussions en cours montrent que les drones (à côté de l'intelligence artificielle) prendront une place très importante dans ce standard à la fois pour améliorer la couverture réseau à différents niveaux mais aussi pour proposer des solutions à des applications industrielles spécifiques ou à des applications d'urgence et de secours [13].

Même si le concept de réseau cellulaire auto-déployable peut être applicable et utile dans différentes situations telles que l'extension de la couverture et la gestion des événements à forte affluence, l'une des principales applications reste la réhabilitation de la communication dans les situations d'urgence [9, 14, 15]. En effet, dans ces situations qui suivent généralement une catastrophe naturelle ou humaine, les réseaux traditionnels sont partiellement ou totalement inopérants. Les réseaux auto-déployables peuvent offrir ici une solution clé pour rétablir facilement et rapidement les communications entre les individus et les premiers intervenants [9].

Dans les réseaux auto-déployables, les stations de base sont déployées rapidement et facilement. Elles peuvent être connectées à un réseau traditionnel pour étendre la couverture, par exemple, ou être déployées de manière autonome pour être utilisées séparément [9, 10]. Dans nos travaux, nous nous concentrons sur ce dernier type de déploiement où il n'y a pas de liaison (backhaul) vers un réseau traditionnel et où toutes les fonctionnalités du réseau cœur peuvent être regroupées et co-localisées avec une ou plusieurs stations de base du réseau.

Un point essentiel apporté par les réseaux auto-déployables est le continuum entre les réseaux d'accès radio (RAN) et le réseau cœur (CN). En effet, les fonctions virtualisées du réseau cœur peuvent être directement hébergées dans une station de base formant le réseau cœur local [16]. Ces fonctions peuvent être hébergées dans chaque

station de base, comme considéré dans [4, 5, 6] ou dans une station de base qui héberge un réseau cœur (CN) local partagé comme dans le cas de [9, 10, 16]. Dans notre thèse, nous considérons principalement ce deuxième cas, où toutes les fonctionnalités du réseau cœur sont regroupées dans une entité appelée, réseau cœur local qui sera co-localisée avec une station de base du réseau. Nous discuterons vers la fin la possibilité de considérer plusieurs réseaux cœur dans un réseau auto-déployable et la possibilité d'adapter facilement nos modèles à ce cas.

1.2 Objectifs de la thèse, verrous traités et aperçu des contributions

Nous discutons dans cette section les principaux verrous scientifiques traités dans la thèse avant de dresser un résumé des contributions.

1.2.1 Verrous traités et objectifs de la thèse

Malgré les avantages des réseaux auto-déployables et du continuum qu'ils offrent entre le réseau d'accès radio et le réseau cœur, ils amènent de nouveaux défis scientifiques. Parmi ces défis, le problème d'allocation de ressources est un verrou crucial car les entités mobiles embarquant des UE, des stations de base, une partie ou la totalité des réseaux cœur sont mobiles avec un trafic et des liens fortement dynamiques ce qui rend les solutions traditionnelles d'allocation de ressources inadaptées [50]. Parmi ces défis, le problème de placement des stations ou des fonctions est aussi d'importance majeure et a attiré récemment beaucoup d'attention. Ce problème a été étudié sous plusieurs angles dans la littérature incluant le déploiement optimal de stations de base embarquées sur des drones [50, 51, 52], la planification de trajectoires des stations de base aériennes [71], le placement des fonctions réseau et le placement de réseaux cœur [10, 72]. Ces travaux parmi d'autres feront l'objet de discussions plus détaillées dans le chapitre suivant.

Dans le cadre de notre thèse, nous nous sommes focalisés principalement sur le problème de placement optimal du réseau cœur (CN) local partagé [9, 10, 72]. En effet, le choix de la station de base qui hébergera le réseau cœur local doit être fait de manière rigoureuse, tout en considérant différents objectifs tels que le flux global de trafic, l'équité entre les stations de base, etc. Aussi un verrou d'importance cruciale est le temps d'exécution des modèles de placement optimal. En effet, le choix de la station de base qui hébergera le réseau cœur local doit être calculé dans un temps raisonnable afin de répondre au besoin de réactivité qui est souvent très important. Aussi,

1.2 Objectifs de la thèse, verrous traités et aperçu des contributions

les réseaux auto-déployables sont souvent très dynamiques ce qui peut entraîner un changement important dans la topologie physique et/ou dans les flux de trafic que chaque station de base souhaite envoyer. Dans ce cas, les modèles de placement optimal doivent être relancés avec une éventuelle migration du réseau cœur. Tout cela rend le temps d'exécution des modèles une question très importante.

Un autre verrou important qui nous a intéressé dans cette thèse concerne l'équité entre les stations de base dans un réseau cellulaire auto-déployable. En effet, plusieurs résultats de nos tests montrent que le placement optimal du réseau cœur en maximisant le flux global de trafic envoyé par les stations de base, cause un déséquilibre important en termes de satisfaction absolue et relative par rapport à la demande de trafic des différentes stations de base. Pour cela, nous nous sommes intéressés à ce verrou en essayant de mieux comprendre le compromis entre la maximisation du flux de trafic et l'équité entre les stations de base.

L'objectif principal de notre thèse était donc de proposer de nouveaux modèles qui permettent de choisir la meilleure station de base pour héberger le réseau cœur local. Nous avons pour but de proposer des solutions optimales et efficaces qui s'exécutent dans un temps plus faible que les autres solutions itératives de la littérature. Dans le même sens, notre objectif était de proposer des solutions qui passe à l'échelle en offrant des temps d'exécution raisonnables même dans le cas de grand réseaux avec des centaines de stations de base. Un sous-objectif était également de proposer des solutions génériques pouvant être adaptées à diverses situations ou contraintes, telles que l'hétérogénéité des demandes de trafic et des flux initiaux provenant des stations de base, des métriques d'optimisation variées, ou encore des contraintes supplémentaires sur les stations de base.

Le deuxième objectif de cette thèse était de tenter de garantir une équité en termes de satisfaction des demandes de trafic entre les stations de base. Pour cela, notre premier sous objectif était de proposer des modèles qui visent à garantir l'équité absolue ou relative entre les stations de base et d'étudier leurs impacts sur le flux global de trafic. Un autre sous objectif était d'analyser et de comprendre la relation entre le meilleur flux qu'on peut atteindre et l'équité ce qui aidera dans l'étude du compromis entre la maximisation du flux de trafic et la gestion de l'équité dans les réseaux cellulaires auto-déployables.

Le troisième objectif de la thèse, fortement lié aux deux premiers, était l'étude de la scalabilité (passage à l'échelle) de nos modèles en fonction de différents paramètres. Ces paramètres incluent le nombre de stations de base dans le réseau, le degré de connectivité entre elles, le nombre de réseaux cœur à déployer, ainsi que les métriques d'optimisation considérées.

1.2.2 Résumé des contributions

Afin d'atteindre nos objectifs, nous avons dans un premier temps proposé un programme non linéaire mixte en nombres entiers qui permet de trouver en une itération (one-shot) la meilleure station de base hébergeant le réseau cœur local tout en optimisant une métrique de performance (flux global de trafic envoyé par les stations de base, des métriques d'équité entre les stations de base, etc.). Nous avons ensuite linéarisé les contraintes non linéaires afin d'atteindre des temps d'exécution très faibles et discuté le modèle linéaire résultant. Notre modèle est générique et permet d'ajuster facilement les métriques à optimiser et/ou d'ajouter ou de modifier les contraintes selon le besoin. En considérant la maximisation du flux global de trafic, nous avons évalué les performances de notre modèle en le comparant à des solutions similaires de la littérature. Les résultats montrent que notre approche conjointe en une itération réduit considérablement le temps d'exécution par rapport aux solutions existantes tout en fournissant les mêmes résultats optimaux.

Grâce à notre formulation générique, nous avons considéré dans un deuxième temps le problème d'équité entre les stations de base. En effet, les résultats ont montré que la maximisation du flux global de trafic cause un déséquilibre important en termes de satisfaction absolue et relative par rapport à la demande de trafic des différentes stations de base. Nous avons donc proposé des variantes efficaces de notre modèle qui permettent d'assurer le placement optimal du réseau cœur tout en garantissant une équité entre les stations de base. Nous avons mené différentes simulations pour évaluer les résultats obtenus et les performances de nos variantes visant à garantir l'équité de satisfaction entre les stations de base. Les résultats de simulations nous ont permis de discuter le compromis entre la maximisation du flux de trafic et la gestion de l'équité entre les stations de base, tout en mettant en avant les avantages de notre approche de modélisation conjointe.

Dans un troisième temps, nous avons enfin étudié plus en détail le passage à l'échelle de nos modèles en fonction du nombre de stations de base, du nombre de réseaux cœur à déployer et selon les métriques d'optimisation à considérer. Les résultats montrent que nos modèles offrent des temps d'exécution très raisonnables même dans le cas de grands réseaux avec des centaines de stations de base.

1.3 Structure détaillée du document

Après avoir présenté et discuté le contexte global et le positionnement de nos travaux dans ce chapitre introductif, le chapitre 2 sera consacré à l'état de l'art. Nous reviendrons dans un premier temps sur les réseaux auto-déployables et les principaux

travaux de la littérature qui ont adressé la conception et le déploiement des réseaux cellulaires auto-déployables. Ensuite, nous nous focalisons sur les papiers qui traitent le problème de placement optimal du réseau cœur qui est le principal problème auquel nous nous attaquons dans cette thèse. Nous présentons également des travaux concernant le placement optimal dans les réseaux conventionnels et le placement optimal dans les environnements NFV. Comme dernier point, nous discuterons quelques travaux de la littérature concernant le placement optimal dans le MEC (Mobile Edge Computing).

Le chapitre 3, présente les détails des notations, des hypothèses, de la formalisation et, principalement, notre solution de placement optimal du réseau cœur. En effet, nous présentons en détail le modèle que nous proposons en commençant par la formulation intuitive et non linéaire du problème et en discutant en détail la linéarisation des contraintes non linéaires et le modèle résultant. Nous finissons ce chapitre par discuter les avantages de notre modèle et quelques extensions possibles.

Le chapitre 4, est consacré à la validation du modèle que nous avons proposé et à l'évaluation de ses performances. Nous y présentons dans un premier temps la configuration et les paramètres de simulation. Nous présentons ensuite et discutons les différents résultats d'évaluation et de comparaison. En particulier, nous évaluons et validons les performances de notre modèle en comparant les solutions optimales obtenues ainsi que le temps d'exécution à ceux de deux autres modèles de la littérature. Nous menons ces comparaisons et évaluations en fonction de la taille du réseau et du degré de connectivité, tout en considérant deux cas : un cas homogène des flux initiaux provenant des stations de base, et un cas hétérogène plus général.

Dans le chapitre 5, nous présentons de nouveaux modèles permettant d'assurer l'équité dans le réseau. Nous commençons par introduire cette direction à travers un exemple concret et une discussion. Ensuite, nous définissons les métriques prises en compte et détaillons les nouveaux modèles proposés. Enfin, nous présentons les résultats de simulation de ces modèles et évaluons leurs performances en termes d'équité, de flux global atteint, et de temps d'exécution. Une discussion sur le compromis entre l'équité et la maximisation du flux de trafic est proposée à cette occasion.

Dans le chapitre 6, nous présentons une étude de la scalabilité de nos modèles, à la fois en fonction de nombre de stations de base dans les réseaux, en fonction du nombre de réseaux cœur à déployer et selon les métriques d'optimisation.

Nous concluons ce document par un bilan global des contributions et des conclusions, et nous discutons quelques extensions et travaux futurs que nous souhaitons explorer.

Publication issue de la thèse :

Titre : One-shot models for local core network placement and traffic optimization in self deployable networks.

Revue : Journal of Network and Computer Applications (Impact Factor 8.7).

Lien : <https://doi.org/10.1016/j.jnca.2023.103721>

Référence complète : Asma Bechkit and Ali Melit, "One-shot models for local core network placement and traffic optimization in self deployable networks," Journal of Network and Computer Applications, vol. 219, p. 103721, 2023.

Chapitre 2

État de l'art sur les réseaux auto-déployables et le placement optimal

Contenu du chapitre

2.1	Conception et expérimentation des réseaux cellulaires auto-déployables	34
2.2	Placement optimal du réseau cœur dans les réseaux cellulaires auto-déployables	36
2.3	Placement optimal des stations de base aériennes	37
2.4	Placement optimal dans les réseaux conventionnels	38
2.5	Placement optimal dans les environnements NFV	41
2.6	Placement optimal dans le Mobile Edge Computing	43
2.7	Conclusions	46

Après avoir discuté du contexte global et introduit le concept des réseaux cellulaires auto-déployables dans le chapitre précédent, nous commençons ce chapitre par présenter quelques principaux travaux représentatifs qui ont adressé la conception et/ou l'expérimentation de ces réseaux cellulaires auto-déployables. Dans une deuxième section, nous présentons plus en détail l'un des défis majeurs des réseaux cellulaires auto-déployables qui est le placement optimal du réseau cœur (CN), et qui constitue le principal problème traité dans cette thèse. Dans les sections qui suivent nous présentons un état de l'art sur des travaux qui se rapprochent de notre problème sans qu'ils soient au cœur de ce dernier. En particulier, nous discuterons quelques travaux représentatifs sur le déploiement de stations de base aériennes et sur le placement optimal dans les réseaux conventionnels, dans les réseaux implémentant le NFV (Network Function Virtualization) et dans le MEC (Mobile Edge Computing).

2.1 Conception et expérimentation des réseaux cellulaires auto-déployables

Comme discuté dans le chapitre précédent, les réseaux cellulaires auto-déployables offrent d’un côté une grande flexibilité avec différents champs d’applications ce qui les rend une technologie d’avenir. De l’autre côté, ils amènent de nouveaux défis techniques et scientifiques qui sont encore au stade de la recherche et qui sont encore loin d’être maîtrisés. Ces deux aspects ont fait que cette technologie a attiré beaucoup d’attention de la part des chercheurs ces dernières années. Dans ce cadre, de nombreux travaux de recherche commencent récemment à s’attaquer aux différents verrous technologiques, en proposant des plateformes expérimentales et/ou des solutions théoriques aux problèmes associés.

Dans ce domaine en constante évolution des réseaux cellulaires auto-déployables, les chercheurs explorent diverses voies pour concevoir et améliorer les performances de ces réseaux. Certains se concentrent sur le développement et le déploiement réel de réseaux cellulaires auto-déployables [3, 4, 5, 6]. Parallèlement, l’attention se porte également sur le défi du placement optimal des réseaux cœur afin d’optimiser les performances en termes de capacité [10], de réduction des délais de transmission [16], etc. D’autres chercheurs s’attaquent au problème de placement physique des stations de base aériennes [50, 51, 52].

L’un des principaux axes de recherche lié à notre thèse concerne la conception et le déploiement dans le monde réel d’architectures cellulaires auto-déployables. Dans [3], les auteurs conçoivent, développent et mettent en place l’un des premiers prototypes de réseaux cellulaires auto-déployables. Ils motivent leur travail par les possibilités offertes par la technologie LTE (4G) combinée à la démocratisation des drones à basse altitude pour un déploiement à la demande. Ils discutent également les défis posés par la conception du réseau d’accès radio (RAN), de la liaison terrestre et du réseau cœur (CN) pour les réseaux cellulaires à base de drones. Dans ce travail, les auteurs présentent une conception de bout en bout, nommée SkyLiTE, en tant qu’un réseau auto-organisé de drones à basse altitude qui fournit une connectivité LTE optimisée [3]. La solution proposée comprend trois éléments : SkyRAN, SkyHaul et SkyCore, qui traitent les défis mutuellement dépendants dans les couches RAN (réseau d’accès), backhaul et CN (réseau cœur) respectivement. Dans ce travail, les auteurs mettent en oeuvre une première version comprenant les modules SkyRAN et SkyCore et l’ont déployée avec succès sur un drone hexacoptère DJI [3].

2.1 Conception et expérimentation des réseaux cellulaires auto-déployables

Par la suite, cette architecture a été étendue et étudiée de manière plus approfondie pour supporter plusieurs drones dans d'autres travaux [4, 5, 6]. Dans [4], les auteurs proposent une architecture Edge fiable et évolutive, nommée SkyCore. Ainsi, une conception adaptée du réseau cœur EPC (Evolved Packet Core, désignant le cœur de réseau dans la nomenclature LTE) est présentée, où toutes ses fonctionnalités sont virtualisées et regroupées dans une entité autonome, légère et indépendante qui est co-localisée avec chaque drone hébergeant une station de base (BS). SkyCore introduit deux nouveaux composants : la refonte logicielle de la fonctionnalité EPC et une communication inter-EPC (drones) efficace, via un réseau défini par logiciel (SDN). En outre, les auteurs considèrent certaines questions spécifiques telles que la mobilité des drones et leur durée de vie limitée, ainsi que leur impact sur les opérations du réseau. Par exemple, lorsqu'un drone doit quitter le réseau pour être rechargé, un schéma de migration de l'état du drone est proposé. SkyCore prend en charge l'interopérabilité avec les stations de base LTE commerciales et les smartphones. Les auteurs de ces travaux ont développé et déployé une version complète du prototype SkyCore sur des drones à voilure tournante DJI Matrice 600 Pro pour créer un réseau LTE à deux drones [4, 5, 6]. Ils ont également comparé SkyCore à une architecture EPC conventionnelle (un EPC standard dans chaque drone LTE). Les résultats montrent la faisabilité de leur architecture et mettent en évidence l'évolutivité et les bonnes performances de SkyCore, principalement en termes de délais du plan de contrôle du réseau [4, 5, 6].

Outre les travaux impliquant à la fois la conception et le développement de solutions de réseaux cellulaires auto-déployables dans des environnements réels, d'autres recherches ont adressé des questions relatives aux réseaux auto-déployables dans le cadre des scénarios simulés. Dans [73], les auteurs abordent l'articulation entre la norme LTE du 3GPP et le réseau auto-déployable, en se concentrant principalement sur l'association des utilisateurs. Ils étudient les normes 3GPP qui ont pour objectif de renforcer les opérations de sécurité publique en utilisant des réseaux LTE (Long Term Evolution) et montrent que la procédure d'association proposée ne prend pas en compte les exigences en matière de qualité de service (QoS). Ensuite, les auteurs proposent des schémas d'association optimisés et plus adaptés à la gestion des situations d'urgence.

Dans [71], les auteurs présentent une solution basée sur le SDN (Software Defined Networking) pour déployer efficacement des stations de base aériennes (ABSs, Aerial Base Stations) contrôlables dans les réseaux cellulaires. Ils proposent également une approche de planification de trajectoire qui réduit la consommation d'énergie tout en tenant compte de la couverture des utilisateurs et de la qualité de la communication. Les auteurs ont réalisé des simulations qui montrent les bonnes performances de leur solution en termes de connectivité et de consommation d'énergie.

2.2 Placement optimal du réseau cœur dans les réseaux cellulaires auto-déployables

L'un des principaux problèmes que nous abordons dans cette thèse concerne le placement optimal du réseau cœur local dans un réseau cellulaire auto-déployable. Cette question a été spécifiquement étudiée dans certains travaux récents. Dans [9, 10], Oueis et al. traitent le problème du placement optimal d'un EPC local (EPC, Evolved Packet Core, désignant le cœur de réseau dans la nomenclature LTE) dans un réseau cellulaire auto-déployable. Ils considèrent un ensemble de stations de base (BS) sans liaison de retour (backhaul) avec un réseau traditionnel et proposent une nouvelle métrique de centralité, appelée centralité de flux. Cette métrique permet de choisir la meilleure station de base pour placer le réseau cœur local, afin de maximiser la quantité de flux de trafic échangé entre les stations de base et le réseau cœur local, tout en respectant la bande passante limitée sur les différents liens.

Les auteurs modélisent le réseau sous la forme d'un graphe géométrique aléatoire pondéré avec des contraintes de capacité pour modéliser la bande passante limitée [10]. Ils proposent ensuite un modèle de flux maximal itératif basé sur la formulation du problème en tant que problème d'optimisation linéaire (LP, Linear Programming) ainsi qu'une approche itérative équivalente basée sur le problème de la coupe minimale (Min-Cut) dans un graphe. Ils supposent que toutes les stations de base envoient le même flux de trafic (cas homogène). Pour calculer le score de centralité de flux d'un nœud du graphe (station de base), ils proposent de calculer le flux maximal de trafic que le nœud peut recevoir de tous les autres nœuds sources du graphe s'il était choisi comme nœud destinataire du flux (puits, hébergeant le réseau cœur). Ils proposent ensuite de choisir le nœud ayant le meilleur score de centralité pour héberger le réseau cœur. Ainsi, les deux solutions proposées sont basées sur la recherche du flux maximal (ou de la valeur de coupe minimale équivalente) pour chaque puits potentiel, puis à choisir le puits potentiel offrant le meilleur flux maximal. Les auteurs réalisent des simulations approfondies dans différentes topologies de graphes tout en comparant leurs solutions à d'autres métriques de centralité et en considérant l'objectif de maximisation du flux global. Les modèles présentés dans [9, 10] sont les principaux modèles auxquels nous allons nous comparer par la suite. Nous reviendrons donc brièvement sur ces travaux en présentant formellement les deux modèles itératifs à base de LP et à base de coupe minimale (Min-Cut) dans le chapitre suivant.

Dans [72], les auteurs étendent leur travail présenté dans [10] en isolant les deux entités de cœur de réseaux LTE : l'entité de gestion de la mobilité (MME, Mobility Management Entity) et la passerelle de service (S-GW, Serving Gateway). Ils partent du

principe que le MME est situé avec une station de base (BS) et que le placement de la passerelle de service (S-GW) peut se faire selon trois scénarios (centralisé, distribué sur toutes les BSs et distribué de manière optimale sur un sous ensemble de BSs). Le premier scénario (appelé une S-GW par défaut) considère qu'une seule S-GW doit être placée dans une station de base de telle sorte à minimiser la consommation de bande passante de la liaison de retour (backhaul). Dans le deuxième scénario, chaque station de base héberge une S-GW et les auteurs supposent que chaque utilisateur (UE) s'attache à la S-GW hébergée dans la station de base à laquelle il est associé. L'objectif est donc uniquement d'optimiser le placement du MME. Dans le troisième scénario, seulement un sous-ensemble de stations de base peut héberger des S-GW et les auteurs proposent un modèle qui permet de déterminer le nombre de S-GW et d'utilisateurs qui leur sont attachés, leur emplacement et l'emplacement du MME tout en minimisant la consommation de la bande passante de la liaison de retour (backhaul). Sous les hypothèses considérées, les auteurs ont montré que les deux derniers scénarios réduisent considérablement la consommation de la liaison de retour par rapport au scénario d'une S-GW par défaut. Ils ont aussi montré que ces deux derniers scénarios offrent des quantités de trafic similaires sur la liaison de retour. En effet, les résultats présentés suggèrent que les S-GW peuvent être réparties sur toutes les BSs avec une perte négligeable par rapport à un placement optimal sur un sous ensemble de BSs.

