Radio access and core functionalities in self-deployable mobile networks

Jad Oueis

▶ **To cite this version:**

Jad Oueis. Radio access and core functionalities in self-deployable mobile networks. Networking and Internet Architecture [cs.NI]. Université de Lyon, 2018. English. NNT : 2018LYSEI095 . tel-02091147v2

HAL Id: tel-02091147
https://theses.hal.science/tel-02091147v2

Submitted on 6 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N°d'ordre NNT : 2018LYSEI095

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
INSA Lyon

Ecole Doctorale EDA 512
Informatique et Mathématiques de Lyon

Spécialité/ discipline de doctorat :
Informatique

Soutenue publiquement le 27/11/2018, par :
Jad Oueis

Radio Access and Core Functionalities