Dans [16], les auteurs proposent une heuristique pour sélectionner la station de base qui devrait héberger le réseau cœur local tout en minimisant le délai entre les autres stations de base et ce réseau cœur local dans un réseau cellulaire auto-déployable. Ils montrent que leur solution offre de meilleurs résultats en termes de délai par rapport aux solutions visant à optimiser la capacité. En effet, ils montrent que les solutions tenant compte de la capacité peuvent conduire à de multiples saturations de liens, ce qui augmente considérablement les délais.

2.3 Placement optimal des stations de base aériennes

Au delà du placement optimal des réseaux cœur, qui est au centre de notre thèse, ainsi que des solutions similaires et comparables à la nôtre, nous avons jugé que d'autres sujets sont d'intérêt et permettront de comprendre des solutions liées aux verrous des réseaux cellulaires auto-déployables ou au problème du placement optimal. Parmi ces travaux connexes, nous avons identifié le placement des stations de base dans un contexte de réseaux auto-déployables, le problème de placement dans les réseaux conventionnels, le placement des fonctions virtualisées et le placement optimal dans le Mobile Edge Computing. Nous commençons par aborder dans cette section des travaux représentatifs traitant le problème de placement des stations de base aériennes.

Ce problème a attiré récemment beaucoup d'intérêt [50, 51, 52]. Shahzadi et al. présentent dans [50] un état de l'art général sur l'intégration de drones dans les réseaux de communication sans fil tout en adressant cette question de placement optimal de stations de base aériennes. Ils traitent entre autres les différents cas d'utilisation de drones dans les réseaux 5G-B5G (5G and beyond 5G), tel que l'utilisation en tant que relais volant, station de base ou équipement utilisateur (UE). Ils discutent aussi l'impact de drones sur l'amélioration de la connectivité et de la portée des communications dans les zones difficiles d'accès. Ils passent en revue quelques solutions de placement optimale 3D et d'allocation de ressources.

Dans un scénario plus précis, les auteurs de [51], proposent un mécanisme de placement intelligent des drones embarquant des stations de base 5G, pour améliorer les opérations du réseau mobile, dans les situations de foule dense ou dans les situations d'urgence. L'objectif est d'améliorer la couverture des utilisateurs mobiles et les performances du réseau de manière générale. Les auteurs proposent d'utiliser un algorithme génétique pour résoudre le problème d'optimisation qu'ils considèrent pour le placement des stations de base aériennes. Les résultats de simulation montrent que la solution proposée permet de couvrir 90 % des utilisateurs mobiles tout en offrant un taux de livraison de paquets de près de 90 % et tout en présentant un taux rapide de convergence.

Huang et al. [52] ont aussi étudié le problème du déploiement 3D optimal de stations de base embarquées sur des drones. Dans [52], les auteurs présentent un problème d'optimisation avec l'objectif de maximiser le nombre d'utilisateurs couverts et d'offrir une qualité de service (QoS) acceptable tout en s'assurant que les drones ne sont pas loin d'une station pour recharger leurs batteries. Ils proposent également un algorithme permettant de résoudre ce problème et de trouver les positions optimales pour placer les drones embarquant des stations de base. Les auteurs réalisent des simulations approfondies qui montrent l'efficacité de l'algorithme proposé.

2.4 Placement optimal dans les réseaux conventionnels

Outre les travaux sur les réseaux cellulaires auto-déployables, le problème du placement des passerelles et des nœuds-puits a été abordé dans la littérature pour différents types de réseaux, notamment les réseaux cellulaires classiques [74], les réseaux ad-hoc sans fil [75, 76] et les réseaux de capteurs sans fil [77, 78]. Même si ces travaux ne sont pas exactement au coeur de notre cas d'usage des réseaux cellulaires auto-déployables, nous avons étudié et résumé quelques travaux représentatifs liés aux problèmes que nous traitons.

Par exemple dans [74], les auteurs abordent la question du déploiement centralisé/distribué des passerelles dans un réseau 5G à l'échelle d'un pays. Ils définissent tout d'abord une métrique du coût du transport des données directement vers une passerelle donnée ou indirectement via plusieurs passerelles distribuées. Sur la base de cette métrique et d'un modèle de trafic donné, ils simulent un réseau à l'échelle d'un pays (Allemagne) et considèrent deux cas : i) une seule passerelle centrale et ii) plusieurs (n) passerelles déployées dans les n plus grandes villes du pays. Les auteurs effectuent des simulations en faisant varier le trafic dans les villes et le nombre de passerelles. Les résultats montrent qu'un petit nombre de passerelles bien réparties donne de bons résultats. A titre d'exemple, selon le modèle de coût, l'utilisation de quatre passerelles à l'échelle du pays, peut entraîner des coûts de transmission qui ne dépassent que de 40% le coût optimal théorique qui nécessiterait des milliers de passerelles distribuées [74].

Dans un cas d'utilisation différent, certains travaux de recherche se sont penchés sur le placement optimal des nœuds-puits (appelées aussi passerelles dans certains travaux) dans les réseaux maillés sans fil (WMN, Wireless Mesh Networks) afin de maximiser la quantité de flux. Les auteurs de [75] proposent trois heuristiques qui permettent de choisir une position offrant une solution proche de l'optimale pour placer une passerelle, tout en maximisant le débit du réseau. Afin de calculer la solution optimale exacte et donc d'évaluer la qualité des trois heuristiques, les auteurs utilisent une recherche exhaustive tout en se référant à un travail antérieur [79] qui développe une solution pour déterminer le débit optimal et la configuration (itinéraires de flux, ordonnancement des liens marche/arrêt et paramètres de la couche physique) compte tenu d'une passerelle connue. Dans [75], les auteurs ont utilisé ce modèle et ils ont appliqué une approche de recherche itérative exhaustive pour trouver la position optimale de la passerelle (offrant le meilleur débit) et la position de la passerelle dans le pire des cas (offrant le plus faible débit). Ils évaluent les performances de leurs trois heuristiques par rapport à ces deux cas en termes de débit du réseau, tout en considérant différents types de réseaux (réseaux en grille, réseaux en grilles régulières sous-compactes, réseaux arbitraires et réseaux en grilles irrégulières). Les résultats montrent principalement que la différence entre le meilleur et le pire placement de passerelle atteint 50% en termes de débit. Ils montrent également que deux heuristiques parmi les trois proposées offrent de très bonnes performances quasi-optimales.

D'autres travaux se concentrent sur le choix d'un nombre minimum de passerelles dans les réseaux maillés sans fil, comme dans [76]. Dans ce travail, les auteurs proposent un algorithme calculant de manière récursive des ensembles dominants pondérés (DSs, Dominating Sets) minimaux pour choisir les positions des passerelles dans un réseau maillé sans fil tout en garantissant une qualité de service (QoS) donnée dans

le réseau. Ils comparent les performances de l'algorithme de base et de l'algorithme récursif pondéré à deux autres solutions de la littérature : l'ensemble dominant itératif glouton (Iterative Greedy Dominating Set) et le placement progressif (Augmenting Placement). Les auteurs évaluent les résultats concernant le nombre de passerelles nécessaires en utilisant différents objectifs de qualité de service. Les résultats montrent que l'algorithme récursif proposé dans ce travail offre des meilleures performances par rapport aux autres solutions. En effet, il permet de placer jusqu'à 50% de passerelles en moins tout en offrant des performances correctes sous différentes contraintes de qualité de service.

Le problème du placement des nœuds-puits présente également un grand intérêt dans les réseaux de capteurs sans fil (WSNs, Wireless Sensor Networks), car il a un impact significatif sur le délai de transmission, l'agrégation des données, la consommation d'énergie, la durée de vie du réseau, etc. Alors que certains travaux se concentrent sur le placement géographique des nœuds-puits pour prolonger la durée de vie du réseau [77], d'autres s'attaquent au placement optimal de plusieurs nœuds-puits dans les réseaux de capteurs sans fil avec agrégation de données [78]. En effet, les auteurs de [78] considèrent un ensemble de nœuds de capteurs déployés et abordent le problème du déploiement des nœuds-puits supplémentaires pour une agrégation de données efficace en termes de latence et d'énergie. Ils proposent une phase de pré-traitement comportant deux sous-tâches : i) l'identification des régions compétitives où le déploiement de nœuds-puits est utile/possible et ii) le calcul du nombre de sauts de chaque nœud vers chaque région compétitive (en utilisant une recherche en largeur d'abord à partir de chaque région compétitive). En utilisant des paramètres de réseau connus et les résultats de la phase de pré-traitement, ils développent deux approches d'optimisation, dans le but de minimiser la latence (délai de communication) ou de minimiser le coût d'énergie de la communication (et donc d'augmenter la durée de vie du réseau). Dans le cas des petits réseaux, ils introduisent une formulation basée sur la programmation linéaire en nombre entiers (ILP) pour le placement optimal des nœuds-puits. Par ailleurs, ils proposent une approche heuristique pour les grands réseaux, qui est basée sur un nouveau paradigme de sélection appelé sélection négative. L'idée principale de cette approche est d'exclure à chaque étape la plus mauvaise position. Les résultats de la simulation montrent que l'ajout de certaines passerelles réduit considérablement la latence et la consommation d'énergie.

2.5 Placement optimal dans les environnements NFV

Comme discuté dans le chapitre précédent, la virtualisation des fonctions réseaux (NFV, Network Function Virtualization) qui vise à découpler les fonctions réseau du matériel dédié en les implémentant sous forme de logiciel a considérablement accéléré et facilité l'émergence d'architectures auto-déployables dans les réseaux cellulaires. Cependant cette technologie amène de nouveaux et multiple défis. Le problème de placement des fonctions réseaux est l'un des principaux défis dans le domaine spécifique des réseaux cellulaires mais aussi dans les réseaux NFV de manière beaucoup plus large [80]. En effet, ce placement impacte considérablement les différentes performances du réseau, tel que la fiabilité, les délais, le coût, etc.

Les auteurs dans [81], présentent un état de l'art riche concernant le placement des différentes VNF (Virtualized Network Functions) et CNF (Containers Network Functions) dans les réseaux 5G. Ils proposent une classification des méthodes et algorithmes existants pour le placement des VNF qui considèrent une seule métrique d'optimisation ou plusieurs métriques combinées. Les auteurs classent les solutions de planification en trois parties : heuristiques, méta-heuristiques et algorithmes d'apprentissage automatique. Ils identifient les métriques de performance et discutent les avantages et les limitations pour chaque catégorie.

Parmi les travaux [82, 83, 84, 85, 86, 87] traitant le placement des fonctions réseau virtualisées dans les réseaux cellulaires, Leyva-Pupo et al. considèrent dans [87] le problème de placement des fonctions du plan utilisateur (UPF, User Plane Function) dans les réseaux de cinquième génération (5G). Ils modélisent le problème sous forme d'un modèle de programmation linéaire en nombres entiers (ILP) multi-objectifs qui vise à optimiser les coûts d'approvisionnement et la qualité de service. Le modèle englobe plusieurs aspects tels que : les exigences spécifiques à l'UPF, les demandes de service, les capacités physiques du réseau, les contraintes de commande des VNF et la topologie de la chaîne (branches simples et multiples).

Au delà des réseaux cellulaires, les problèmes de placement des fonctions réseau virtualisées a été largement étudié dans les réseaux de manière générale [88, 89, 90, 91, 92, 93, 94, 95, 96]. Dans [97], les auteurs présentent un état de l'art approfondi des stratégies et des algorithmes existants concernant le placement des fonctions réseau virtualisées dans les réseaux. Ils organisent ces solutions en différentes catégories, allant des cadres génériques de virtualisation des fonctions réseau basées sur des machines virtuelles aux stratégies de placement des fonctions réseau virtualisées spécifiques. De son côté, Demirci et al. présentent dans [98] un autre état de l'art traitant le placement optimal des fonctions réseaux virtualisées dans les réseaux SDN (Software Defined Networks).

Les auteurs de [99] proposent une solution utilisant l'apprentissage automatique pour le placement des fonctions réseau virtualisées. Les auteurs proposent d'abord un programme linéaire en nombre entiers (ILP) qui prend en compte la latence, le coût de violation de l'objectif de niveau de service, l'utilisation du matériel et le coût de réajustement de la fonction. Ensuite, le réseau de substrat est partitionné en un ensemble de clusters disjoint à la demande. Enfin, un algorithme basé sur les clusters et piloté par les données est proposé pour le placement et le réajustement des fonctions virtualisées. Les résultats des simulations montrent que la solution proposée réduit considérablement le temps de réajustement et diminue l'utilisation du matériel par rapport aux solutions de la littérature (K-means, K-medoids et la migration sans clustering). Dans [100], les auteurs proposent un modèle d'apprentissage hybride heuristique-méta-heuristique pour le problème de placement des fonctions de réseau virtuel en ligne. Ce modèle est constitué de deux parties : une offline (hors ligne) exécutée une seule fois avant l'opération NFV principale et une online (en ligne). Pour la composante hors ligne, une méta-heuristique basée sur un algorithme génétique est proposée pour apprendre les hyperparamètres du modèle en ligne et minimiser la consommation totale d'énergie dans l'infrastructure NFV. Pour la composante en ligne, une heuristique multi-critère appelée MCH est proposée. Les auteurs montrent grâce à différentes simulations, l'efficacité de la solution proposée par rapport à la littérature. Dans [101], les auteurs adressent un autre sujet complémentaire et proposent un mécanisme de déploiement et de contrôle automatique des services réseau (NS, Network services) dans les environnement NFV. Ils proposent une méthode pour décrire et déployer les composants NSs efficacement dans le réseau NFV et pour contrôler leur comportement. Un prototype a été mis en œuvre pour valider la fonctionnalité du système et pour l'évaluation de ses performances.

Le chaînage dans les réseaux NFV fait référence à la manière dont différentes fonctions réseau virtuelles sont organisées et inter-connectées. Il constitue aussi un enjeu important dans ces réseaux. Les auteurs dans [102] formalisent le problème de placement et de chaînage des fonctions de réseau virtuel pour les applications en temps réel sous la forme d'un programme linéaire en nombres entiers (ILP). Ils proposent ensuite une solution heuristique afin de résoudre le problème tout en offrant une scalabilité. Selon les résultats, l'heuristique proposée réduit le temps d'exécution jusqu'à 80% par rapport au modèle optimal à base d'ILP. Les auteurs de [103] étudient aussi le problème de placement et de chaînage efficace des fonctions réseau virtuelles. Les auteurs expliquent que les solutions actuelles à ce problème sont soit longues à mettre en œuvre et incapables de gérer des requêtes en temps réel, soit utilisent des méthodes heuristiques pour trouver des solutions approximatives. Pour résoudre ce problème, ils proposent une méthode de résolution exacte et efficace basée sur la décomposition

du problème. Ils améliorent cette approche de solution exacte en utilisant une heuristique de branchement pour explorer plus efficacement l'espace de recherche. Les auteurs montrent que leurs approches améliorent les performances par rapport aux solutions existantes.

Dans [104], les auteurs étudient à la fois le routage du trafic et le placement des fonctions réseau virtuelles pour prendre en charge les services de multidiffusion sur un réseau physique supportant la virtualisation des fonctions réseau. Le problème est formulé sous la forme d'un programme linéaire mixte en nombres entiers (MILP) visant à minimiser les coûts de provisionnement de fonctions et de liaisons tout en respectant les contraintes de ressources du réseau physique, les contraintes de conservation de flux et les règles de placement des fonctions réseaux. Étant donné la complexité de résoudre le problème formulé, les auteurs proposent des algorithmes heuristiques à faible complexité pour trouver une solution efficace à la fois pour les scénarios de routage à trajet unique et à trajets multiples. Dans [105], les auteurs adressent aussi un problème d'optimisation conjointe du placement des fonctions virtualisées, de l'allocation des CPUs et du routage des flux dans les scénarios où le changement de trafic est dynamique. Les auteurs formulent le problème comme un programme non linéaire en nombres entiers mixte (MINLP). Ils proposent une méthode basée sur la relaxation pour résoudre le problème sur des petites échelles. Pour les réseaux à grande échelle, ils présentent des heuristiques basées sur un algorithme glouton à base de recuit simulé pour trouver des solutions quasi-optimales dans des temps raisonnables. Les auteurs évaluent les algorithmes sur des topologies et des modèles de trafic réels.

2.6 Placement optimal dans le Mobile Edge Computing

Le Mobile Edge Computing (MEC) est une architecture informatique qui consiste à déplacer une partie des ressources de calcul et de stockage vers des serveurs situés plus près des utilisateurs finaux [106]. Cela évite de traiter les données et d'effectuer les calculs dans des centres de données distants et permet de réduire la latence et d'améliorer l'expérience utilisateur. Il permet également d'étendre les capacités de calcul et de stockage. Le MEC peut être implémenté à différents niveaux dans les réseaux mobiles cellulaires. Il peut être directement implémenté au niveau des infrastructures des opérateurs de réseau mobile [107] (ce qui est répandue et ce qui nous intéresse le plus), des fournisseurs de services Cloud, des fournisseurs de services tiers et même des terminaux mobiles eux-mêmes. Il est à noter que l'implémentation du Mobile Edge Computing, se base souvent sur la virtualisation des fonctions réseau (NFV) que nous avons discuté et sur les conteneurs.

Les problèmes de placement des serveurs et des services Edge sont très importants dans ce type d'infrastructure. En effet, un mauvais emplacement des serveurs Edge pourra entraîner des latences importantes, une surconsommation énergétique, un déséquilibre de charge entre les serveurs [108], etc. Ainsi ce problème a été étudié dans différents travaux [109, 110, 111, 112, 113] de la littérature avec différents objectifs comme la latence (délai d'accès entre l'utilisateur mobile et le serveur Edge), la qualité de service, l'équilibrage de charge, la consommation énergétique, le coût global de déploiement, etc. Même si les problèmes de placement de serveurs et services dans le Mobile Edge Computing semblent loin de notre problème, nous avons décidé d'étudier brièvement quelques travaux représentatifs qui peuvent être utiles, même indirectement. Cela est d'intérêt en particulier dans le cas d'extension de nos travaux pour considérer la latence entre les stations de base et le réseau cœur local, les contraintes de calcul et d'énergie au sein des réseaux cœur à déployer et enfin l'équilibrage de charge entre plusieurs réseaux cœur locaux.

Wang et al. formulent dans [108] le problème de placement des serveurs Edge dans un environnement MEC comme un problème d'optimisation multi-objectifs sous contraintes. Le but est de choisir les meilleurs emplacements des serveurs Edge avec un objectif d'équilibrer la charge entre les serveurs Edge et de minimiser le délai d'accès à ces derniers. Les auteurs proposent un modèle à base de programmation mixte en nombres entiers pour trouver la solution optimale. Ils mènent des évaluations sur un ensemble de données réels provenant du réseau MEC de Shanghai Telecom. Les résultats montrent que les modèles proposés offrent de meilleures performances que les solutions de la littérature en terme de délai d'accès et d'équilibrage de charge dans le MEC. Li et al. [114] formulent aussi ce problème comme un problème d'optimisation multi-objectifs avec contraintes mais en se focalisant principalement sur la consommation d'énergie. Ils proposent une solution à base d'optimisation par essaims de particules et évaluent leurs solutions sur les données réelles de Shanghai Telecom. Les résultats montrent que leur algorithme peut réduire la consommation d'énergie de plus de 10% et améliorer l'utilisation des ressources informatiques de plus de 15% par rapport à d'autres algorithmes. De leur côté, Jiang et al. [115] étudient le problème du placement efficace des serveurs Edge en considérant des environnements dynamiques où les états du réseau varient dans le temps ainsi que les coûts associés au placement. Ils proposent deux algorithmes basés sur l'apprentissage par renforcement profond et montrent que les solutions proposées permettent un placement intelligent et dynamique des serveurs Edge, tout en offrant de meilleures performances par rapport aux solutions existantes.

Poularakis et al. [116] abordent le problème de placement efficace des services et de l'acheminement des demandes d'utilisateurs mobiles vers ces services. Les auteurs considèrent des contraintes multidimensionnelles (stockage, calcul et communication) et proposent une heuristique afin d'atteindre des solutions proches de l'optimale. Les résultats d'évaluation montrent que l'approche proposée permet une utilisation efficace des ressources disponibles pour maximiser le nombre de demandes servies. Ma et al. étudient dans leur travail [117] le placement prédictif des services dans le MEC avec une prédiction des informations à court terme. Ils explorent un mécanisme pour prédire les informations à court terme sur la mobilité de l'utilisateur et proposent une solution de placement prédictif de service Edge sous contraintes de coûts budgétaires à long terme. Les auteurs effectuent une évaluation approfondie en utilisant une trace de données du monde réel. Les résultats montrent que leur modèle offre des bonnes performances en permettant de réduire la latence perçue par les utilisateurs tout en maintenant des coûts faibles et stables.

D'autres recherches adressent conjointement le problème du déploiement des serveurs et des services Edge dans un environnement MEC [118]. Zhang et al. proposent par exemple [118] un modèle conjoint de déploiement des serveurs et de placement des services dans le but de maximiser le profit global de tous les serveurs Edge, sous les contraintes liées au nombre de serveurs, aux relations entre les serveurs Edge et les stations de base, à la capacité de stockage et à la capacité de calcul de chaque serveur Edge. Les auteurs ont mené des évaluations approfondies sur des données du monde réel et les résultats montrent que leur algorithme offre de bons résultats par rapport à d'autres méthodes de la littérature.

Wang et al. [119] considèrent enfin un environnement MEC multi-drones, où un nombre de drones sont déployés en tant que serveurs Edge pour les utilisateurs mobiles à grande échelle. Pour cet environnement, les auteurs proposent une méthode d'optimisation à deux niveaux, appelée ToDeTaS, qui vise à optimiser conjointement le déploiement des drones et l'ordonnancement des tâches, dans le but de minimiser la consommation d'énergie du système. Les performances de l'approche proposée ont été étudiée sur dix instances avec un nombre d'utilisateurs allant jusqu'à 1000. Les résultats ont montré l'efficacité de l'optimisation à deux niveaux et de la solution proposée selon différentes métriques de performance (nombre moyen de tâches accomplies, pourcentage moyen d'exécutions réussies, consommation d'énergie du système, etc.) [119].

Sous domaine d’étude	Exemples représentatifs d’articles étudiés	Conférence/Revue
Communication via systèmes aériens Réseaux cellulaires & UAV/robots	Mozaffari et al. 19 [1] Mishra et al. 21 [13]	IEEE Comm. Surv. & Tuto. IEEE Comm. Stand. Mag.
Conception & développement de prototypes de réseaux cellulaires auto-déployables	Moradi et al. 18 [4] Moradi et al. 19 [5] Moradi et al. 21 [6]	ACM MobiCom ACM GetMobile Comm. of the ACM
Placement optimal du réseau cœur dans les réseaux cellulaires auto-déployables	Oueis et al. 19 [10] Oueis et al. 19 [72] Zhang et al. 21 [16]	Elsev. Comput. Commun. IEEE WCNC IEEE WCNC
Placement optimal des stations de base aériennes	Shahzadi et al. 21 [50] Zhao et al. 22 [51]	Elsev. JNCA Elsev. Comput. Commun.
Placement optimal dans les réseaux de capteurs, ad-hoc & cellulaires	Kiess et al. 14 [74] Wong et al. 04 [78]	IEEE GLOBECOM IEEE LCN
Placement optimal dans les environnements NFV	Attaoui et al. 23 [81] Golkarifard et al. 21 [85]	IEEE TNSM Elsev. Comput. Netw.
Placement optimal dans le Mobile Edge Computing	S. Wang et al. 19 [108] Y. Wang et al. 20 [119]	JPDC IEEE Trans. on Cybernetics

TABLE 2.1 – Classification représentative des travaux étudiés

2.7 Conclusions

Dans ce chapitre, nous avons présenté un résumé des travaux représentatifs en rapport avec les problématiques abordées dans notre thèse. Nous nous sommes focalisés dans les deux premières parties sur le cœur de notre problème à savoir la conception et le développement de réseaux cellulaires auto-déployables et le problème particulier du placement optimal du réseau cœur. Nous avons ensuite parcouru différents travaux connexes liés au placement des stations de base aériennes, au placement des passerelles dans les différents réseaux conventionnels, et au placement dans les environnement NFV (Network Function Virtualization) et MEC (Mobile Edge Computing). Le tableau 2.1 résume les sous-domaines étudiés ainsi que quelques articles représentatifs de chacun d’eux.

Cette revue de la littérature montre l’intérêt important et croissant que ce domaine suscite dans les travaux de recherche. En effet, la très grande majorité des travaux étudiés sont très récents et sont publiés dans des conférences et revues du premier plan. Par ailleurs, cette revue de la littérature met également en évidence la complexité du sujet ainsi que la diversité des technologies, des objectifs et des contraintes à prendre en compte dans ce type d’environnement.

Chapitre 3

De nouveaux modèles pour le placement optimal des réseaux cœur

Contenu du chapitre

3.1	Notations et énoncé du problème	48
3.2	Modèles itératifs de la littérature	50
3.3	Notre modèle optimisant le placement du réseau cœur en maximisant les flux	52
3.3.1	Premier modèle intuitif (non-linéaire)	53
3.3.2	Linéarisation des contraintes de notre modèle	54
3.4	Extensions du modèle	56
3.5	Conclusions	58

Dans ce chapitre, nous présentons et formalisons le principal problème auquel nous nous attaquons dans cette thèse : le placement optimal des réseaux cœur dans les réseaux cellulaires auto-déployables. Nous discutons également des hypothèses et présentons les notations utilisées dans le reste de ce document. Nous introduisons brièvement les deux modèles itératifs de la littérature auxquels nous nous comparons. Ensuite, nous présentons notre modèle intuitif de placement optimal, qui implique certaines contraintes non linéaires. Nous expliquons en détail la linéarisation de ces contraintes avant de présenter le modèle résultant et d'en discuter les avantages. Nous concluons ce chapitre en mentionnant quelques extensions possibles du modèle proposé, dont certaines seront approfondies dans les chapitres suivants.

3.1 Notations et énoncé du problème

Comme cela a été discuté dans le premier chapitre introductif et dans l'état de l'art, l'utilisation des réseaux auto-déployables se confronte à des différents verrous scientifiques et techniques. Un des verrous les plus importants concerne le placement optimal du réseau cœur [9, 10, 16, 72]. Plus concrètement, étant donné un réseau cellulaire auto-déployable avec plusieurs stations de base, ayant des demandes de trafic hétérogènes, la question à laquelle nous nous attaquons est d'identifier de manière efficace la(les) meilleure(s) station(s) de base pour héberger le(s) réseau(x) cœur (CN) partagé(s). Sans perte de généralité, nous considérons d'abord un seul réseau cœur partagé.

Comme dans les travaux connexes de la littérature, nous proposons de modéliser le réseau auto-déployable sous la forme d'un graphe orienté pondéré $G(V, E)$ où V , l'ensemble des sommets, représente les stations de base (BS) et E , l'ensemble des arcs, modélise les liens de communication. Nous notons $n = |V|$ le nombre de sommets (stations de base). Le mot "nœuds" est aussi utilisé dans ce document pour désigner les sommets du graphe qui correspondent aux stations de base, numérotées de 1 à n .

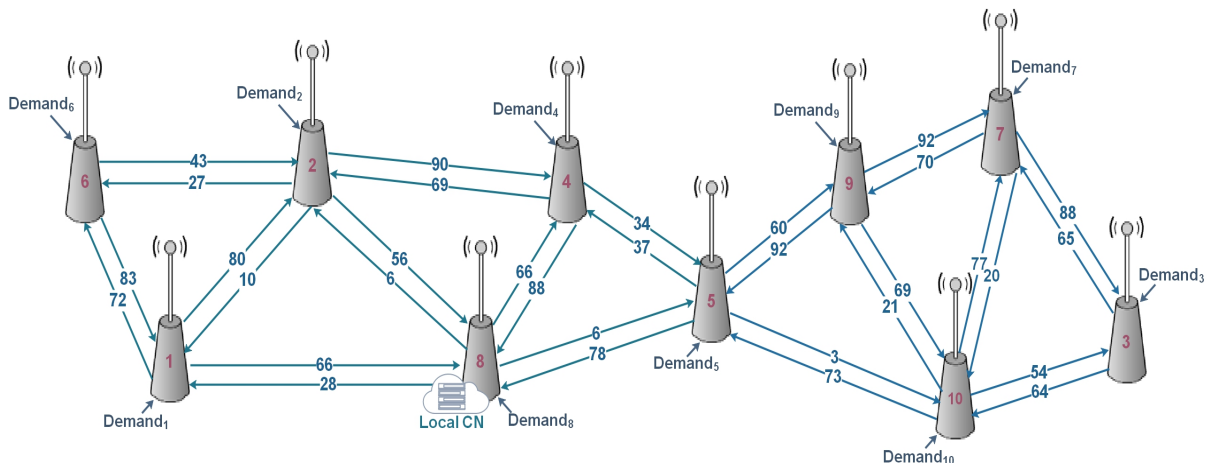


FIGURE 3.1 – Exemple d'un réseau auto-déployable avec 10 stations de base interconnectées et un réseau cœur local co-localisé avec une des stations de base

Afin de modéliser la bande passante limitée des liens de communication entre les stations de base, nous supposons que chaque arc $(i, j) \in E$ dans le graphe a une capacité limitée (représentant la quantité maximale de flux de trafic qui peut passer par le lien de la station de base i vers la station de base j), que nous notons $c(i, j) = c_{i,j}$.

Sans perte de généralité, nous supposons que les stations de base n'ont pas de contraintes de capacité de routage et que chaque station peut héberger le réseau cœur local. Nous notons que ces hypothèses peuvent être facilement assouplies en ajoutant des contraintes adéquates, comme nous le discuterons vers la fin de ce chapitre.

3.1 Notations et énoncé du problème

Paramètre	Signification
V	Ensemble de sommets (nœuds) corresp. aux stations de base, numérotées de 1 à n
E	Ensemble des arcs (liens de communication)
S	Ensemble de sommets (stations de base) autre que la destination (modèles itératifs)
n	Nombre de sommets (stations de base)
$c_{i,j} = c(i,j)$	Capacité du lien entre la station i et la station j (quantité maximale de flux de trafic)
$Demand_i$	Demande de flux de trafic provenant de la station de base i . Cela représente le flux propre à la station de base i qu'elle souhaite envoyer au réseau cœur à travers le réseau.
f_{ij}	Variable de décision représentant le flux de trafic depuis la station de base i vers la station de base j
f_{ij}^*	Valeur optimale du flux de trafic entre la station de base i et la station de base j
M	Très grande valeur fixe utilisée pour la linéarisation du problème
Q_i	Flux maximal de trafic qu'une station de base i peut router
CN_{max}	Nombre maximal de réseaux cœur (CN) à placer dans le réseau
y_i	Variable de décision indiquant si on doit positionner le réseau cœur à la station de base i ou pas

TABLE 3.1 – Principales notations utilisées

En tant que modèle de trafic, nous considérons que chaque station de base $i \in V$ a une demande de flux de trafic à envoyer au réseau cœur local (à travers le réseau formé) que nous notons $Demand_i$. Cette demande représente le flux propre à la station de base i qu'elle souhaite envoyer au réseau cœur. Nous souhaitons considérer le cas le plus général où cette demande de trafic peut être différente d'une station de base à l'autre. En effet cette demande de trafic dépend en pratique du nombre d'utilisateurs attachés à cette station de base ainsi que leurs flux de trafic respectifs. La figure 3.1 présente un exemple d'un réseau de 10 stations de base inter-connectées avec des capacités de communications asymétriques. Chaque station de base i souhaite envoyer un propre flux de trafic $Demand_i$ qui peut être hétérogène d'une station à l'autre.

En fonction de différentes métriques d'optimisation (maximisation du flux global envoyé par les stations de base, à titre d'exemple), l'objectif des modèles que nous visons sera de trouver en même temps la meilleure station de base qui devrait héberger le réseau cœur ainsi que les flux optimaux. Notre but est aussi de proposer des modèles génériques et scalables offrant des solutions rapidement même dans le cas des grands réseaux. Cela permettra de relancer les modèles en cas de changements de configurations et de procéder à des éventuelles migrations dans des temps raisonnables.

Afin de comparer les performances de notre modèle à celles de la littérature [9, 10] sous les mêmes hypothèses et objectifs, nous considérons tout d'abord comme métrique d'optimisation la maximisation du flux global de trafic que les stations de base peuvent envoyer au réseau cœur via le réseau formé. Le problème revient donc à trou-

ver le meilleur emplacement du réseau cœur qui maximise la somme des flux initiaux que les noeuds (stations de base) peuvent injecter dans le graphe associé en respectant les différentes contraintes. Nous montrons ensuite que notre modèle peut être facilement généralisé pour considérer d'autres métriques d'optimisation comme l'équité.

Les principales notations utilisées dans le reste de cette thèse sont résumées dans le tableau 3.1. Quelques constantes et variables du tableau seront introduites dans les prochaines sections au fur et à mesure de la présentation de notre modèle.

3.2 Modèles itératifs de la littérature

Les deux principaux modèles auxquels nous nous comparons sont les modèles itératifs à base de programme linéaire (LP) et à base de coupe minimale (Min-Cut) proposés dans [9, 10]. Les auteurs modélisent le réseau sous forme d'un graphe géométrique aléatoire pondéré et ils proposent des résolutions itératives. La première solution proposée est basée sur un programme linéaire (LP) qui permet de calculer la quantité maximal de trafic qu'un nœud d peut recevoir de tous les autres nœuds sources du graphe $G(V, E)$ s'il était choisi comme nœud destinataire hébergeant le réseau cœur. Ce modèle que nous appelons $LP(G(V, E), d, S)$ est présenté ci-dessous et suppose que toutes les stations de base i doivent injecter dans le réseau un flux $z_i(d)$ homogène appelé $\lambda(d)$ (Contrainte 3.2) que le modèle cherche à maximiser (Objectif 3.1). S représente ici l'ensemble des stations de base autre que la destination d . Ce modèle exige que tous les flux (notés $f_{i,j}$) doivent respecter la contrainte de capacité des liens (Contrainte 3.3), que toutes les stations de base intermédiaires doivent respecter les contraintes de conservation de flux (Contrainte 3.4) et que le nœud destinataire d doit récupérer le $(n - 1) \cdot \lambda(d)$ unités injectées dans le réseau (Contrainte 3.5) [9, 10].

Modèle $LP(G(V, E), d, S)$:

Objectif : Maximiser $\lambda(d)$ (3.1)

Sous les contraintes :

$$\forall i \in S, \quad z_i(d) = \lambda(d) \quad (3.2)$$

$$\forall (i, j) \in E, \quad f_{i,j} \leq c_{i,j} \quad (3.3)$$

$$\forall i \in S, \quad \sum_{k \in S} f_{k,i} = \sum_{j \in S} f_{i,j} \quad (3.4)$$

$$\sum_{i \in V} f_{i,d} = (n - 1) \cdot \lambda(d) \quad (3.5)$$

Au delà du fait que le modèle proposé nécessite que les stations de base envoient les mêmes flux initiaux de trafic (ce que nous appelons cas homogène des flux initiaux), il

nécessite de fixer la station de base destinataire d pour résoudre le problème et trouver le flux maximal. Ainsi, pour trouver la meilleure station de base, les auteurs proposent d'exécuter ce modèle pour toutes les stations de base potentielles (voir le pseudo algorithme 1) et de choisir le nœud (station de base) offrant le meilleur flux maximal pour héberger le réseau cœur.

Algorithm 1 Modèle itératif à base de LP [9, 10]

- 1: **Inputs** : $G(V, E)$
 - 2: **Outputs** : d^* (meilleure station de base), flux optimaux
 - 3: **for** d (destination potentielle) dans V **do**
 - 4: $S = V \setminus d$
 - 5: Résoudre le modèle $LP(G(V, E), d, S)$ présenté ci-dessous
 - 6: soit $\bar{\lambda}(d)$, la solution optimale donnée par la résolution de $LP(G(V, E), d, S)$
 - 7: **end for**
 - 8: $d^* = \operatorname{argmax}_{d \in V} \bar{\lambda}(d)$
-

Étant donné la dualité entre le problème de flux maximal (flot maximal) et de coupe minimale (Min-Cut), les auteurs proposent aussi une approche itérative équivalente basée sur le calcul de la coupe minimale dans un graphe. De même, l'approche proposée est basée sur une résolution itérative, tandis que le flux maximal de chaque configuration possible est calculé en utilisant un algorithme de calcul de coupe minimale au lieu du modèle LP. Le pseudo algorithme 2 fournit un aperçu de la solution itérative à base de Min-Cut et plus de détails peuvent être trouvés dans [9, 10].

Algorithm 2 Modèle itératif à base de Min-Cut [9, 10]

- 1: **Inputs** : $G(V, E)$
 - 2: **Outputs** : d^* (meilleure station de base), flux optimaux
 - 3: **for** d (destination potentielle) dans V **do**
 - 4: $S = V \setminus d$
 - 5: Résoudre le problème de coupe minimale dans le graphe G avec S comme sources et d comme destinataire
 - 6: soit $\bar{\lambda}(d)$, le flux maximal homogène par station de base
 - 7: **end for**
 - 8: $d^* = \operatorname{argmax}_{d \in V} \bar{\lambda}(d)$
-

Les auteurs de [9, 10] ne précisent pas un algorithme de calcul de coupe minimale à utiliser. Dans notre cas, nous considérons la méthode par défaut utilisée dans le package Python NetworkX¹ qui est basée sur l'algorithme "highest-label preflow-push" offrant une bonne complexité et que nous discuterons dans le chapitre suivant.

1. <https://networkx.org/>

3.3 Notre modèle optimisant le placement du réseau cœur en maximisant les flux

Les modèles proposés dans la littérature pour identifier le meilleur réseau cœur dans un réseau cellulaire auto-déployable se basent sur une approche itérative. En particulier, les deux solutions discutées dans la section précédente [9, 10] et celle proposée dans [72] sont basées sur des résolutions itératives. Une solution similaire était considérée dans [79] précédemment pour identifier la meilleure passerelle dans le cas des réseaux maillés sans fil. Ce travail a été brièvement discuté dans le chapitre d'état de l'art. Ces solutions itératives consistent notamment à proposer un modèle efficace qui optimise une métrique donnée en ayant une destination fixe et connue et à identifier le meilleur emplacement de la passerelle/du réseau cœur de manière itérative en considérant à chaque itération une destination potentielle. Par exemple, les deux solutions discutées dans la section précédente sont basées sur la fixation de la destination à chaque étape afin de calculer le flux maximal, et ensuite de choisir la meilleure destination qui permet d'atteindre le meilleur flux maximal.

Une des contributions de notre thèse est de proposer un nouveau modèle "one-shot" qui, contrairement aux modèles itératifs, optimise à la fois le meilleur emplacement du réseau cœur ainsi que les valeurs des flux en une seule itération. Pour ce faire, nous utilisons la programmation en nombres entiers mixtes, et on définit les variables de décisions binaires, y_i , qui sont égales à 1 si la station de base (nœud) i doit être sélectionnée comme étant la meilleure pour héberger le réseau cœur local et à 0 dans le cas contraire. Sans perte de généralité, nous supposons que nous essayons de localiser le réseau cœur dans une seule station de base et nous discuterons d'autres extensions à la fin de ce chapitre.

Nous considérons le cas général où les stations de base peuvent avoir des demandes de trafic différentes et des flux initiaux de trafic différents. Nous notons $f_{i,j}$ comme variables de décision représentant le flux de trafic depuis le nœud i vers le nœud j . Nous désignons par $f^*(i, j)$ les valeurs optimales du flux. Dans un objectif de simplification, nous proposons d'ajouter au graphe G un nœud virtuel 0 et des arcs allant du nœud virtuel 0 vers chacun des autres nœuds (station de base) i . Chaque arc $(0, i)$ aura une capacité égale à la demande de flux de trafic de la station de base i (c.-à-d $c(0, i) = Demand_i$). Ainsi, $f_{0,i}$ représente le flux initial de trafic qu'une station de base i va injecter dans le réseau et qui doit être inférieur ou égal à sa demande de flux ($Demand_i$). Nous précisons ici que le flux initial ($f_{0,i}$) fait référence, dans ce manuscrit, au **flux propre à la station de base i** , qui sera réellement envoyé vers le réseau cœur via le réseau formé dans une solution donnée.

3.3.1 Premier modèle intuitif (non-linéaire)

Le premier modèle intuitif que nous proposons consiste à maximiser la somme des flux initiaux envoyés par les stations de base (provenant des stations de base) i) en respectant les contraintes de capacité sur tous les liens, ii) en assurant la conservation du flux à chaque station de base et iii) en vérifiant que la station de base sélectionnée pour héberger le réseau cœur peut récupérer tous les flux initiaux injectés. Nous proposons donc le programme mixte en nombres entiers suivant :

$$\text{Objectif : Maximiser } \sum_{i \in [1, n]} f_{0,i} \quad (3.6)$$

Sous les contraintes :

$$\forall i \in [1, n], f_{0,i} \leq Demand_i \quad (3.7)$$

$$\forall i \in [1, n], \forall j \in [1, n], f_{i,j} \leq c_{i,j} \quad (3.8)$$

$$\forall i \in [1, n], \sum_{k \in [0, n]} f_{k,i} - \sum_{j \in [1, n]} f_{i,j} = \left(\sum_{i \in [1, n]} f_{0,i} \right) \cdot y_i \quad (3.9)$$

$$\sum_{i \in [1, n]} y_i = 1 \quad (3.10)$$

Variables de décision :

$$y_i \in \{0, 1\}, i \in [1, n];$$

$$f_{ij} \in \mathbb{R}^+, i \in [0, n], j \in [1, n]$$

Dans ce modèle de base, l'équation (3.6) représente l'objectif du modèle, qui est ici de maximiser la somme totale de tous les flux envoyés depuis le nœud virtuel vers tous les autres nœuds (qui correspondent aux flux initiaux des stations de base). Comme nous le discuterons plus tard, nous pouvons choisir d'autres métriques à optimiser selon les besoins comme l'équité à titre d'exemple. De son côté, la contrainte (3.7), assure que chaque flux initial de toute station de base i ne dépasse pas la demande de flux de trafic qu'elle souhaite envoyé au réseau cœur ($Demand_i$). La contrainte (3.8) permet de garantir les contraintes de capacité sur chaque lien de communication. Autrement dit elle assure que le flux qui traverse chaque lien de communication entre une station de base i et une station de base j est inférieur à la capacité maximale sur ce lien.

La contrainte (3.9) est très importante et assure que le flux est conservatif dans tous les nœuds intermédiaires et que la destination qui sera choisie par le modèle (i tel que $y_i = 1$) récupérera toutes les unités de flux. En effet, cette contrainte est équivalente à la suivante :

$$\forall i \in [1, n],$$

$$\sum_{k \in [0, n]} f_{k,i} - \sum_{j \in [1, n]} f_{i,j} = \begin{cases} 0, & \text{Si } y_i = 0 \\ \sum_{l \in [1, n]} f_{0,l}, & \text{Si } y_i = 1 \end{cases} \quad (3.11)$$

Cela veut dire que si une station de base (sommet du graphe) n'est pas l'emplacement optimal du réseau cœur ($y_i = 0$), la somme des flux entrant (y compris celui venant du sommet fictif 0) doit être égale à la somme de flux sortant. Par contre, si la station de base est l'emplacement optimal pour héberger le réseau cœur ($y_i = 1$) du réseau, la somme du flux entrant moins la somme du flux sortant doit être égale à la somme des flux envoyés par toutes les stations de base, ce qui veut dire que le réseau cœur récupère tout le flux injecté. Nous notons que l'équation (3.9) comme son équivalente 3.11 sont non linéaires ce qui rend les problèmes très difficiles à résoudre.

Enfin, la contrainte (3.10) assure qu'il n'y a qu'une seule et unique destination dans le réseau. Comme nous discuterons plus tard, nous considérons une seule destination dans un premier temps, car cela est plus simple et aussi nous permet de se comparer équitablement par rapport aux autres solutions de la littérature. Cependant cette contrainte peut facilement être généralisée comme nous le discuterons plus tard.

Il est important à noter que dans le cas où une seule destination est exigée (contrainte 3.10), notre modèle garantit implicitement que le réseau est connexe. En effet, aucune solution réalisable ne peut être trouvée dans le cas d'un graphe non connexe (réseau non connecté). Si une solution est trouvée, le graphe est connexe et le réseau associé est forcément connecté.

3.3.2 Linéarisation des contraintes de notre modèle

Le problème présenté ci-dessus n'est pas linéaire en raison de l'équation (3.9). En effet, y_i et $f_{0,i}$ sont des variables de décision. Une solution peut consister à résoudre le problème de manière itérative en fixant la potentielle station de base qui doit héberger le réseau cœur à chaque itération et en choisissant ensuite la destination qui offre le meilleur flux maximal à envoyer par les stations de base. Il en résulte une solution très similaire à celle proposée dans [9, 10]. Dans ce travail, nous proposons une solution plus efficace en linéarisant l'équation (3.9) ce qui nous permettra de parvenir à un programme linéaire mixtes en nombres entiers (MILP pour Mixed-Integer Linear Programming en anglais) permettant de résoudre le problème. Pour ce faire, nous proposons de linéariser l'équation (3.9) comme suit :

$$\forall i \in [1, n], \quad \sum_{k \in [0, n]} f_{k,i} - \sum_{j \in [1, n]} f_{i,j} \leq M \cdot y_i \quad (3.12)$$

$$\forall i \in [1, n], \quad \sum_{k \in [0, n]} f_{k,i} - \sum_{j \in [1, n]} f_{i,j} \geq 0 \quad (3.13)$$

Où M est un grand nombre qui doit être supérieur ou égal à la valeur maximale possible de $\sum_{i \in [1, n]} f_{0,i}$. Pour simplifier, nous choisissons $M = \sum_{i \in [1, n], j \in [1, n]} c_{i,j}$.

D'une part, lorsque $y_i = 0$ (nœuds qui ne sont pas le destinataire optimal), les équations (3.12) et (3.13) combinées garantissent la conservation du flux. D'autre part, lorsque $y_i = 1$, les deux équations relâchent la contrainte au niveau du destinataire (station de base qui doit héberger le réseau cœur). Lorsqu'elles sont combinées aux équations de conservation du flux dans les autres nœuds, la seule valeur possible de $(\sum_{k \in [0, n]} f_{k,i} - \sum_{j \in [1, n]} f_{i,j})$ qui garantit une solution réalisable du problème MILP global est $\sum_{i \in [1, n]} f_{0,i}$. Par conséquent, le problème linéaire mixte en nombres entiers équivalent au programme non linéaire peut être donné comme suit :

Objectif : Maximiser $\sum_{i \in [1, n]} f_{0,i}$

Sous les contraintes :

Contraintes de demande de flux de trafic : (3.7)

Contraintes de capacité : (3.8)

Contraintes linéarisées de conservation de flots : (3.12)(3.13)

Contrainte liée au nombre de CN : (3.10)

Variabes de décision :

$y_i \in \{0, 1\}, i \in [1, n];$

$f_{ij} \in \mathbb{R}^+, i \in [0, n], j \in [1, n]$

Notre modèle ainsi linéarisé présente plusieurs atouts dont les plus importants peuvent être résumés comme suit :

- **Efficacité en termes de temps d'exécution :** Grâce à la linéarisation du modèle "one-shot", notre intuition était d'atteindre des temps d'exécution efficaces en comparaisons aux autres modèles itératifs proposés dans la littérature. Afin de

confirmer ce point, nous avons mené plusieurs simulations dont les résultats sont présentés et discutés dans le prochain chapitre.

- **Possibilité de considérer des demandes et des flux initiaux hétérogènes** : En effet, notre modèle intègre nativement l'hétérogénéité des demandes de flux de trafic ($Demand_i$) ainsi que celle des flux initiaux réellement envoyés par les stations de base vers le réseau cœur ($f_{0,i}$).
- **Généricité** : En effet, notre modèle permet de considérer différentes métriques d'optimisation et de contraintes. D'un côté, il s'adapte facilement à différentes métriques, y compris des métriques d'équité, sur lesquelles nous reviendrons en détail dans le chapitre 5. D'autre part, plusieurs extensions peuvent être mises en œuvre en mettant à jour les contraintes existantes ou en ajoutant de nouvelles, comme nous l'illustrons dans la section suivante.

3.4 Extensions du modèle

Comme nous venons de mentionner, la généralité du modèle proposé facilite son extension et son adaptation selon les caractéristiques et les objectifs du déploiement. Parmi d'autres extensions ou adaptations possibles, nous résumons dans cette section quatre dont les deux dernières sont exploitées et évaluées dans les chapitres suivants.

Contrainte de routage sur les stations de base

Dans certaines situations, quelques stations de base peuvent présenter des contraintes de routage en fonction de leurs ressources limitées de calcul, de mémoire ou d'énergie. Ces contraintes peuvent aussi être hétérogènes. Nous notons Q_i le flux maximal de trafic qu'une station de base i peut router. Afin de prendre en compte ces éventuelles contraintes, il suffit d'ajouter les deux contraintes suivantes 3.14 et 3.15 qui assurent respectivement que le flux entrant à chaque station de base i ne dépasse pas Q_i et que le flux sortant de la station de base i ne dépasse pas Q_i .

$$\forall i \in [1, n], \quad \sum_{k \in [0, n]} f_{k,i} \leq Q_i \quad (3.14)$$

$$\forall i \in [1, n], \quad \sum_{k \in [1, n]} f_{i,k} \leq Q_i \quad (3.15)$$

Limitation des stations de base pouvant accueillir le réseau cœur

Pour les mêmes raisons citées ci-dessous certaines stations de base peuvent être facilement exclues de l'ensemble des stations de base potentielles pour héberger le

réseau cœur (à cause de contraintes de ressources par exemple). Afin de considérer cette éventuelle limite, il suffit d'ajouter des contraintes forçant les y_i des stations de base inéligibles à zéro.

Considération d'autres fonctions d'optimisation

Comme cela sera discuté plus en détail dans le chapitre 5, il est parfois nécessaire de changer les objectifs d'optimisation selon les besoins d'utilisation. Il est facile d'adapter ces objectifs dans notre modèle. À titre d'exemple, afin de maximiser la somme de la satisfaction proportionnelle des flux par rapport aux demandes de stations de base, il suffit de remplacer la fonction d'optimisation 3.6 par la fonction 3.16. D'autres métriques plus complexes seront étudiées dans le chapitre 5 lorsque nous nous intéressons au problème d'équité entre les stations de base.

$$\text{Objectif : Maximiser } \sum_{i \in [1, n]} \frac{f_{0,i}}{c_{0,i}} \quad (3.16)$$

Prise en compte de plusieurs réseaux cœur

Comme discuté précédemment, les fonctions du réseau cœur peuvent être hébergées dans chaque station de base, comme considéré dans [4, 5, 6] ou dans une seule station de base qui héberge un réseau cœur local comme dans notre cas et le cas étudié dans [9, 10, 16]. Néanmoins, il pourrait être envisageable et opportun dans certaines situations de prévoir un nombre réduit de réseaux cœur avec des procédures de coordination adéquate afin de partager la charge dans un réseau auto-déployable. Si on note CN_{max} le nombre maximal de réseaux cœur à déployer, notre modèle peut être facilement étendu à ce cas en changeant l'équation 3.10 par la contrainte 3.17.

$$\sum_{i \in [1, n]} y_i \leq CN_{max} \quad (3.17)$$

Il est à noter ici que nous proposons une inégalité et non pas une égalité car selon la fonction objective considérée, il est parfois mieux de placer moins de réseaux cœur et l'inégalité laisse au modèle la possibilité de choisir à la fois le nombre optimal de réseaux cœur, leurs emplacements et les flux optimaux. Dans le cas où on souhaite déployer exactement CN_{max} réseaux cœur, il suffit de remplacer l'inégalité par une égalité.

3.5 Conclusions

Dans ce chapitre, nous avons proposé un modèle permettant de choisir la meilleure station de base pour héberger le réseau cœur, tout en maximisant le flux global envoyé par les stations de base vers le réseau cœur dans un réseau auto-déployable. Notre modèle s'exécute en une seule itération ("one-shot") évitant une recherche itérative de la meilleure station de base. Notre premier modèle intuitif étant non linéaire, nous avons linéarisé les contraintes non linéaires et présenté et discuté le modèle résultant. Nous avons aussi montré la généralité de notre modèle et discuté quelques extensions possibles.

Grâce à notre modélisation "one-shot" évitant le calcul itératif, et à la linéarisation des contraintes non linéaires, notre intuition est de pouvoir réduire le temps de calcul par rapport aux solutions itératives ce qui permettrait un meilleur passage à l'échelle tout en offrant les mêmes solutions optimales. Cette évaluation de performances et les comparaisons aux solutions de la littérature feront l'objet du prochain chapitre.

Chapitre 4

Évaluation de performances des modèles proposés

Contenu du chapitre

4.1	Configuration et paramètres des simulations	60
4.2	Validation et évaluation des performances dans le contexte homogène des flux initiaux des stations de base	61
4.3	Validation et évaluation des performances dans le contexte hétérogène des demandes de flux et des flux initiaux	63
4.3.1	Évaluation de performances en fonction de la taille du réseau .	64
4.3.2	Évaluation de performances en fonction de la connectivité du réseau	66
4.4	Conclusions	67

Dans ce chapitre, nous nous focalisons sur l'implémentation de nos modèles, l'évaluation de leurs performances et la comparaison avec des solutions de la littérature. Dans un premier temps, nous présentons la configuration de simulation et les outils utilisés pour implémenter nos solutions. Nous présentons à cette occasion les paramètres par défaut utilisés dans l'évaluation de performance. Dans un deuxième temps, nous présentons et discutons les différents résultats obtenus en termes de positions optimales du réseau cœur et en termes de temps de calcul dans le cas où le flux initial des stations de base est homogène. Nous comparons à cette occasion notre solution à deux solutions de la littérature. Nous analysons les résultats obtenus et nous tirons quelques conclusions. La dernière partie de ce chapitre est dédiée aux cas où les demandes de flux de trafic ainsi que les flux initiaux peuvent être hétérogènes. Là aussi, nous comparons notre modèle aux solutions de la littérature que nous adaptons au cas hétérogène et nous présentons et discutons les différents résultats.

4.1 Configuration et paramètres des simulations

Afin de valider notre modèle et de mettre en évidence ses fonctionnalités et ses avantages, nous effectuons des simulations approfondies tout en comparant nos solutions à deux solutions existantes. Sans perte de généralité, nous considérons dans nos simulations une topologie de graphe géométrique aléatoire telle que considéré dans plusieurs travaux de la littérature [9, 10]. Dans une telle topologie, n nœuds (stations de base) sont aléatoirement déployés dans un carré unité (un carré dont les côtés ont une longueur de 1). Deux sommets (nœuds) qui correspondent aux stations de base peuvent communiquer si la distance qui les sépare ne dépasse pas un rayon donné appelé ρ .

Afin de simuler des liens asymétriques, nous considérons, pour chaque arrête, deux arcs unidirectionnels avec des capacités différentes. Nous notons que n et ρ caractérisent complètement le graphe géométrique aléatoire, où ρ a un impact direct sur le degré de connectivité des nœuds et donc sur le nombre moyen de liens de communication (arcs dans le graphe géométrique aléatoire dirigé). Évidemment, plus la valeur de ρ est élevée, plus le degré de connectivité est plus élevé. Une valeur de $\rho = 1$ correspond à un graphe complet (un réseau où chaque station de base peut communiquer directement avec n'importe quelle autre station de base). Nous considérons différentes valeurs de ρ dans nos simulations mais nous ne considérons que les graphes connectés (stations de base qui peuvent communiquer entre elles directement ou indirectement via un routage multi-saut).

Toutes les simulations ont été effectuées sous Python 3.6 en utilisant l'optimiseur CPLEX d'IBM ILOG pour résoudre les modèles MILP (Mixed-Integer Linear Programs) déjà introduits et les modèles MIQP (Mixed-Integer Quadratic Programs) qui seront introduits plus loin. Nous considérons dans ce chapitre un réseau de 10 à 100 stations de base avec ρ allant de 0.2 à 1. Un nombre plus important de stations de base sera considéré plus loin afin d'étudier la scalabilité de nos modèles dans des réseaux de très grande taille. Nous notons également que les capacités des liens sont tirées aléatoirement entre 10 et 100 unités selon une distribution uniforme. Tous les résultats présentés dans ce chapitre sont des moyennes de 30 réalisations aléatoires (valeur par défaut) et les intervalles de confiance avec un taux de confiance de 98% sont tracés dans les différentes courbes. Dans les chapitres suivants, il a parfois été nécessaire d'aller au-delà de 30 réalisations aléatoires, et cela sera mentionné dans les différentes simulations concernées. Nous résumons dans le tableau 4.1 les paramètres de simulations par défaut utilisés dans nos tests.

4.2 Validation et évaluation des performances dans le contexte homogène des flux initiaux des stations de base

Paramètre	Valeur par défaut
n	100 (à noter que plusieurs simulations font varier ce paramètre. 100 est la valeur par défaut si aucune autre valeur n'est mentionnée)
Nombre de répétitions	30
Taux de confiance	98%
ρ	0.2 (à noter que plusieurs simulations font varier ce paramètre. 0.2 est la valeur par défaut si aucune autre valeur n'est mentionnée)
$c(i, j)$	Génération selon une distribution aléatoire uniforme entre 10 et 100
$Demand_i$	Génération selon une distribution aléatoire uniforme entre 10 et 100
Q_i	∞
CN_{max}	1

TABLE 4.1 – Paramètres de simulations par défaut

4.2 Validation et évaluation des performances dans le contexte homogène des flux initiaux des stations de base

Pour une comparaison équitable de notre approche avec les modèles proposés dans [10], nous considérons dans un premier temps que toutes les stations de base doivent injecter dans le réseau un même flux initial homogène que le modèle cherche à maximiser, comme considéré dans l'article [10]. Dans ce cas, notre modèle peut être adapté et simplifié comme suit (Nous forçons les flux initiaux à être identiques et relâchons les contraintes sur les demandes de trafic) :

$$\textbf{Objectif : Maximiser } \Lambda \tag{4.1}$$

Sous les contraintes :

$$\forall i \in [1, n], f_{0,i} = \Lambda \tag{4.2}$$

Contraintes de capacité : (3.8)

Contraintes linéarisées de conservation de flots : (3.12)(3.13)

Contrainte liée au nombre de destinations : (3.10)

Variables de décision :

$$y_i \in \{0, 1\}, i \in [1, n];$$

$$\Lambda \in \mathbb{R}^+, f_{ij} \in \mathbb{R}^+, i \in [0, n], j \in [1, n]$$

Étant donné n , la taille du réseau (nombre de stations de base), nous notons ici que les solutions réalisables garantissent que nous récupérons toutes les unités injectées à la destination grâce à nos équations linéaires (3.12) et (3.13). Nous évaluons ce modèle en faisant varier le nombre de nœuds de 10 à 100, et en fixant le rayon ρ à 0.2.

Chapitre 4. Évaluation de performances des modèles proposés

Nombre de nœuds	Modèle itératif LP [10]			Notre modèle		
	Flux max moyen	Temps exe. moyen (s)	Meilleure CN (1 ^{ère} inst.)	Flux max moyen	Temps exe. moyen (s)	Meilleure CN (1 ^{ère} inst.)
10	6.168	0.170	1	6.168	0.074	1
20	10.376	0.405	4	10.376	0.098	4
30	14.543	0.791	8	14.543	0.114	8
50	17.552	2.345	21	17.552	0.185	21
70	17.914	5.331	49	17.914	0.284	49
100	16.966	14.058	23	16.966	0.636	23

TABLE 4.2 – Résultats de la comparaison entre le modèle itératif basé sur LP et notre modèle, en termes de flux maximal, de temps d’exécution et du meilleur emplacement du réseau cœur (moyennes sur 30 itérations - cas des flux initiaux homogènes)

En comparant notre modèle en une itération (one-shot) avec le modèle itératif basé sur LP proposé dans [10], nous donnons d’abord dans le tableau 4.2 le meilleur emplacement du réseau cœur de la première instance, le flux maximal moyen et le temps d’exécution moyen sur 30 itérations aléatoires.

Les résultats montrent que le meilleur réseau cœur choisi par les deux modèles est le même dans les instances sélectionnées. Nous notons que très peu d’instances aboutissent à multiples configurations optimales offrant la même valeur optimale avec un emplacement différent du réseau cœur. Le tableau 4.2 montre également que le flux maximal moyenné sur toutes les itérations est exactement le même. Nous avons vérifié que cette valeur optimale est la même pour toutes les instances. À noter que le flux maximal spécifié ici est celui envoyé par chaque station de base, et il est identique entre toutes les stations de base (cas des flux initiaux homogènes).

Cependant, les résultats montrent que notre modèle réduit de manière significative le temps d’exécution et que plus le nombre de nœuds est élevé, plus le taux d’amélioration est important. Par exemple, notre modèle permet de réduire le temps d’exécution moyen par un facteur de 2.30 (resp. 12.68 et 22.10) pour 10 (resp. 50 et 100) nœuds (voir le tableau 4.2). Les résultats concernant le temps d’exécution en fonction du nombre de nœuds sont également représentés dans la figure 4.1 avec des intervalles de confiance calculés à un niveau (taux) de confiance de 98%.

Nous remarquons également à partir de ces résultats que le modèle proposé permet un meilleur passage à l’échelle en fonction du nombre de nœuds par rapport au modèle itératif. Par exemple, en multipliant le nombre de nœuds par 5 (de 20 à 100), le temps d’exécution est multiplié par 34.71 dans le modèle itératif, alors que le temps d’exécution de notre modèle est multiplié par 6.49. Ce résultat est obtenu principalement grâce à la modélisation linéaire en une seule itération proposée dans notre approche.

4.3 Validation et évaluation des performances dans le contexte hétérogène des demandes de flux et des flux initiaux

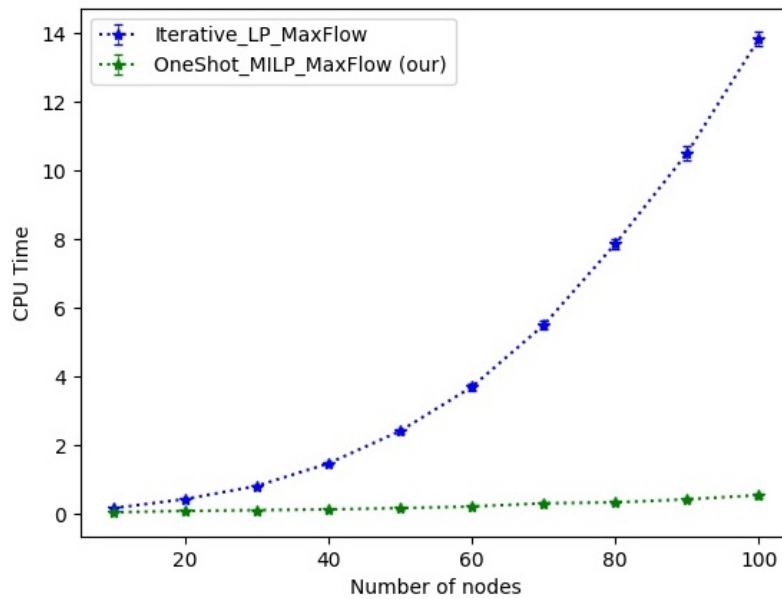


FIGURE 4.1 – Temps moyen d’exécution (en secondes) en fonction du nombre de nœuds n (moyennes sur 30 itérations - cas des flux initiaux homogènes)

4.3 Validation et évaluation des performances dans le contexte hétérogène des demandes de flux et des flux initiaux

Nous reconsidérons maintenant le cas plus général où des demandes de flux de trafic hétérogènes sont prises en compte, et où les flux initiaux des stations de base peuvent être différents (nous appelons ce contexte le cas hétérogène). Dans ce cas, nous utilisons notre modèle linéaire proposé dans la section 3.3.2 alors que l’objectif est de maximiser la somme des flux initiaux envoyés par les stations de base.

Pour une comparaison équitable, nous adaptons le modèle LP itératif proposé dans [10] pour considérer les demandes de trafic et le cas hétérogène des flux initiaux. Nous y parvenons en ajoutant un nœud source virtuel avec des arcs vers les autres nœuds et en contraignant le flux initial de chaque nœud ($f_{0,i}$) à ne pas dépasser $Demand_i$.

Nous considérons également ici le modèle itératif à base de coupe minimale (Min-Cut) [10]. Nous rappelons que cette approche est également basée sur une résolution itérative, tandis que le flux maximal de chaque configuration possible est calculé en utilisant un algorithme de coupe minimale (Min-Cut) plutôt que du modèle LP proposé. Les auteurs de [10] ne précisent pas un algorithme Min-Cut à utiliser. Dans notre cas, nous considérons la méthode par défaut utilisée dans le package Python NetworkX¹ qui est basée sur l’algorithme "highest-label preflow-push". Cet algorithme est largement utilisé grâce à son temps d’exécution efficace de $O(n^2\sqrt{m})$ où n est le nombre

1. <https://networkx.org/>

de nœuds et m est le nombre des arcs. Nous notons que cet algorithme nécessite une source et une destination fixes pour fonctionner, ce qui signifie que son utilisation itérative dans notre cas se fait en fixant la destination à chaque étape ce qui donne une complexité globale de $O(n^3\sqrt{m})$.

Nombre de nœuds	Modèle itératif LP [10]		Modèle Min-Cut [10]		Notre modèle	
	Flux max. moyen	Temps exe. moyen (s)	Flux max. moyen	Temps exe. moyen (s)	Flux max. moyen	Temps exe. moyen (s)
10	302.633	0.163	302.633	0.014	302.633	0.033
20	454.633	0.409	454.633	0.069	454.633	0.039
30	622.800	0.793	622.800	0.224	622.800	0.050
50	1008.600	2.308	1008.600	0.915	1008.600	0.083
70	1296.700	5.246	1296.700	2.262	1296.700	0.160
100	1733.067	13.488	1733.067	5.840	1733.067	0.278

TABLE 4.3 – Résultats de la comparaison du modèle itératif basé sur LP, du modèle itératif basé sur Min-Cut et de notre modèle, en termes de flux maximal et de temps d'exécution en fonction de n (moyennes sur 30 itérations - cas hétérogène)

4.3.1 Évaluation de performances en fonction de la taille du réseau

Pour le cas hétérogène, nous avons comparé notre solution basée sur le modèle MILP proposé dans la section 3.3.2 au modèle itératif à base de LP et au modèle itératif à base de Min-Cut [10]. Nous considérons les capacités des liens uniformément réparties entre 10 et 100 unités, et nous générons aléatoirement des demandes hétérogènes de trafic des stations de base entre 10 et 100 selon une distribution uniforme. Comme nous l'avons fait précédemment, nous avons vérifié que les valeurs optimales obtenues sont les mêmes en utilisant les trois modèles, et nous avons évalué le temps d'exécution des trois modèles en fonction du nombre de nœuds (n) dans un premier temps. Nous avons fixé ici ρ à 0.2 et les résultats sont moyennés sur 30 itérations.

Les résultats résumés d'abord dans le tableau 4.3 confirment le mérite de notre modèle. En effet, on observe que le flux maximal moyenné sur toutes les itérations est exactement le même (ainsi que le flux maximal sur toutes les instances). On note que le flux maximal spécifié ici est la somme de tous les flux initiaux optimaux envoyés par les stations de base. On remarque par ailleurs que notre solution réduit considérablement le temps d'exécution par rapport à la solution itérative à base de Min-Cut et encore plus par rapport au modèle itératif à base de LP. A titre d'exemple, pour un réseau de 100 stations de base, notre modèle permet de réduire le temps d'exécution moyen par un facteur de 21 par rapport au modèle itératif à base de Min-Cut et par un facteur de 48.51 par rapport au modèle itératif à base de LP (voir le tableau 4.3).

4.3 Validation et évaluation des performances dans le contexte hétérogène des demandes de flux et des flux initiaux

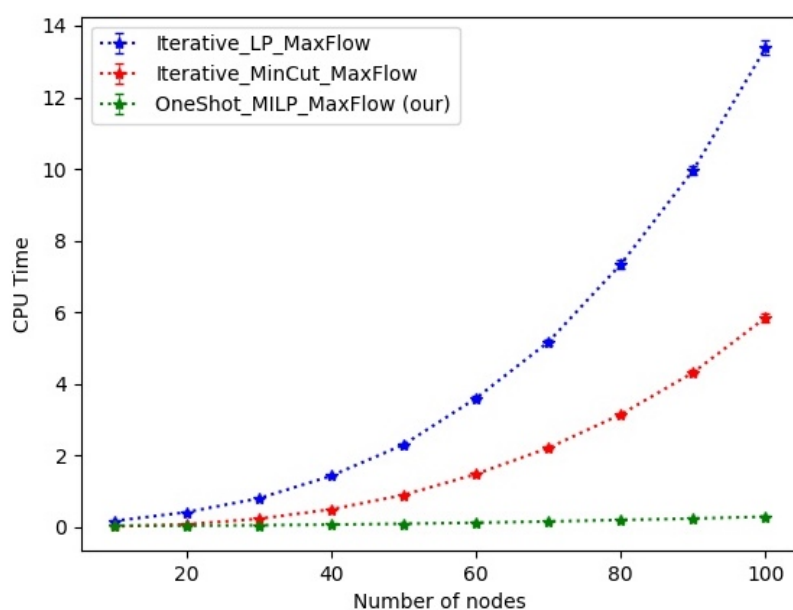


FIGURE 4.2 – Temps moyen d’exécution (en secondes) en fonction du nombre de nœuds n (moyennes sur 30 itérations - cas hétérogène)

La figure 4.2 trace les mêmes résultats concernant le temps d’exécution en fonction du nombre de stations de base avec des intervalles de confiance calculés à un niveau (taux) de confiance de 98%.

Les résultats démontrent que les performances de notre modèle par rapport à celui basé sur une approche itérative à base de LP sont très similaires à celles obtenues dans le cas homogène des flux initiaux. Ils montrent également que le modèle itératif à base de Min-Cut offre une complexité intermédiaire en fonction du nombre de nœuds. En effet, les résultats montrent que le temps d’exécution moyen est presque le même lorsque le nombre de stations de base est très faible. Cependant, ce temps augmente de manière significative dans le modèle itératif à base de Min-Cut et de manière encore plus significative dans le modèle itératif à base de LP par rapport à notre modèle. En effet, ce temps atteint pour 100 nœuds 0.28 secondes en utilisant notre modèle contre 5.84 secondes en utilisant le modèle itératif Min-Cut et 13.49 secondes en utilisant le modèle itératif LP. Ces résultats montrent que la complexité de notre modèle en fonction du nombre de stations de base est bien meilleure que les deux autres et confirment les bonnes performances de notre solution.

4.3.2 Évaluation de performances en fonction de la connectivité du réseau

Afin d'évaluer et de comparer les performances de notre modèle en termes de degré de connectivité du réseau, nous évaluons le temps d'exécution en fonction du rayon de connectivité. Dans ce cas, nous faisons varier les valeurs du rayon ρ de 0.2 à 1.0, tout en fixant le nombre de nœuds à 50.

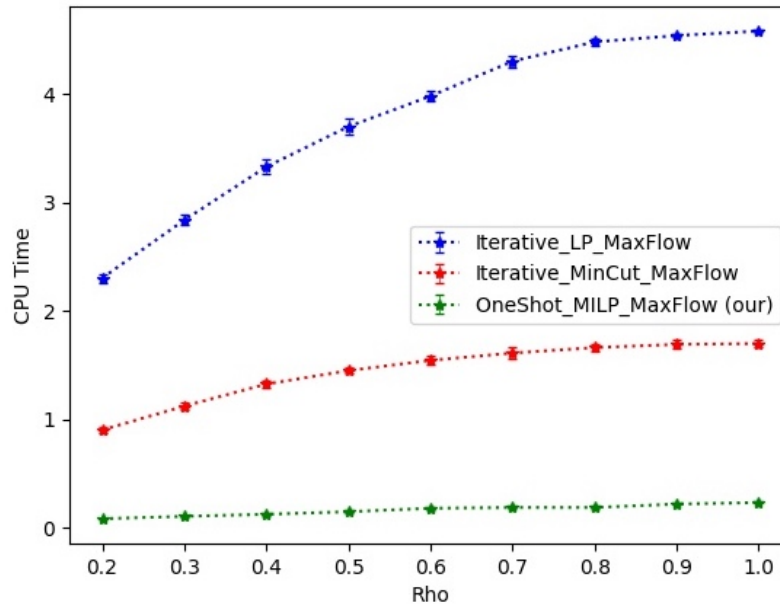


FIGURE 4.3 – Temps moyen d'exécution (en secondes) en fonction du rayon de connectivité ρ ($n = 50$ - cas hétérogène)

La figure 4.3 présente la moyenne des résultats sur 30 itérations avec des intervalles de confiance calculés à un niveau de confiance de 98%. Ils montrent que plus la valeur de ρ est élevée, plus le temps d'exécution est important pour tous les modèles. Cela était prévisible, car des valeurs ρ plus élevées entraînent un plus grand nombre de liens dans le réseau et, par conséquent, un temps d'exécution plus élevé.

Les résultats plus détaillés sont présentés dans le tableau 4.4. Ils montrent que notre modèle offre de meilleures performances avec un temps d'exécution moyen allant de 0.08 à 0.22 secondes pour ρ allant de 0.2 à 1.0. Ce temps d'exécution est très faible par rapport au temps d'exécution de la solution itérative Min-Cut allant de 0.90 à 1.67 secondes et de la solution itérative LP allant de 2.26 à 4.57 secondes.

Cependant, nous remarquons ici que le taux d'amélioration offert par notre modèle par rapport au modèle itératif Min-Cut (resp. au modèle itératif LP) diminue de 11.25 (resp. 28.25) pour $\rho = 0.2$ à 7.59 (resp. 20.77) quand $\rho = 1.0$. Même si l'amélioration reste importante, nous avons voulu comprendre et analyser cette baisse de fac-

ρ	Modèle itératif LP [10]		Modèle Min-Cut [10]		Notre modèle	
	Flux max. moyen	Temps exe. moyen (s)	Flux max. moyen	Temps exe. moyen (s)	Flux max. moyen	Temps exe. moyen (s)
0.2	930.367	2.259	930.367	0.899	930.367	0.084
0.3	1287.467	2.825	1287.467	1.123	1287.467	0.101
0.4	1473.500	3.275	1473.500	1.311	1473.500	0.124
0.5	1713.633	3.747	1713.633	1.469	1713.633	0.154
0.6	1847.500	4.016	1847.500	1.591	1847.500	0.182
0.7	1829.767	4.285	1829.767	1.620	1829.767	0.197
0.8	1930.033	4.477	1930.033	1.672	1930.033	0.186
0.9	1907.433	4.548	1907.433	1.686	1907.433	0.220
1.0	1945.100	4.571	1945.100	1.675	1945.100	0.225

TABLE 4.4 – Résultats de la comparaison du modèle itératif basé sur LP, du modèle itératif basé sur Min-Cut et de notre modèle en termes de flux maximal et de temps d'exécution en fonction de ρ (moyennes sur 30 itérations - cas hétérogène)

teur d'amélioration. En analysant le tableau 4.4, on remarque tout d'abord que l'augmentation du rayon de connectivité ρ (et donc du taux de connectivité) augmente le flux maximal de manière significative dans un premier temps mais que l'amélioration commence à se resserrer à partir d'un certain seuil. Par ailleurs en analysant quelques instances de solutions, nous avons remarqué que des rayons élevés de connectivité facilitent énormément les possibilités de routage de flux car le nombre de liens directs entre les stations de base augmente de manière significative. Cela facilite implicitement la recherche de la meilleure station de base pour accueillir le réseau cœur.

Il convient de mentionner que les valeurs élevées de ρ ne sont pas habituelles dans les scénarios réels. Par exemple, $\rho = 1$ correspond à un graphe complet, c'est-à-dire à un réseau entièrement connecté où chaque station de base peut communiquer directement avec chaque autre station de base, ce qui est le cas extrême.

4.4 Conclusions

Ce chapitre a été consacré à l'évaluation des performances du modèle proposé dans le chapitre précédent. Après avoir présenté la configuration de simulation et les outils utilisés pour implémenter nos solutions, nous avons exposé et discuté les différents résultats obtenus en termes de positions optimales et de temps de calcul, tant dans le cas où les flux initiaux des stations de base sont homogènes que dans le cas plus général où les demandes de flux des stations de base ainsi que les flux initiaux peuvent être hétérogènes. Nous avons également comparé notre solution à deux approches de la littérature basées sur une résolution itérative.

Chapitre 4. Évaluation de performances des modèles proposés

Les résultats obtenus montrent que notre approche arrive à atteindre les mêmes solutions optimales dans des temps d'exécution très faibles. Nous avons étudié les améliorations selon la taille des réseaux et la connectivité de ces derniers et les résultats montrent des facteurs de gain importants. Ces améliorations s'expliquent par le fait que les autres modèles itératifs fixent à chaque itération un nœud comme destination, puis choisissent la meilleure destination potentielle. En revanche, notre modèle, que nous avons réussi à linéariser, permet de sélectionner en une seule itération la meilleure station de base pour localiser le réseau cœur, tout en optimisant une fonction objective donnée.

Chapitre 5

Gestion de l'équité entre les stations de base

Contenu du chapitre

5.1 Motivations	70
5.2 Métriques prises en compte	71
5.3 Modèles optimaux d'équité	72
5.3.1 Maximiser la satisfaction proportionnelle moyenne	72
5.3.2 Minimiser la variance des satisfactions proportionnelles	72
5.3.3 Maximiser la satisfaction proportionnelle minimale	73
5.4 Évaluation des modèles optimaux d'équité	74
5.4.1 Évaluation des performances en termes de métriques de flux et d'équité	74
5.4.2 Évaluation du temps d'exécution	76
5.5 Conclusions	78

Après avoir validé notre modèle de placement optimal maximisant le flux global envoyé par les stations de base et évalué ses performances, nous nous attaquons dans ce chapitre au deuxième objectif de la thèse concernant l'équité entre les stations de base. En effet, une maximisation du flux global envoyé par les stations de base (somme des flux initiaux) peut engendrer un déséquilibre important de satisfactions absolues et relatives entre les stations de base. Nous explorons dans ce chapitre ce volet en commençant par exposer quatre métriques en termes de flux et d'équité. Nous présentons ensuite trois variantes de notre modèle initial qui nous permettent de mieux comprendre et mieux adresser le problème d'équité. Nous présentons et analysons les résultats de simulation en fonction des quatre métriques. A la lumière de ces résultats, nous discutons aussi le compromis entre la maximisation du flux de trafic et l'équité entre les stations de base et nous finissons ce chapitre par une évaluation du temps d'exécution des quatre modèles étudiés dans ce chapitre.

5.1 Motivations

Dans les chapitres précédents, nous avons proposé un modèle en une itération qui permet de trouver à la fois la meilleure station de base pour héberger le réseau cœur local et de maximiser le flux global (somme des flux initiaux) envoyé par les stations de base. La métrique de maximisation du flux global est largement utilisée dans la littérature dans de multiples applications, mais elle peut conduire à un partage inéquitable entre le flux des différentes stations de base, à la fois en termes de flux absolu et du flux relatif par rapport à la demande des stations de base. Nous avons dressé ce constat à plusieurs reprises lors la validation de notre modèle maximisant le flux global et le déséquilibre était souvent très marqué. A titre d'exemple, le tableau 5.1 montre les demandes de trafic des stations de base et les flux initiaux optimaux pour un réseau de 10 stations de base (avec $\rho = 0.3$). On remarque clairement un déséquilibre à la fois en termes de flux optimaux absolus et de taux relatif de satisfaction par rapport aux demandes de chaque station.

BS Id	$c_{0,i} = Demand_i$	$f_{0,i}^*$	Taux de satisfaction proportionnelle $\frac{f_{0,i}^*}{Demand_i}$
1	24	4	16.66 %
2	63	41	65.08%
3	22	0	0%
4	52	52	100%
5	94	94	100%
6	85	10	11.76%
7	78	78	100%
8	16	1	6.25%
9	78	5	6.41%
10	57	57	100 %

TABLE 5.1 – Exemple de trafic optimal avec maximisation du flux global dans le réseau ($\rho = 0.3$ et $n = 10$). La position optimale du réseau cœur est la station 5 dans cet exemple

Motivés par ce fait et encouragés par l'efficacité de notre modèle en terme de temps d'exécution, nous avons souhaité considérer l'équité dans nos modèles. L'équité dans l'ingénierie du trafic a été déjà étudiée dans la littérature pour les réseaux conventionnels avec des sources et des destinations fixes et connues, et plusieurs métriques d'équité ont été prises en compte [120, 121, 122, 123]. Sans perte de généralité, nous nous intéressons à la satisfaction proportionnelle par rapport aux demandes de flux des stations de base et à l'équité entre stations de base en termes de ces taux de satisfaction. Nous montrerons dans ce qui suit que notre modèle "one-shot" peut facilement s'adapter afin de calculer efficacement les meilleures solutions qui optimisent des métriques d'équité tout en identifiant le meilleur emplacement du réseau cœur local.

5.2 Métriques prises en compte

Dans le reste de ce chapitre, nous considérons en particulier les trois métriques suivantes en plus de celle du flux global (somme des flux initiaux des stations de base).

Satisfaction proportionnelle moyenne :

Cette métrique ne prend pas en compte les valeurs absolues du flux, mais la valeur relative moyenne proportionnelle à chaque demande de trafic d'une station de base. Cette métrique peut être facilement définie comme suit :

$$\frac{1}{n} \sum_{i \in [1, n]} \frac{f_{0,i}}{c_{0,i}} \quad (5.1)$$

Variance de la satisfaction proportionnelle :

Cette métrique calcule la variance de la satisfaction proportionnelle entre toutes les stations de base. Elle donne une idée de l'équité en termes de satisfaction proportionnelle. En effet, plus cette mesure est faible, plus les taux de satisfaction proportionnelle entre toutes les stations de base sont proches. Cette métrique est définie comme suit :

$$\left(\frac{1}{n} \sum_{i \in [1, n]} \left(\frac{f_{0,i}}{c_{0,i}} \right)^2 \right) - \left(\frac{1}{n} \sum_{i \in [1, n]} \frac{f_{0,i}}{c_{0,i}} \right)^2 \quad (5.2)$$

Satisfaction proportionnelle minimale :

Cette métrique représente la satisfaction proportionnelle minimale entre toutes les stations de base. Elle représente le cas le plus défavorable, ce qui est intéressant à analyser car des valeurs très faibles signifient que certaines stations de base auront une très mauvaise satisfaction proportionnelle même si le flux global est élevé. Cette mesure est définie comme suit :

$$\text{MIN}_{i \in [1, n]} \left\{ \frac{f_{0,i}}{c_{0,i}} \right\} \quad (5.3)$$

5.3 Modèles optimaux d'équité

Sur la base des trois métriques supplémentaires définies ci-dessus, nous adaptons notre modèle "one-shot" pour calculer les solutions optimales concernant chaque métrique, comme nous l'expliquons dans ce qui suit.

5.3.1 Maximiser la satisfaction proportionnelle moyenne

Ici, notre modèle est toujours le même que celui présenté dans la section 3.3.2, à l'exception de la substitution de la fonction objective par la satisfaction proportionnelle moyenne des stations de base. Nous notons que l'on peut omettre la division par n puisque la fonction d'optimisation est équivalente. Le modèle résultant est donc donné comme suit :

$$\text{Objectif : Maximiser } \sum_{i \in [1, n]} \frac{f_{0,i}}{c_{0,i}} \quad (5.4)$$

Sous contraintes :

$$(3.7)(3.8)(3.12)(3.13)(3.10)$$

Variables de décision :

$$y_i \in \{0, 1\}, i \in [1, n];$$

$$f_{ij} \in \mathbb{R}^+, i \in [0, n], j \in [1, n]$$

5.3.2 Minimiser la variance des satisfactions proportionnelles

Là encore, on peut simplement remplacer la fonction objective par la minimisation de la variance des satisfactions proportionnelles. Cependant, cela peut conduire à une solution triviale où tous les flux initiaux $f_{0,i}$ sont égaux à zéro. Nous ajoutons donc un autre terme égal à la somme de tous les flux multipliée par une très petite valeur, notée ϵ . Cela permet d'éliminer la solution triviale où tous les flux $f_{0,i}$ sont égaux à zéro. Nous notons que le modèle se transforme dans ce cas en un programme quadratique mixtes en nombres entiers (MIQP) à cause des termes quadratiques dans la fonction objective. Le modèle MIQP résultant est donné par :

$$\begin{aligned}
 \text{Objectif : Minimiser } & \frac{1}{n} \sum_{i \in [1, n]} \left(\frac{f_{0,i}}{c_{0,i}} \right)^2 \\
 & - \left(\frac{1}{n} \sum_{i \in [1, n]} \frac{f_{0,i}}{c_{0,i}} \right)^2 - \epsilon \sum_{i \in [1, n]} (f_{0,i})
 \end{aligned} \tag{5.5}$$

Sous contraintes :

$$(3.7)(3.8)(3.12)(3.13)(3.10)$$

Variables de décision :

$$\begin{aligned}
 y_i & \in \{0, 1\}, i \in [1, n]; \\
 f_{ij} & \in \mathbb{R}^+, i \in [0, n], j \in [1, n]
 \end{aligned}$$

5.3.3 Maximiser la satisfaction proportionnelle minimale

L'équité max-min est une des métriques d'équité les plus utilisées dans les réseaux conventionnels avec des sources et des destinations fixes et connues. Dans notre cas, notre modèle peut être adapté en utilisant la modélisation max-min classique tout en considérant la satisfaction proportionnelle. Nous définissons pour cela la variable auxiliaire \mathcal{J} qui est forcée d'être inférieure ou égale à tous les taux de satisfaction proportionnelle des stations de base (grâce à la contrainte 5.7) et que nous cherchons à maximiser dans la fonction objective. Le modèle résultant est alors donné par :

$$\text{Objectif : Maximiser } \mathcal{J} \tag{5.6}$$

Sous contraintes :

$$\begin{aligned}
 \forall i \in [1, n], \quad & \frac{f_{0,i}}{c_{0,i}} \geq \mathcal{J} \\
 & (3.7)(3.8)(3.12)(3.13)(3.10)
 \end{aligned} \tag{5.7}$$

Variables de décision :

$$\begin{aligned}
 y_i & \in \{0, 1\}, i \in [1, n]; \\
 f_{ij} & \in \mathbb{R}^+, i \in [0, n], j \in [1, n]
 \end{aligned}$$

Variables auxiliaires :

$$\mathcal{J} \in \mathbb{R}^+$$

5.4 Évaluation des modèles optimaux d'équité

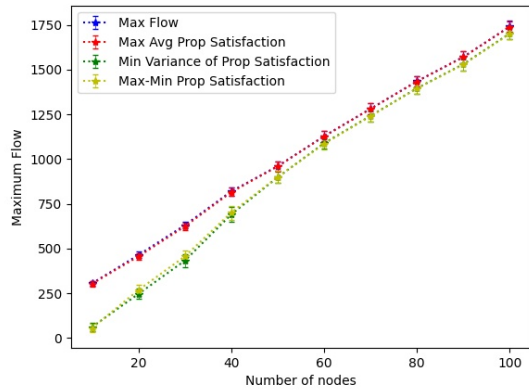
Nous avons implémenté les trois modèles additionnels ainsi que le modèle initial maximisant le flux global des stations de base, en suivant la même configuration de simulation présentée dans la section 4.1. Afin de comprendre le compromis entre la maximisation du flux global et l'équité entre les stations de base, nous comparons les quatre modèles en fonction des quatre métriques d'optimisation. Les résultats moyennés sur 100 itérations sont tracés dans la figure 5.1 avec des intervalles de confiance calculés à un niveau (taux) de confiance de 98%. Les résultats en termes de temps d'exécution que nous discuterons dans un second temps sont tracés dans la figure 5.2.

5.4.1 Évaluation des performances en termes de métriques de flux et d'équité

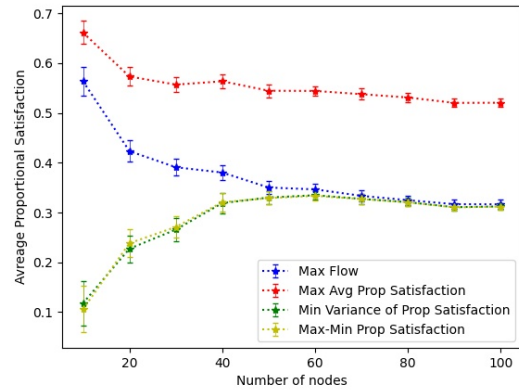
Nous traçons dans la figure 5.1 les quatre métriques tout en comparant les quatre modèles d'optimisation correspondants. Les résultats montrent tout d'abord que chaque modèle offre la meilleure performance en ce qui concerne sa métrique d'optimisation, ce qui est attendu. En outre, les résultats nous amènent à étudier les modèles en deux groupes, avec quelques comparaisons intéressantes intra- et inter- groupes. Le premier groupe comprend les modèles de flux maximal (tracé en bleu) et de satisfaction proportionnelle moyenne maximale (tracé en rouge), tandis que le deuxième groupe comprend les modèles de variance minimale (tracé en vert foncé) et les modèles de satisfaction proportionnelle max-min (tracé en vert clair).

Concernant le premier groupe, nous remarquons d'une part que les modèles de flux maximal et de satisfaction proportionnelle moyenne maximale donnent des résultats très similaires en termes de flux global, tout en conduisant tous les deux à des valeurs de satisfaction proportionnelle minimale très faibles. Cela signifie que certaines stations de base sont peu satisfaites, même si le flux global est optimal ou proche de l'optimal. En revanche, les deux modèles diffèrent en termes de la satisfaction proportionnelle moyenne et de la variance de la satisfaction proportionnelle. En effet, le premier modèle (flux maximal) est plus performant que le deuxième en termes de la variance de la satisfaction proportionnelle, atteignant un ratio allant jusqu'à 1.25 (voir Figure 5.1c). Cependant, le deuxième modèle (satisfaction proportionnelle moyenne maximale) offre une bien meilleure satisfaction proportionnelle moyenne, qui est multipliée par un ratio allant jusqu'à 1.65 (voir Figure 5.1b).

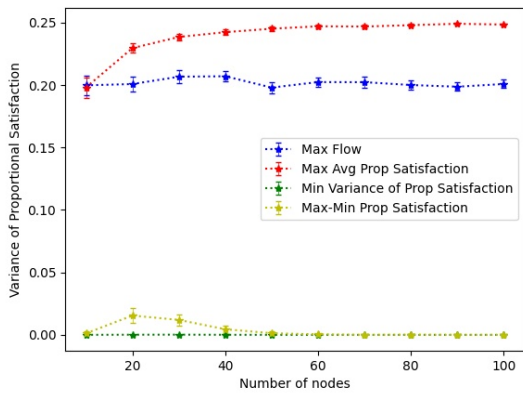
5.4 Évaluation des modèles optimaux d'équité



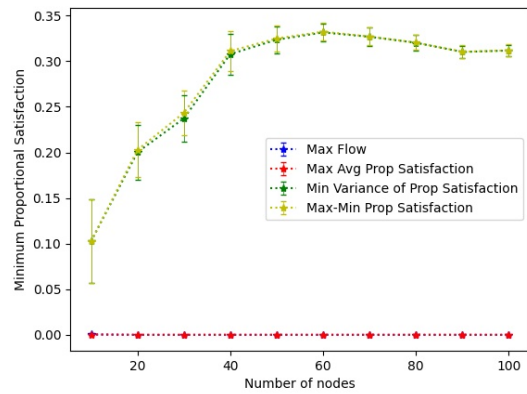
(a) Flux global vs taille du réseau (n)



(b) Satisfaction proportionnelle moyenne vs taille du réseau (n)



(c) Variance de la satisfaction proportionnelle vs taille du réseau (n)



(d) Satisfaction proportionnelle minimale vs taille du réseau (n)

FIGURE 5.1 – Comparaison des quatre modèles (Max Flow, Max Avg. Satisf., Min Norm. Var., Max-Min Satisf.) en fonction des quatre métriques correspondantes

Concernant le deuxième groupe, nous remarquons que le modèle minimisant la variance de la satisfaction proportionnelle et le modèle max-min maximisant la satisfaction proportionnelle minimale fournissent des résultats très similaires en termes des quatre métriques étudiées. Les deux modèles ne présentent qu'une légère différence en termes de variance de la satisfaction proportionnelle lorsque le nombre de nœuds est compris entre 20 et 40. La proximité des performances peut s'expliquer par le fait que le modèle max-min permet d'offrir un taux de satisfaction proportionnelle similaire, très proche du taux optimal, à toutes les stations ce qui se traduit par une variance très faible. Ces résultats montrent que les deux modèles sont souvent équivalents dans notre cas, mais nous recommandons le modèle max-min pour son temps d'exécution, comme détaillé dans la section suivante.

Enfin, la comparaison des deux groupes montre deux classes d'objectifs contradictoires : i) maximiser le flux global des stations de base ou la satisfaction proportionnelle moyenne des stations de base d'un côté et ii) maximiser l'équité en termes de satisfaction proportionnelle entre les stations de base de l'autre côté. Nous pensons donc qu'il est très important de concevoir la fonction d'optimisation en fonction des objectifs de l'application et, dans la plupart des cas, envisager une fonction d'optimisation multi-objectifs dans laquelle plusieurs métriques peuvent être combinées. Nous discuterons dans la section des perspectives quelques pistes que nous souhaitons explorer dans ce cadre.

En examinant les détails des résultats obtenus dans la figure 5.1, nous remarquons principalement que lorsque le nombre de nœuds est faible (inférieur à 50 dans notre cas), la différence est très significative en termes de flux global ainsi qu'en termes de satisfaction proportionnelle moyenne. Cependant, tous les résultats montrent que lorsque le nombre de nœuds est élevé (supérieur à 50 dans notre cas), les performances sont proches en termes de flux global alors que la différence reste importante en termes d'équité de la satisfaction proportionnelle. En effet, les résultats montrent par exemple que le modèle max-min offre un bon flux global lorsque le nombre de nœuds est élevé. Dans ce cas, nous pensons qu'il faut accorder plus d'importance à l'équité entre les taux de satisfaction proportionnelle lors de la conception des modèles d'optimisation multi-objectifs.

5.4.2 Évaluation du temps d'exécution

Outre l'évaluation des performances en termes de flux et de métriques d'équité, nous avons aussi évalué le temps d'exécution des quatre modèles et les résultats sont présentés dans la figure 5.2. Tous les résultats sont moyennés sur 100 itérations et nous traçons les intervalles de confiance calculés à un niveau de confiance de 98%. Les résultats numériques détaillés sont aussi présentés dans le tableau 5.2.

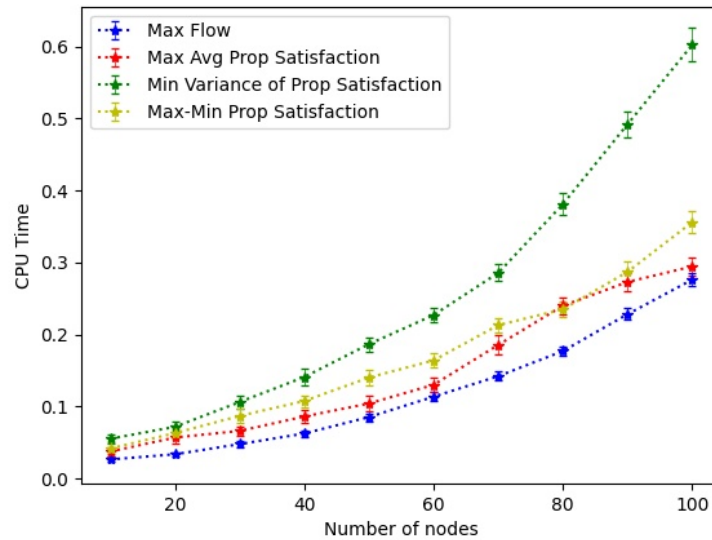


FIGURE 5.2 – Temps moyen d'exécution (en secondes) des quatre modèles en fonction de la taille du réseau (moyennes sur 100 itérations)

D'une part, les résultats montrent que les trois modèles (flux maximal, satisfaction proportionnelle moyenne maximale et max-min) offrent des temps d'exécution similaires. Nous remarquons cependant que le modèle max-min implique n contraintes supplémentaires, ce qui explique son temps d'exécution légèrement plus élevé. D'autre part, les résultats montrent que le modèle minimisant la variance proportionnelle prend plus de temps que les trois autres modèles. Ceci s'explique par la fonction d'optimisation quadratique utilisée dans ce cas. Une étude de scalabilité plus approfondie est proposée dans le chapitre suivant.

Nombre de nœuds	Max flow	Max Avg Prop. Satisf.	Min Variance	Max-Min Prop. Satisf.
10	0.026	0.037	0.055	0.041
20	0.033	0.057	0.072	0.063
30	0.048	0.066	0.106	0.087
40	0.062	0.086	0.141	0.107
50	0.085	0.104	0.186	0.140
60	0.113	0.130	0.227	0.164
70	0.142	0.185	0.286	0.213
80	0.177	0.240	0.381	0.235
90	0.228	0.273	0.492	0.287
100	0.276	0.294	0.602	0.356

TABLE 5.2 – Résultats de la comparaison du temps d'exécution des quatre modèles en fonction de la taille du réseau n (moyennes sur 100 itérations)

5.5 Conclusions

Dans ce chapitre, nous nous sommes focalisés sur la gestion de l'équité entre les stations de base. Après avoir expliqué l'intérêt d'explorer cette équité, nous avons proposé trois variantes de notre modèle : l'une permettant de maximiser la satisfaction proportionnelle moyenne, une autre visant à minimiser la variance des satisfactions proportionnelles, et une dernière permettant de maximiser la satisfaction proportionnelle minimale.

Les résultats de simulation montrent que les deux derniers modèles sont souvent équivalents dans notre cas. Ils confirment également que : i) maximiser le flux global ou la satisfaction proportionnelle moyenne des stations de base et ii) maximiser l'équité en termes de satisfaction proportionnelle entre les stations de base sont deux objectifs contradictoires. Nous avons remarqué que les degrés de divergence entre ces deux objectifs dépendent également du nombre de nœuds dans le réseau. Des pistes de modèles multi-objectifs à deux phases seront discutées comme perspectives de nos travaux de thèse.

Chapitre 6

Étude de la scalabilité des modèles de placement optimal

Contenu du chapitre

6.1 Scalabilité des modèles maximisant le flux selon le nombre de stations de base	80
6.2 Scalabilité des modèles maximisant le flux selon le nombre de réseaux cœur	82
6.3 Étude de la scalabilité en fonction des métriques d'optimisation . . .	86
6.4 Conclusions	88

Dans les deux chapitres précédents, le principal objectif était respectivement la validation/comparaison de notre modèle maximisant le flux global et ensuite l'étude de l'équité entre les stations de base avec un placement optimal du réseau cœur. Nous avons dans ces deux chapitres considéré des réseaux allant jusqu'à 100 stations de base tout en se focalisant sur la validation des modèles et l'étude des métriques de flux, de satisfaction proportionnelle et d'équité. Dans ce chapitre, nous reviendrons sur la scalabilité (passage à l'échelle) de nos modèles plus en détail. Nous abordons ce sujet de trois axes : i) le passage à l'échelle du modèle principal (maximisant le flux global) pour des grands réseaux allant jusqu'à des centaines de stations de base, ii) le passage à l'échelle en considérant plusieurs réseaux cœur à déployer et enfin iii) le passage à l'échelle en fonction de la fonction objective (critère d'optimisation).

6.1 Scalabilité des modèles maximisant le flux selon le nombre de stations de base

Afin d'évaluer la scalabilité de nos modèles de maximisation du flux global en fonction du nombre de nœuds, nous avons commencé par effectuer les mêmes comparaisons réalisées en chapitre 4 mais en variant le nombre de nœuds de 50 à 400 nœuds. Nous comparons ici les trois modèles : notre modèle avec un objectif de maximiser le flux global (somme des flux initiaux) que les stations de base peuvent envoyer, le modèle itératif à base de programme linéaire (LP) et le modèle itératif à base de coupe minimale (Min-Cut). Nous notons que nous considérons ici le cas hétérogène avec un $\rho = 0.2$.

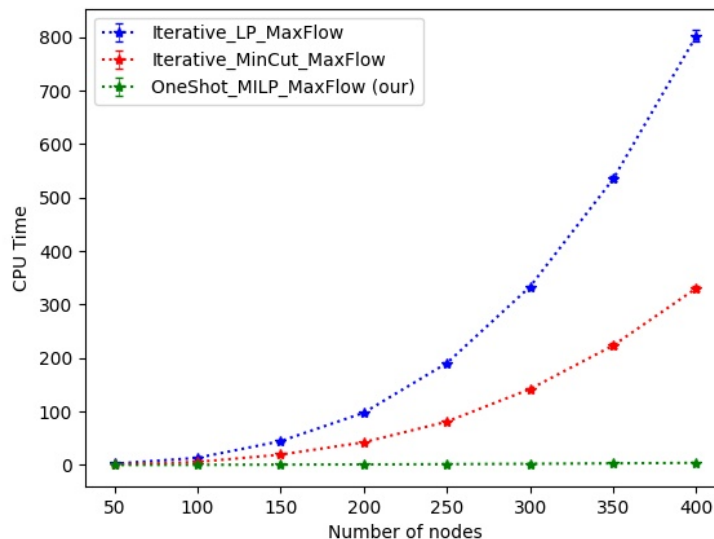


FIGURE 6.1 – Temps moyen d'exécution (en secondes) pour un nombre de nœuds, n , allant de 50 à 400 ($\rho = 0.2$ - moyennes sur 30 itérations)

Les résultats présentés dans la figure 6.1 sont moyennés sur 30 itérations et présentent des intervalles de confiance calculés à un niveau (taux) de confiance de 98%. Ils montrent les mêmes tendances que les résultats présentés dans le chapitre 4.

En effet, ils montrent que le modèle itératif à base de coupe minimale offre une faible complexité en fonction du nombre de nœuds par rapport au modèle itératif à base de programme linéaire. Les résultats montrent aussi que notre modèle "one-shot" offre une meilleure scalabilité par rapport aux deux autres modèles avec une plus faible complexité en fonction du nombre de nœuds. On observe par ailleurs, que le temps d'exécution de notre modèle reste très raisonnable même pour des très grands réseaux avec des centaines de stations de base. En effet, il ne dépasse pas les 5 secondes en

6.1 Scalabilité des modèles maximisant le flux selon le nombre de stations de base

Nombre de nœuds	Modèle itératif LP [10]	Modèle Min-Cut [10]	Notre modèle
	Temps exe. moyen (secondes)	Temps exe. moyen (secondes)	Temps exe. moyen (secondes)
50	2.348	0.910	0.084
100	13.330	5.765	0.283
150	44.583	19.316	0.642
200	97.722	42.396	1.021
250	190.578	81.052	1.557
300	333.440	142.059	2.234
350	534.888	223.520	3.073
400	802.432	330.580	4.002

TABLE 6.1 – Étude de la scalabilité en fonction de la taille du réseau : comparaison entre le modèle itératif basé sur LP, le modèle itératif basé sur Min-Cut et notre modèle (moyennes sur 30 itérations - cas hétérogène)

moyenne même avec un réseau de 400 nœuds alors que le modèle itératif à base de coupe minimale s'exécute en 330 secondes en moyenne et le modèle itératif à base de programme linéaire en 802 secondes en moyenne pour la même taille du réseau (Voir le tableau 6.1).

A la lumière de ces résultats, nous nous concentrons dans ce qui suit exclusivement sur notre modèle basé sur la modélisation "one-shot", que nous évaluons en fonction de ρ pour des grands réseaux, puis conjointement en fonction de n et ρ .

ρ	0.2	0.4	0.6	0.8	1
Temps d'exécution moyen (s)	6.81	12.17	16.89	19.78	21.12

TABLE 6.2 – Étude de scalabilité en fonction de ρ ($n = 500$ - moyennes sur 50 itérations)

Dans le tableau 6.2, nous présentons le temps d'exécution moyen sur 50 itérations en fonction de ρ pour un réseau constitué de 500 stations de base. Là aussi les résultats sont satisfaisants et montrent que le temps d'exécution moyen de notre modèle maximisant le flux global ne dépasse pas 22 secondes dans le pire des cas correspondant à un graphe complet. Les résultats confirment par ailleurs que le temps d'exécution augmente avec le taux de connectivité mais que plus le réseau est connecté plus l'augmentation est moins importante.

Nous nous sommes aussi intéressés à la scalabilité conjointe en fonctions de nombre de stations de base n et du rayon de connectivité ρ . Nous traçons dans la figure 6.2 le temps d'exécution en fonction de n et ρ . Ces résultats confirment la tendance observée dans le tableau 6.2 et permettent de bien visualiser l'évolution de temps d'exécution en fonction de n et ρ conjointement.

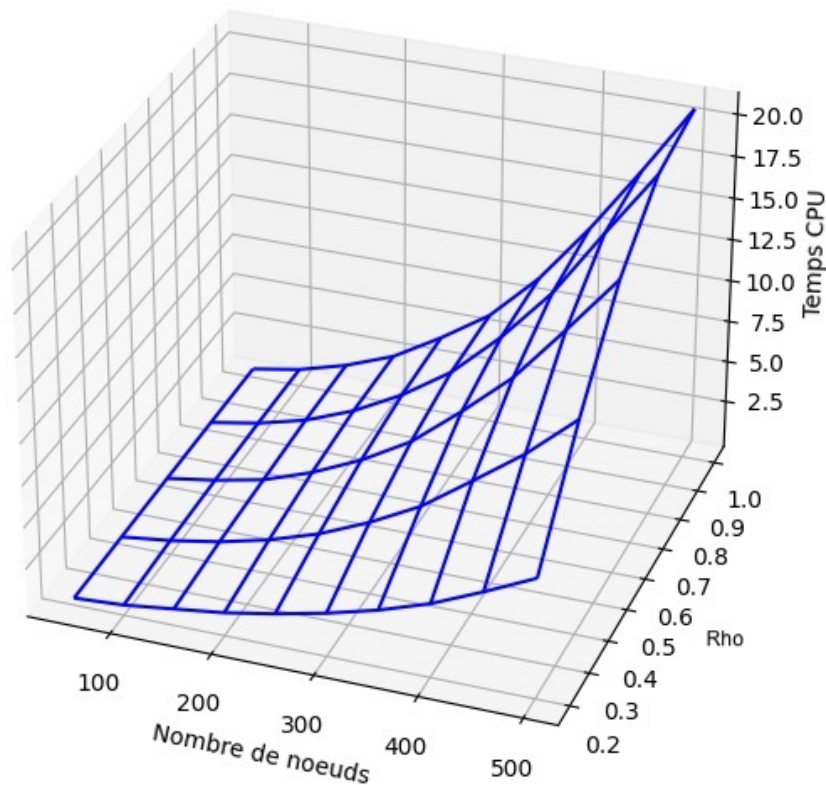


FIGURE 6.2 – Temps moyen d’exécution (en secondes) en fonction du nombre de nœuds n et du rayon de connectivité ρ

6.2 Scalabilité des modèles maximisant le flux selon le nombre de réseaux cœur

Dans cette section, nous allons étudier les performances de notre modèle en fonction du nombre des réseaux cœur à déployer. En effet, nous avons jusque là évalué exclusivement le cas où un seul réseau cœur doit être déployé mais comme abordé précédemment, certains travaux considèrent un réseau cœur hébergé dans chaque station de base [4, 5, 6]. Dans certaines situations, il pourrait être envisageable et opportun de prévoir un nombre réduit de réseaux cœur, avec des procédures de coordination adéquates, afin de partager la charge dans un réseau auto-déployable.

Afin de mener ces évaluations, nous considérons simplement notre modèle linéarisé présenté en section 3.3 en remplaçant l’équation 3.10 par l’équation 3.17. Il est à rappeler ici que nous considérons une inégalité et non pas une égalité car selon la fonction objective considérée, il est parfois mieux de placer moins de réseaux cœur et l’inégalité laisse au modèle la possibilité de choisir à la fois le nombre optimal de réseaux cœur, leurs emplacements et les flux optimaux.

6.2 Scalabilité des modèles maximisant le flux selon le nombre de réseaux cœur

CN	50	100	150	200	250	300	350	400	450	500
1	0.113	0.279	0.644	1.052	1.616	2.304	3.340	4.272	5.617	7.111
2	0.129	0.320	0.705	1.177	1.792	2.539	3.633	4.468	6.261	7.607
3	0.144	0.395	0.905	1.548	2.373	3.332	4.556	5.935	8.663	9.831
4	0.125	0.374	0.832	1.365	2.311	3.619	6.048	8.080	10.363	17.145
5	0.115	0.371	0.793	1.313	2.054	3.076	4.982	5.587	7.446	10.451

TABLE 6.3 – Étude de la scalabilité en fonction du nombre de réseaux cœur et du nombre de nœuds ($\rho = 0.2$ - moyennes sur 50 itérations)

Nous avons considéré dans nos simulations $\rho = 0.2$ et un objectif de maximisation du flux global (somme des flux initiaux des stations de base). Pour cette configuration, nous avons évalué, pour n allant de 50 à 500 et pour un nombre maximal de réseaux cœur (CN_{max}) allant de 1 à 5, à la fois le temps d'exécution et le flux maximal obtenu. Nous commençons par analyser le temps d'exécution qui est notre objectif principal dans ce chapitre. Les résultats de la simulation moyennés sur 50 itérations sont présentés dans le tableau 6.3 et tracés dans les deux figures 6.3 et 6.4.

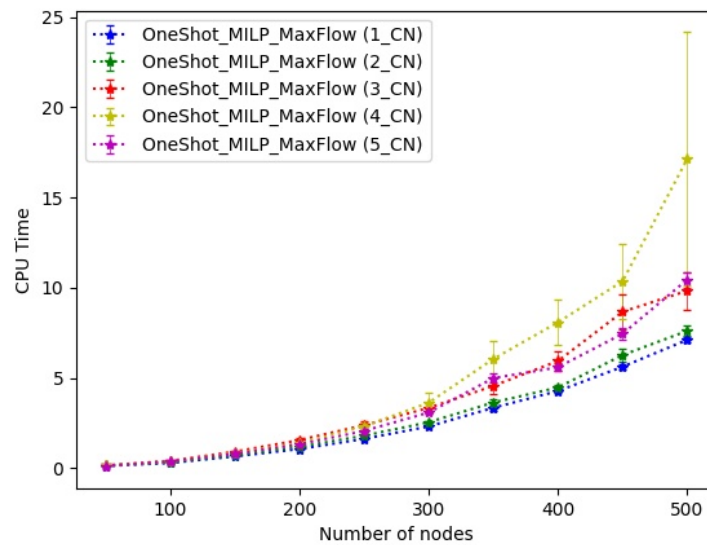


FIGURE 6.3 – Temps moyen d'exécution (secondes) en fonction du nombre de nœuds n pour différents nombres de réseaux cœur ($\rho = 0.2$ - moyennes sur 50 itérations)

A partir du tableau 6.3 et de la figure 6.3, nous notons d'abord que le temps d'exécution des modèles augmentent avec l'augmentation du nombre de nœuds et cela quelque soit le nombre de réseaux cœur. Cela était attendu car pour tout CN_{max} donné, l'augmentation du nombre de stations de base implique une augmentation dans le nombre de variables et dans le nombre de contraintes du modèle d'optimisation.

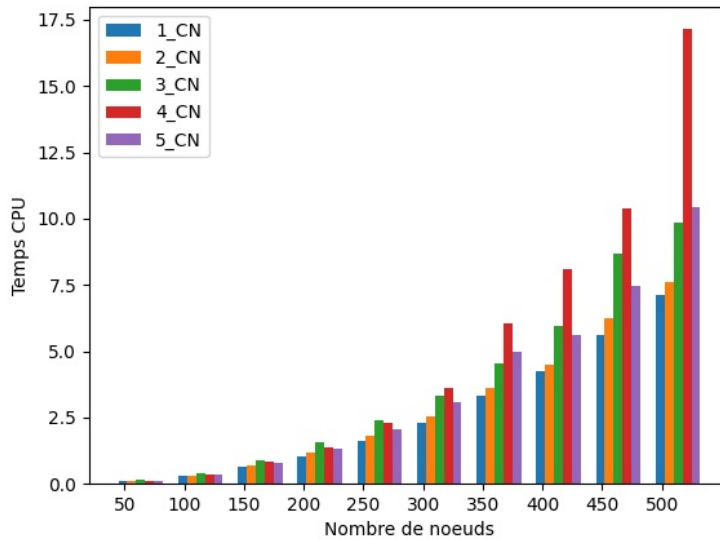


FIGURE 6.4 – Temps moyen d’exécution (en secondes) en fonction du nombre de nœuds n et du nombre de réseaux cœur ($\rho = 0.2$ - moyennes sur 50 itérations)

Néanmoins, les résultats numériques confirmés dans le graphique de la figure 6.3 montrent que pour un nombre de nœuds donné, le temps d’exécution commence à baisser à partir d’un certain seuil du nombre maximal de réseaux cœur. Afin de mieux visualiser ce constat, nous avons tracé dans la figure 6.4 les mêmes résultats groupés par nombre de nœuds. Ce graphique montre que pour un nombre de stations de base allant de 50 à 250, dès que le nombre maximal de réseaux cœur à déployer dépasse 3, le temps d’exécution commence légèrement à diminuer. Pour un nombre de stations de base allant de 300 à 500, ce seuil est de 4 ce qui veut dire le temps d’exécution commence légèrement à diminuer à partir de $CN_{max} = 4$. Notre intuition pour expliquer ce constat est le fait que l’augmentation du nombre CN_{max} sans contraintes supplémentaires fait que le flux maximal possible augmente vite ce qui fait que beaucoup de liens deviennent saturés dans les solutions réalisables tout en limitant l’espace de recherche possible.

Afin de confirmer cette intuition, nous nous sommes aussi intéressés à l’évolution de la solution optimal en termes de flux global maximal en fonction du nombre de nœuds et du nombre de réseaux cœur. Les résultats sont résumés dans le tableau 6.4 et présentés aussi dans la figure 6.5. Ils montrent que le flux maximal augmente avec l’augmentation du nombre de réseaux cœur possibles à déployer. Ceci était attendu, mais les résultats montrent que l’augmentation se resserre progressivement avec l’augmentation du nombre de réseaux cœur. Cela confirme la saturation de plusieurs liens

6.3 Scalabilité des modèles maximisant le flux selon le nombre de réseaux cœur

CN	50	100	150	200	250	300	350	400	450	500
1	989.3	1724.9	2399.6	3130.3	3798.7	4416.3	5136.9	5794.4	6431.4	7038.2
2	1861.6	3329.1	4667.9	6091.6	7416.8	8671.3	10042.0	11389.4	12644.6	13861.9
3	2551.4	4821.6	6839.3	8951.9	10915.4	12823.3	14841.1	16862.2	18739.4	20554.0
4	2730.8	5456.9	8213.7	10846.9	13603.9	16262.4	18957.7	21558.2	24275.6	26818.2
5	2731.4	5464.2	8230.5	10857.9	13625.2	16352.8	19067.3	21713.9	24579.9	27255.2

TABLE 6.4 – Flux maximal en fonction du nombre de nœuds et de réseaux cœur

de communications (en atteignant la capacité maximale) ce qui réduit les possibilités d'augmenter le flux maximal en ajoutant de nouveaux réseaux cœur.

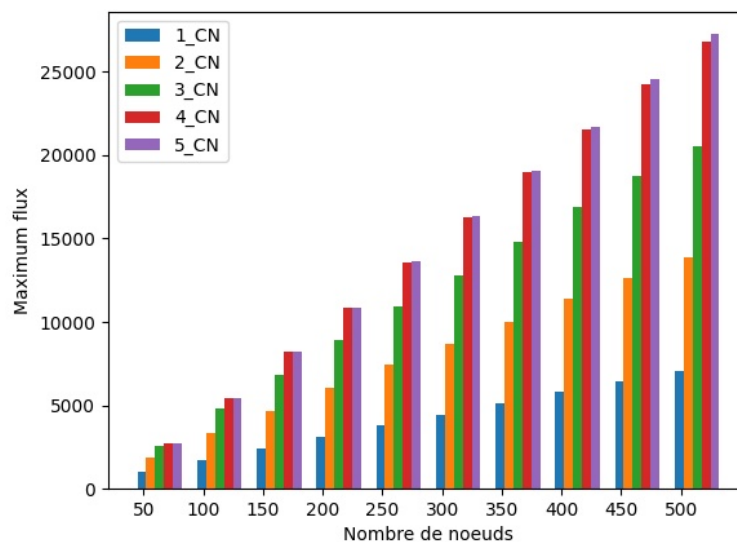


FIGURE 6.5 – Flux maximal en fonction du nombre de nœuds et de réseaux cœur

Nous notons que ces conclusions sont valables dans le cas d'une maximisation du flux global envoyé par les stations de base, le scénario considéré dans cette section. Dans le cas où d'autres fonctions objectives sont utilisées, il est possible que le temps d'exécution évolue différemment en fonction du nombre maximal de CN_{max} et d'autres évaluations sont nécessaires dans ce cas.

Il est enfin important de rappeler que le déploiement de plusieurs réseaux cœur nécessite forcément un coût supplémentaire et une coordination supplémentaire. L'objectif de notre étude vise principalement i) à montrer la faisabilité d'adapter notre modèle et de l'exécuter dans des temps très raisonnables même dans le cas où nous souhaitons déployer plusieurs réseaux cœur et ii) à étudier l'évolution du temps d'exécution en fonction du nombre maximal de réseaux cœur à déployer CN_{max} .

6.3 Étude de la scalabilité en fonction des métriques d'optimisation

Après avoir étudié l'impact de l'augmentation du nombre de nœuds et du nombre de réseaux cœur sur les performances de notre modèle, nous avons souhaité étudier l'impact des métriques d'optimisation choisies dans la fonction objective. En effet, dans le cadre de notre étude sur l'équité, nous avons examiné, dans le chapitre 5, les temps d'exécution de quatre modèles distincts, en considérant des réseaux allant jusqu'à 100 nœuds. Nous nous sommes concentrés sur l'interprétation des résultats en termes de métriques d'équité associées.

Dans ce chapitre, nous souhaitons reprendre ces quatre métriques à titre d'exemple afin d'observer dans quelle mesure le choix de la fonction objective impacte le temps d'exécution, notamment dans le cas de réseaux à très grande taille. Pour cela nous avons mené plusieurs simulations en considérant des réseaux allant de 50 à 500 nœuds avec un ρ égal à 0.2. Les résultats obtenus sont résumés dans le tableau 6.5 et tracés dans la figure 6.6.

Nombre de nœuds	50	100	150	200	250	300	350	400	450	500
Max Flow	0.09	0.27	0.68	1.07	1.62	2.41	3.27	4.31	5.59	6.88
Max Avg. Satisf.	0.11	0.30	0.68	1.10	1.72	2.51	3.45	4.39	5.89	7.24
Min Norm. Var.	0.19	0.64	1.97	4.60	10.11	21.34	37.31	58.96	92.37	139.70
Max-Min Satisf.	0.13	0.37	0.79	1.35	2.21	3.30	4.72	6.69	9.52	11.97

TABLE 6.5 – Temps d'exécution moyen des quatre modèles (Max Flow, Max Avg. Satisf., Min Norm. Var., Max-Min Satisf.) en fonction du nombre de nœuds

Les résultats obtenus montrent clairement que les modèles (Max Flow et Max Avg. Satisf.) maximisant respectivement le flux global et la satisfaction proportionnelle moyenne offrent des temps d'exécution très similaires et qui augmentent avec l'augmentation du nombre de nœuds dans le réseau. A titre d'exemple pour un doublement du nombre de nœuds de 250 à 500 le temps d'exécution du modèle de flux maximal est multiplié par 4.24 alors que le temps d'exécution du modèle de satisfaction proportionnelle moyenne maximale est multiplié par 4.21. Cela est attendu car les deux modèles présentent le même nombre de variables et de contraintes.

Les résultats montrent que le modèle de satisfaction proportionnelle max-min offre aussi une évolution raisonnable en fonction de nombre de nœuds même si le temps d'exécution est légèrement plus élevé que les deux autres modèles. Ici aussi on remarque que le temps d'exécution est multiplié par un facteur de 5.41 pour un doublement de nombre de nœuds de 250 à 500.

6.3 Étude de la scalabilité en fonction des métriques d'optimisation

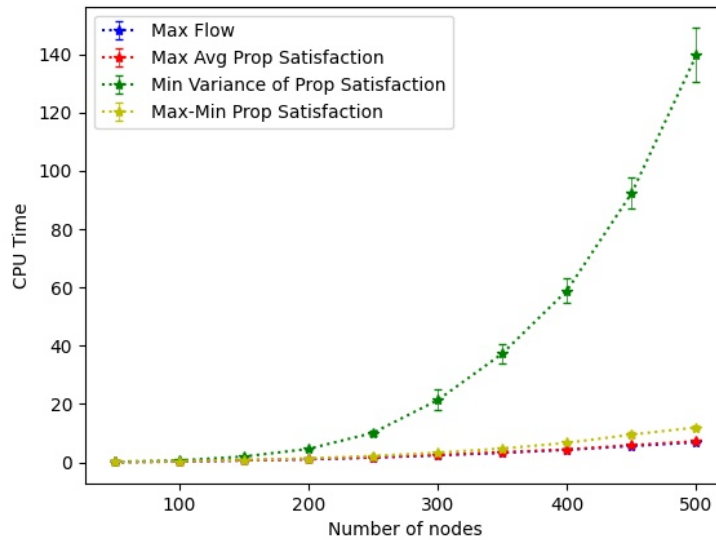


FIGURE 6.6 – Temps moyen d'exécution (en secondes) des quatre modèles en fonction de la taille du réseau (nombre de nœuds)

Par contre, les résultats montrent que le modèle de variance minimale offre quand à lui les performances les moins bonnes. En effet, ce modèle offre la plus mauvaise scalabilité en fonction du nombre de nœuds. Cela est confirmé visuellement dans la figure 6.6 et il est beaucoup plus prononcé que ce que nous avons observé dans la section 5.4.2 en considérant un nombre de nœuds allant jusqu'à 100. A titre d'exemple, le temps d'exécution est multiplié par 13.82 pour un doublement du nombre de nœuds de 250 à 500. Cela s'explique par la fonction d'optimisation quadratique utilisée dans ce cas.

Il est à noter donc que la considération de certaines fonctions objectives pourrait entraîner des temps d'exécution plus importants ce qui pourrait limiter la possibilité de ré-exécuter les modèles fréquemment en cas de changement de topologie. Comme nous le discuterons dans la section des conclusions et des perspectives, cela pourra justifier l'exploration d'heuristiques offrant une solution proche de l'optimal dans des temps d'exécution rapides dans le cas où certaines fonctions objectives (notamment quadratique) sont utilisées avec des réseaux de très grande taille. Pour le cas particulier de l'équité, nous rappelons néanmoins que le modèle de satisfaction proportionnelle max-min offre des solutions très similaires à celle du modèle de variance minimale tout en offrant des bien meilleurs temps d'exécution et une plus faible complexité en fonction de la taille du réseau. Il est donc recommandé de privilégier son utilisation dans notre cas.

6.4 Conclusions

Dans ce chapitre, nous avons étudié la scalabilité de nos modèles en fonction du nombre de nœuds, du nombre maximal de réseaux cœur à déployer et en considérant plusieurs critères d'optimisation. Les résultats montrent tout d'abord que le modèle "one-shot" maximisant le flux global des stations de base offre une très bonne scalabilité en fonction du nombre de nœuds et en fonction du nombre de réseaux cœur. Les résultats montrent que l'augmentation du nombre de réseaux cœur augmente le temps d'exécution dans un premier temps mais celui-ci commence à baisser légèrement à partir d'un certain seuil.

Les résultats montrent enfin que le critère d'optimisation impacte l'évolution du temps de calcul en fonction des autres paramètres. A titre d'exemple, l'utilisation d'une fonction objective quadratique augmente significativement la complexité de résolution en fonction du nombre de nœuds.

Conclusions et perspectives

Dans la suite de ce document, nous concluons en commençant par présenter un résumé du contexte de notre thèse, de nos contributions et des principaux résultats obtenus. Nous présentons ensuite quelques perspectives et travaux futurs en se focalisant sur quatre axes qui nous semblent pertinents à explorer dans le futur.

Résumé des contributions

Le domaine des réseaux cellulaires a connu des évolutions majeures ces dernières années. D'une part, la démocratisation des systèmes mobiles comme les robots et les drones offre la possibilité d'embarquer des stations de base ou même le réseau cœur sur ces unités mobiles. D'autre part, le domaine des réseaux cellulaires témoigne d'une prolifération croissante de la virtualisation des fonctions de réseau (NFV) et de la mise en réseau définie par logiciel (SDN) qui offrent plus de flexibilité dans le réseau cœur. Ces deux faits, associés aux capacités d'auto-organisation offertes par des normes récentes telles que la 5G, ont donné naissance au concept de réseau cellulaire auto-déployable, qui permet un déploiement plus souple et plus rapide des réseaux cellulaires. Ce concept est encore plus opportun compte tenu de l'émergence des réseaux privés et de la place qu'occupent les drones dans la future normalisation 6G.

Dans les réseaux auto-déployables, les stations de base sont déployées rapidement et facilement. Elles peuvent être connectées à un réseau traditionnel pour étendre la couverture, ou déployées de manière indépendante pour un usage autonome. Dans cette thèse, nous nous sommes concentrés sur ce dernier type de déploiement où il n'y a pas de liaison de retour vers un réseau traditionnel et où toutes les fonctionnalités du réseau cœur (CN) peuvent être regroupées et colocalisées avec une ou plusieurs stations de base du réseau.

Même si le concept de réseau cellulaire auto-déployable peut être applicable et utile dans différentes situations comme l'extension de la couverture et la gestion des événements à forte affluence, l'une des principales applications reste la réhabilitation de la communication dans les situations d'urgence. En effet, dans ces situations qui

suivent généralement une catastrophe naturelle ou d'origine humaine, les réseaux traditionnels sont partiellement ou totalement inopérants. Les réseaux auto-déployables peuvent offrir ici une solution clé afin de restaurer facilement et rapidement les communications entre les humains et les premiers intervenants.

Un point essentiel apporté par les réseaux auto-déployables est le continuum entre les réseaux d'accès radio et le réseau cœur. En effet, les fonctions virtualisées du réseau cœur peuvent être directement hébergées dans une station de base formant le réseau cœur local. Ces fonctions peuvent être hébergées dans chaque station de base ou dans une station de base qui héberge un réseau cœur local partagé. Dans ce travail, nous avons principalement considéré ce deuxième cas, où toutes les fonctionnalités du réseau cœur sont regroupées dans une unité (CN local) colocalisée avec une station de base du réseau. Néanmoins, nous avons aussi considéré vers la fin de notre thèse le cas d'un réseau auto-déployable avec plusieurs réseaux cœur.

Malgré les avantages offerts par les réseaux auto-déployables, ceux-ci soulèvent également de nouveaux défis scientifiques. L'un de ces multiples défis concerne le placement optimal des réseaux cœur locaux partagés. En effet, le choix des stations de base hébergeant le(s) réseau(x) cœur partagé(s) doit être effectué dans un délai raisonnable afin de garantir la réactivité nécessaire. De plus, ce choix doit être fait de manière rigoureuse, tout en considérant différents objectifs tels que la maximisation du flux global de trafic, la gestion de l'équité entre les stations de base, etc.

Cela a constitué le cœur de notre thèse en examinant le problème du placement efficace des réseaux cœur locaux. Nous proposons pour cela un nouveau modèle générique qui permet de localiser simultanément le réseau cœur local tout en optimisant une mesure de performance donnée. Nous avons commencé par modéliser conjointement le problème sous la forme d'un programme non linéaire mixte en nombres entiers. Notre modèle permet de trouver à la fois le meilleur emplacement du réseau cœur local et de maximiser le flux que les stations de base peuvent envoyer vers le réseau cœur à travers le réseau. Nous avons linéarisé les contraintes non linéaires de notre modèle initial et discuté des atouts de notre formulation finale en termes d'efficacité du temps d'exécution, de capacité à intégrer des demandes et des flux initiaux hétérogènes, ainsi que de sa généralité.

Nous avons ensuite mené des simulations extensives afin de comparer notre solution à d'autres solutions issues de la littérature. Nous avons en particulier considéré un modèle itératif à base de coupe minimale et un modèle itératif à base de programmation linéaire. Les résultats sont présentés et discutés dans le contexte des flux initiaux homogènes, ainsi que dans un contexte hétérogène plus général. Ils montrent que notre approche conjointe en une itération réduit considérablement le temps d'exé-

cution par rapport aux solutions existantes, tout en fournissant les mêmes résultats optimaux dans tous les scénarios étudiés. Les résultats montrent également que notre approche s'adapte bien à la taille du réseau et au rayon de connectivité.

Après avoir évalué les performances de nos nouveaux modèles, nous nous sommes attaqués à un verrou très important concernant l'équité entre les stations de base. Nous avons pour cela proposé des variantes de notre modèle pour prendre en compte l'équité en termes de taux de satisfaction proportionnelle. Nous avons montré que notre modèle peut facilement s'adapter à de nouvelles métriques d'optimisation tout en prenant en compte trois métriques supplémentaires. Nous avons également mis en oeuvre et évalué les quatre modèles résultants en termes de flux global, de satisfaction proportionnelle moyenne, de variance de la satisfaction proportionnelle et de taux minimum de satisfaction proportionnelle. Nous avons montré que la minimisation de la variance de la satisfaction proportionnelle est équivalente et similaire à l'optimisation de l'équité max-min en termes de métriques d'équité et de flux tout en offrant un temps de calcul plus faible. Nous avons également analysé et discuté les résultats, qui confirment les deux objectifs contradictoires : maximisation du flux global ou de la satisfaction proportionnelle moyenne, versus équité entre les taux de satisfaction proportionnelle. Nous avons montré que ces objectifs contradictoires sont encore plus prononcés lorsque le nombre de stations de base est faible.

Nous avons enfin étudié la scalabilité de nos modèles à la fois en termes de nombre de nœuds, en termes de nombre de réseaux cœur à déployer et en considérant plusieurs critères d'optimisation. Les résultats ont montré tout d'abord que notre modèle maximisant le flux global des stations de base offre une très bonne scalabilité en fonction du nombre de nœuds et en fonction du nombre de réseaux cœur. Les résultats ont aussi montré que l'augmentation du nombre de réseaux cœur augmentent le temps d'exécution dans un premier temps mais celui-ci commence à baisser légèrement à partir d'un certain seuil. Les résultats ont enfin montré que le critère d'optimisation impacte l'évolution du temps de calcul en fonction des autres paramètres. A titre d'exemple, l'utilisation d'une fonction objective quadratique augmente significativement la complexité de résolution en fonction du nombre de nœuds.

Perspectives et travaux futurs

Dans le cadre de travaux futurs, nous prévoyons d'évaluer des fonctions d'optimisation multi-objectifs qui prennent en compte à la fois le flux global des stations de base et les métriques d'équité. Nos solutions devraient prendre en compte explicitement ou implicitement le nombre de stations de base, puisque celui-ci est impor-

tant, comme nous l'avons montré dans cette thèse. Différents axes seront explorés, allant de la simple fonction objective combinée à des solutions plus sophistiquées. Entre autres, nous visons à proposer et à évaluer un modèle en deux étapes qui optimise une métrique donnée sans diminuer excessivement les autres métriques de performance. Grâce à l'efficacité de notre formulation, nous pouvons par exemple exécuter le modèle optimal pour déterminer le flux global maximal possible et exécuter un second modèle qui maximise le taux minimum de satisfaction proportionnelle tout en contraignant le flux de trafic à ne pas tomber en dessous d'un certain pourcentage du flux maximal possible.

Outre l'optimisation multi-objectif, nous souhaitons prendre en compte davantage de paramètres dans nos modèles, comme la latence par exemple. En effet, ce paramètre est très important dans les applications sensibles et un bon modèle doit trouver le bon emplacement du réseau cœur qui maximise le flux et/ou l'équité tout en minimisant le délai moyen ou le délai maximum (pire des cas) pour acheminer le trafic des stations de base au réseau cœur. Là aussi différents axes seront explorés, allant de la simple fonction objective combinée à des solutions plus sophistiquées. On peut de même penser à évaluer un modèle en deux étapes qui optimise le flux maximum sans présenter des délais de transmissions excessivement élevés. On peut de l'autre côté proposer des modèles qui minimisent une agrégation des latences (moyenne, maximum, etc.) sans trop dégrader le flux par rapport au flux maximal possible.

Dans le cas où plusieurs réseaux cœur peuvent être déployés, une autre perspective intéressante sera d'étudier l'équilibrage de charge entre les différents réseaux cœur. Nous pouvons dans un premier temps analyser les différences entre la charge des différents réseaux cœur. Nous pouvons ensuite proposer des modèles qui permettent d'équilibrer la charge entre ces réseaux cœur. Cette question est importante afin d'éviter un déséquilibre dans l'utilisation de ressources. Par contre, cette métrique peut s'avérer contradictoire avec la maximisation du flux et ou avec l'équité en termes de satisfaction proportionnelle des différentes stations de base. Le compromis entre les deux est donc très important à étudier et à analyser.

Un dernier axe que nous souhaitons aborder est de proposer des solutions basées sur des heuristiques et de les comparer à nos solutions en termes de temps d'exécution et de qualité des résultats. Ces heuristiques peuvent concerner les trois axes mentionnés précédemment (modèle combinant flux maximal et équité, les modèles prenant en compte la latence et les modèles visant un équilibrage entre les réseaux cœur). Selon les résultats obtenus sur ces trois axes, la priorité sera de concevoir des heuristiques pour les problèmes où la résolution exacte offre une faible scalabilité.

Bibliographie

- [1] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks : Applications, challenges, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.
- [2] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, and S. Allsopp, "Designing and implementing future aerial communication networks," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 26–34, 2016.
- [3] K. Sundaresan, E. Chai, A. Chakraborty, and S. Rangarajan, "Skylite : End-to-end design of low-altitude uav networks for providing lte connectivity," *ArXiv*, vol. abs/1802.06042, 2018.
- [4] M. Moradi, K. Sundaresan, E. Chai, S. Rangarajan, and Z. M. Mao, "Skycore : Moving core to the edge for untethered and reliable uav-based lte networks," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, (ACM MobiCom)*, p. 35–49, 2018.
- [5] M. Moradi, K. Sundaresan, E. Chai, S. Rangarajan, and Z. M. Mao, "Skycore : Moving core to the edge for untethered and reliable uav-based lte networks," *GetMobile : Mobile Computing and Communications*, vol. 23, no. 1, p. 24–29, 2019.
- [6] M. Moradi, K. Sundaresan, E. Chai, S. Rangarajan, and Z. M. Mao, "Skycore : Moving core to the edge for untethered and reliable uav-based lte networks," *Communications of the ACM*, vol. 64, no. 1, p. 116–124, 2021.
- [7] V.-G. Nguyen, A. Brunstrom, K.-J. Grinnemo, and J. Taheri, "Sdn/nfv-based mobile packet core network architectures : A survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1567–1602, 2017.
- [8] M. Bagaa, T. Taleb, A. Laghrissi, A. Ksentini, and H. Flinck, "Coalitional game for the creation of efficient virtual core network slices in 5g mobile systems," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 469–484, 2018.
- [9] J. Oueis, *Radio access and core functionalities in self-deployable mobile networks*. Ph.d. thesis, Université de Lyon, Nov. 2018.

-
- [10] J. Oueis, V. Conan, D. Lavaux, H. Rivano, R. Stanica, and F. Valois, "Core network function placement in self-deployable mobile networks," *Computer Communications*, vol. 133, pp. 12–23, 2019.
- [11] H. Fourati, R. Maaloul, L. Chaari, and M. Jmaiel, "Comprehensive survey on self-organizing cellular network approaches applied to 5g networks," *Computer Networks*, vol. 199, p. 108435, 2021.
- [12] W. Sarver, "Private cellular networks, are they finally ready to deliver?," *ComSoc Technology News (CTN)*, 2022.
- [13] D. Mishra, A. M. Vegni, V. Loscrí, and E. Natalizio, "Drone networking in the 6g era : A technology overview," *IEEE Communications Standards Magazine*, vol. 5, no. 4, pp. 88–95, 2021.
- [14] M. Y. Selim and A. E. Kamal, "Post-disaster 4g/5g network rehabilitation using drones : Solving battery and backhaul issues," in *2018 IEEE Global Communications Conference Workshops (IEEE Globecom Workshops)*, pp. 1–6, 2018.
- [15] K. Gomez, L. Goratti, T. Rasheed, and L. Reynaud, "Enabling disaster-resilient 4g mobile communication networks," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 66–73, 2014.
- [16] Z. Zhang, R. Stanica, and F. Valois, "Delay-based core network placement in self-deployable mobile networks," in *2021 IEEE Wireless Communications and Networking Conference (IEEE WCNC)*, pp. 1–6, 2021.
- [17] T. Rappaport, *Wireless Communications : Principles and Practice*. USA : Prentice Hall PTR, 2nd ed., 2001.
- [18] X. Lagrange, P. Godlewski, and S. Tabbane, *Réseaux GSM : des principes à la norme*. Réseaux et télécommunications, Hermès Science Publications, 2000.
- [19] X. Sanchez and H. Thioune, *UMTS*. Réseaux et télécommunications, Hermès Science Publications, 2000.
- [20] Y. Bouguen, E. Hardouin, and F.-X. Wolff, *LTE et les réseaux 4G*. Editions Eyrolles, 2012.
- [21] V. Sachan, *Fundamentals of 5G Wireless Communications*. Sachan Series, Independently Published, 2020.
- [22] M. Chafii, L. Bariah, S. Muhaidat, and M. Debbah, "Twelve scientific challenges for 6G : Rethinking the foundations of communications theory," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 868–904, 2023.
- [23] P. Marsch, O. Bulakci, O. Queseth, and M. Boldi, *5G System Design : Architectural and Functional Considerations and Long Term Research*. Wiley Publishing, 1st ed., 2018.
-

-
- [24] E. Dahlman, S. Parkvall, and J. Skold, *5G NR : The Next Generation Wireless Access Technology*. USA : Academic Press, Inc., 1st ed., 2018.
- [25] M. Manini, *Allocation de ressources et ordonnancement dans les réseaux de 5ème génération*. PhD thesis, Rennes 1, 2021.
- [26] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems : Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.
- [27] D. B. da Costa, Q. Zhao, M. Chafii, F. Bader, and M. Debbah, "6g : Vision, applications, and challenges," *Fundamentals of 6G Communications and Networking*, pp. 15–69, 2023.
- [28] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski, "Cellular, wide-area, and non-terrestrial iot : A survey on 5g advances and the road toward 6g," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1117–1174, 2022.
- [29] R. Pujol, *Association des utilisateurs dans les réseaux mobiles flexibles et agiles*. PhD thesis, Université de Lyon, 2022.
- [30] J. Gielis, A. Shankar, and A. Prorok, "A critical review of communications in multi-robot systems," *Current robotics reports*, vol. 3, no. 4, pp. 213–225, 2022.
- [31] N. S. Labib, M. R. Brust, G. Danoy, and P. Bouvry, "The rise of drones in internet of things : A survey on the evolution, prospects and challenges of unmanned aerial vehicles," *IEEE Access*, vol. 9, pp. 115466–115487, 2021.
- [32] F. Ahmed, J. Mohanta, A. Keshari, and P. S. Yadav, "Recent advances in unmanned aerial vehicles : a review," *Arabian Journal for Science and Engineering*, vol. 47, no. 7, pp. 7963–7984, 2022.
- [33] C. Guerber, *Sécurisation des communications dans un réseau ad hoc au sein d'un essaim de drones*. PhD thesis, INSA de Toulouse, 2022.
- [34] H. Shakhathreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (uavs) : A survey on civil applications and key research challenges," *Ieee Access*, vol. 7, pp. 48572–48634, 2019.
- [35] R. Shakeri, M. A. Al-Garadi, A. Badawy, A. Mohamed, T. Khattab, A. K. Al-Ali, K. A. Harras, and M. Guizani, "Design challenges of multi-uav systems in cyber-physical applications : A comprehensive survey and future directions," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3340–3385, 2019.
-

-
- [36] N. Elmeseiry, N. Alshaer, and T. Ismail, "A detailed survey and future directions of unmanned aerial vehicles (uavs) with potential applications," *Aerospace*, vol. 8, no. 12, p. 363, 2021.
- [37] X. Hu and R. H. Assaad, "The use of unmanned ground vehicles (mobile robots) and unmanned aerial vehicles (drones) in the civil infrastructure asset management sector : Applications, robotic platforms, sensors, and algorithms," *Expert Systems with Applications*, vol. 232, p. 120897, 2023.
- [38] M. Erdelj, O. Saif, E. Natalizio, and I. Fantoni, "Uavs that fly forever : Uninterrupted structural inspection through automatic uav replacement," *Ad Hoc Networks*, vol. 94, p. 101612, 2019.
- [39] R. Hewett and S. Puangpontip, "On controlling drones for disaster relief," *Procedia Computer Science*, vol. 207, pp. 3703–3712, 2022.
- [40] S. Hayat, E. Yanmaz, T. X. Brown, and C. Bettstetter, "Multi-objective UAV path planning for search and rescue," in *IEEE ICRA*, pp. 5569–5574, 2017.
- [41] T. Villa, L. Gonzalez, B. Miljevic, Z. Ristovski, and L. Morawska, "An overview of small unmanned aerial vehicles for air quality measurements : Present applications and future perspectives," *Sensors*, vol. 16, p. 1072, 07 2016.
- [42] J. Saunders, S. Saeedi, and W. Li, "Autonomous aerial robotics for package delivery : A technical review," *Journal of Field Robotics*, vol. 41, p. 3–49, July 2023.
- [43] G. U. Sai Theja, M. S. Murari, M. F. Singha, R. Patgiri, and A. Choudhury, "A survey on surveillance using drones," in *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, p. 250–257, Association for Computing Machinery, 2022.
- [44] J. Kim, S. Kim, C. Ju, and H. I. Son, "Unmanned aerial vehicles in agriculture : A review of perspective of platform, control, and applications," *Ieee Access*, vol. 7, pp. 105100–105115, 2019.
- [45] S. M. Adams and C. J. Friedland, "A survey of unmanned aerial vehicle (uav) usage for imagery collection in disaster research and management," in *9th international workshop on remote sensing for disaster response*, vol. 8, pp. 1–8, 2011.
- [46] C. Caillouet and N. Mitton, "Optimization and communication in uav networks," *Sensors*, vol. 20, no. 18, 2020.
- [47] S. Hayat, E. Yanmaz, and R. Muzaffar, "Survey on uav networks for civil applications : A communications viewpoint," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2624–2661, 2016.
- [48] Z. Z. Sharma, P. Vanjani, N. Paliwal, C. M. W. Basnayaka, D. N. K. Jayakody, H.-C. Wang, and P. Muthuchidambaranathan, "Communication and networking

-
- technologies for uavs : A survey," *Journal of Network and Computer Applications*, vol. 168, p. 102739, 2020.
- [49] D. Mishra and E. Natalizio, "A survey on cellular-connected uavs : Design challenges, enabling 5g/b5g innovations, and experimental advancements," *Computer Networks*, vol. 182, p. 107451, 2020.
- [50] R. Shahzadi, M. Ali, H. Z. Khan, and M. Naeem, "Uav assisted 5g and beyond wireless networks : A survey," *Journal of network and computer applications*, vol. 189, p. 103114, 2021.
- [51] Z. Zhao, P. Cumino, C. Esposito, M. Xiao, D. Rosário, T. Braun, E. Cerqueira, and S. Sargento, "Smart unmanned aerial vehicles as base stations placement to improve the mobile network operations," *Computer communications*, vol. 181, pp. 45–57, 2022.
- [52] H. Huang, A. V. Savkin, M. Ding, and M. A. Kaafar, "Optimized deployment of drone base station to improve user experience in cellular networks," *Journal of Network and Computer Applications*, vol. 144, pp. 49–58, 2019.
- [53] X. Lin, V. Yajnanarayana, S. D. Muruganathan, S. Gao, H. Asplund, H.-L. Maattanen, M. Bergstrom, S. Euler, and Y.-P. E. Wang, "The sky is not the limit : Lte for unmanned aerial vehicles," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 204–210, 2018.
- [54] Y. Li and M. Chen, "Software-defined network function virtualization : A survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.
- [55] J. G. Herrera and J. F. Botero, "Resource allocation in nfv : A comprehensive survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, 2016.
- [56] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization : State-of-the-art and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 236–262, 2016.
- [57] B. Umrao and D. Yadav, "Algorithms for functionalities of virtual network : a survey," *The Journal of Supercomputing*, vol. 77, 07 2021.
- [58] R. Masoudi and A. Ghaffari, "Software defined networks : A survey," *Journal of Network and computer Applications*, vol. 67, pp. 1–25, 2016.
- [59] A. K. Rangiseti and B. R. Tamma, "Software defined wireless networks : A survey of issues and solutions," *Wireless Personal Communications*, vol. 97, pp. 6019–6053, 2017.
-

-
- [60] Y. Zhang, L. Cui, W. Wang, and Y. Zhang, "A survey on software defined networking with multiple controllers," *Journal of Network and Computer Applications*, vol. 103, pp. 101–118, 2018.
- [61] J. Zheng, C. Tian, H. Dai, Q. Ma, W. Zhang, G. Chen, and G. Zhang, "Optimizing nfv chain deployment in software-defined cellular core," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 2, pp. 248–262, 2019.
- [62] Z. A. Qazi, M. Walls, A. Panda, V. Sekar, S. Ratnasamy, and S. Shenker, "A high performance packet core for next generation cellular networks," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, pp. 348–361, 2017.
- [63] Z. Zaidi, V. Friderikos, Z. Yousaf, S. Fletcher, M. Dohler, and H. Aghvami, "Will sdn be part of 5g?," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3220–3258, 2018.
- [64] R. Silva, D. Santos, F. Meneses, D. Corujo, and R. L. Aguiar, "A hybrid sdn solution for mobile networks," *Computer Networks*, vol. 190, p. 107958, 2021.
- [65] L. Ma, X. Wen, L. Wang, Z. Lu, and R. Knopp, "An sdn/nfv based framework for management and deployment of service based 5g core network," *China Communications*, vol. 15, no. 10, pp. 86–98, 2018.
- [66] C. H. T. Arteaga, F. B. Anaconda, K. T. T. Ortega, and O. M. C. Rendon, "A scaling mechanism for an evolved packet core based on network functions virtualization," *IEEE Transactions on Network and Service Management*, vol. 17, no. 2, pp. 779–792, 2019.
- [67] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "Nfv and sdn—key technology enablers for 5g networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2468–2478, 2017.
- [68] B. Gerő, D. Jocha, R. Szabó, J. Czentye, D. Haja, B. Németh, B. Sonkoly, M. Szalay, L. Toka, C. J. B. Cano, *et al.*, "The orchestration in 5g exchange—a multi-provider nfv framework for 5g services," in *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pp. 1–2, 2017.
- [69] I. Leyva-Pupo and C. Cervelló-Pastor, "An intelligent scheduling for 5g user plane function placement and chaining reconfiguration," *Computer Networks*, p. 110037, 2023.
- [70] O. Sami Oubbati, M. Atiquzzaman, T. Ahamed Ahanger, and A. Ibrahim, "Softwarization of uav networks : A survey of applications and future trends," *IEEE Access*, vol. 8, pp. 98073–98125, 2020.
-

-
- [71] E. Bozkaya and B. Canberk, "Sdn-enabled deployment and path planning of aerial base stations," *Computer Networks*, vol. 171, p. 107125, 2020.
- [72] J. Oueis, R. Stanica, and F. Valois, "Virtualized local core network functions placement in mobile networks," in *2019 IEEE Wireless Communications and Networking Conference (IEEE WCNC)*, pp. 1–6, 2019.
- [73] R. Pujol, R. Stanica, and F. Valois, "Lte user association for self-deployable networks in disaster management scenarios," in *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pp. 1–4, 2019.
- [74] W. Kiess and A. Khan, "Centralized vs. distributed : On the placement of gateway functionality in 5g cellular networks," in *2014 IEEE Global Communications Conference (IEEE GLOBECOM)*, pp. 4788–4793, 2014.
- [75] S. Muthaiah and C. Rosenberg, "Single gateway placement in wireless mesh networks," in *Proc. of 8th International IEEE Symposium on Computer Networks (ISCN'08)*, 2008.
- [76] B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, "Gateway placement optimization in wireless mesh networks with qos constraints," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2127–2136, 2006.
- [77] D. Das, Z. Rehena, S. Roy, and N. Mukherjee, "Multiple-sink placement strategies in wireless sensor networks," in *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)*, pp. 1–7, 2013.
- [78] J. Wong, R. Jafari, and M. Potkonjak, "Gateway placement for latency and energy efficient data aggregation [wireless sensor networks]," in *29th Annual IEEE International Conference on Local Computer Networks (IEEE LCN)*, pp. 490–497, 2004.
- [79] A. Karnik, A. Iyer, and C. Rosenberg, "Throughput-optimal configuration of fixed wireless networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1161–1174, 2008.
- [80] J. Sun, Y. Zhang, F. Liu, H. Wang, X. Xu, and Y. Li, "A survey on the placement of virtual network functions," *Journal of Network and Computer Applications*, vol. 202, p. 103361, 2022.
- [81] W. Attaoui, E. Sabir, H. Elbiaze, and M. Guizani, "Vnf and cnf placement in 5g : Recent advances and future trends," *IEEE Transactions on Network and Service Management*, 2023.
- [82] C. R. De Mendoza, B. Bakhshi, E. Zeydan, and J. Manges-Bafalluy, "Near optimal vnf placement in edge-enabled 6g networks," in *2022 25th Conference on Innovation in Clouds, Internet and Networks (ICIN)*, pp. 136–140, IEEE, 2022.
-

-
- [83] J. Cao, Y. Zhang, W. An, X. Chen, J. Sun, and Y. Han, "Vnf-fg design and vnf placement for 5g mobile networks," *Science China Information Sciences*, vol. 60, pp. 1–15, 2017.
- [84] A. Baumgartner, V. S. Reddy, and T. Bauschert, "Mobile core network virtualization : A model for combined virtual core network function placement and topology optimization," in *Proceedings of the 2015 1st IEEE conference on Network Softwarization (NetSoft)*, pp. 1–9, 2015.
- [85] M. Golkarifard, C. F. Chiasserini, F. Malandrino, and A. Movaghar, "Dynamic vnf placement, resource allocation and traffic routing in 5g," *Computer Networks*, vol. 188, p. 107830, 2021.
- [86] S. Retal and A. Idrissi, "Virtual network functions placement system for 5g mobile network architecture," *International Journal of Internet Technology and Secured Transactions*, vol. 10, no. 1-2, pp. 4–23, 2020.
- [87] I. Leyva-Pupo and C. Cervelló-Pastor, "Efficient solutions to the placement and chaining problem of user plane functions in 5g networks," *Journal of Network and Computer Applications*, vol. 197, p. 103269, 2022.
- [88] A. Mouaci, É. Gourdin, I. Ljubić, and N. Perrot, "Virtual network functions placement and routing problem : Path formulation," in *2020 IFIP Networking Conference (Networking)*, pp. 55–63, IEEE, 2020.
- [89] J. Crichigno, D. Oliveira, M. Pourvali, N. Ghani, and D. Torres, "A routing and placement scheme for network function virtualization," in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 26–31, IEEE, 2017.
- [90] S. Lange, A. Grigorjew, T. Zinner, P. Tran-Gia, and M. Jarschel, "A multi-objective heuristic for the optimization of virtual network function chain placement," in *2017 29th International Teletraffic Congress (ITC 29)*, vol. 1, pp. 152–160, IEEE, 2017.
- [91] D. Johansson, A. Kassler, and J. Taheri, "On the energy cost of robustness and resiliency for virtual network function placement," in *2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pp. 1–7, 2018.
- [92] X. Shang, Z. Li, and Y. Yang, "Placement of highly available virtual network functions through local rerouting," in *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 80–88, 2018.
- [93] N. He, S. Yang, F. Li, S. Trajanovski, L. Zhu, Y. Wang, and X. Fu, "Leveraging deep reinforcement learning with attention mechanism for virtual network function placement and routing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 34, no. 4, pp. 1186–1201, 2023.
-

-
- [94] B. Addis, D. Belabed, M. Bouet, and S. Secci, "Virtual network functions placement and routing optimization," in *2015 IEEE 4th International Conference on Cloud Networking (CloudNet)*, pp. 171–177, 2015.
- [95] D. Li, P. Hong, K. Xue, and J. Pei, "Virtual network function placement and resource optimization in nfv and edge computing enabled networks," *Computer Networks*, vol. 152, pp. 12–24, 2019.
- [96] R. Moosavi, S. Parsaeefard, M. A. Maddah-Ali, V. Shah-Mansouri, B. H. Khalaj, and M. Bennis, "Energy efficiency through joint routing and function placement in different modes of sdn/nfv networks," *Computer Networks*, vol. 200, p. 108492, 2021.
- [97] A. Laghrissi and T. Taleb, "A survey on the placement of virtual resources and virtual network functions," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1409–1434, 2018.
- [98] S. Demirci and S. Sagiroglu, "Optimal placement of virtual network functions in software defined networks : A survey," *Journal of Network and Computer Applications*, vol. 147, p. 102424, 2019.
- [99] O. A. Wahab, N. Kara, C. Edstrom, and Y. Lemieux, "Maple : A machine learning approach for efficient placement and adjustment of virtual network functions," *Journal of Network and Computer Applications*, vol. 142, pp. 37–50, 2019.
- [100] S. R. Zahedi, S. Jamali, and P. Bayat, "A power-efficient and performance-aware online virtual network function placement in sdn/nfv-enabled networks," *Computer Networks*, vol. 205, p. 108753, 2022.
- [101] M. Chen, S. Zhang, H. Deng, B. Chen, C. Xing, and B. Xu, "Automatic deployment and control of network services in nfv environments," *Journal of Network and Computer Applications*, vol. 164, p. 102677, 2020.
- [102] A. Alleg, R. Kouah, S. Moussaoui, and T. Ahmed, "Virtual network functions placement and chaining for real-time applications," in *2017 IEEE 22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–6, 2017.
- [103] I. A. Ikhelef, J. A. Warwicker, S. Rebennack, M. Y. Saidi, and K. Chen, "Efficient decomposition-based methods for optimal vnf placement and chaining," in *2023 24th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pp. 89–94, 2023.
- [104] O. Alhussein, P. T. Do, Q. Ye, J. Li, W. Shi, W. Zhuang, X. Shen, X. Li, and J. Rao, "A virtual network customization framework for multicast services in nfv-enabled core networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 6, pp. 1025–1039, 2020.
-

-
- [105] J. Sun, F. Liu, H. Wang, and D. O. Wu, "Joint vnf placement, cpu allocation, and flow routing for traffic changes," *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1208–1222, 2023.
- [106] E. Ahmed and M. H. Rehmani, "Mobile edge computing : Opportunities, solutions, and challenges," *Future Generation Computer Systems*, vol. 70, pp. 59–63, 2017.
- [107] E. Cau, M. Corici, P. Bellavista, L. Foschini, G. Carella, A. Edmonds, and T. M. Bohnert, "Efficient exploitation of mobile edge computing for virtualized 5g in epc architectures," in *2016 4th IEEE international conference on mobile cloud computing, services, and engineering (MobileCloud)*, pp. 100–109, 2016.
- [108] S. Wang, Y. Zhao, J. Xu, J. Yuan, and C.-H. Hsu, "Edge server placement in mobile edge computing," *Journal of Parallel and Distributed Computing*, vol. 127, pp. 160–168, 2019.
- [109] I. Hadžić, Y. Abe, and H. C. Woithe, "Server placement and selection for edge computing in the epc," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 671–684, 2018.
- [110] Z. He, K. Li, and K. Li, "Cost-efficient server configuration and placement for mobile edge computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 9, pp. 2198–2212, 2021.
- [111] P. Roy, S. Sarker, M. A. Razzaque, M. M. Hassan, S. A. AlQahtani, G. Aloj, and G. Fortino, "Ai-enabled mobile multimedia service instance placement scheme in mobile edge computing," *Computer Networks*, vol. 182, p. 107573, 2020.
- [112] T. Bahreini and D. Grosu, "Efficient placement of multi-component applications in edge computing systems," in *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*, pp. 1–11, 2017.
- [113] S. Bi, L. Huang, and Y.-J. A. Zhang, "Joint optimization of service caching placement and computation offloading in mobile edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4947–4963, 2020.
- [114] Y. Li and S. Wang, "An energy-aware edge server placement algorithm in mobile edge computing," in *2018 IEEE International Conference on Edge Computing (EDGE)*, pp. 66–73, 2018.
- [115] X. Jiang, P. Hou, H. Zhu, B. Li, Z. Wang, and H. Ding, "Dynamic and intelligent edge server placement based on deep reinforcement learning in mobile edge computing," *Ad Hoc Networks*, vol. 145, p. 103172, 2023.
- [116] K. Poularakis, J. Llorca, A. M. Tulino, I. Taylor, and L. Tassiulas, "Joint service placement and request routing in multi-cell mobile edge computing networks,"
-

-
- in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 10–18, 2019.
- [117] H. Ma, Z. Zhou, and X. Chen, “Predictive service placement in mobile edge computing,” in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 792–797, 2019.
- [118] X. Zhang, Z. Li, C. Lai, and J. Zhang, “Joint edge server placement and service placement in mobile-edge computing,” *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 11261–11274, 2022.
- [119] Y. Wang, Z.-Y. Ru, K. Wang, and P.-Q. Huang, “Joint deployment and task scheduling optimization for large-scale mobile users in multi-uav-enabled mobile edge computing,” *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3984–3997, 2020.
- [120] E. Danna, S. Mandal, and A. Singh, “A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering,” in *2012 Proceedings of IEEE International Conference on Computer Communications (IEEE INFOCOM)*, pp. 846–854, 2012.
- [121] E. Danna, A. Hassidim, H. Kaplan, A. Kumar, Y. Mansour, D. Raz, and M. Segalov, “Upward max min fairness,” in *2012 Proceedings of IEEE International Conference on Computer Communications (IEEE INFOCOM)*, pp. 837–845, 2012.
- [122] E. Amaldi, A. Capone, S. Coniglio, and L. G. Gianoli, “Network optimization problems subject to max-min fair flow allocation,” *IEEE Communications Letters*, vol. 17, no. 7, pp. 1463–1466, 2013.
- [123] L. Yu, J. Sonchack, and V. Liu, “Cebinae : Scalable in-network fairness augmentation,” in *Proceedings of the ACM SIGCOMM 2022 Conference (ACM SIGCOMM)*, p. 219–232, 2022.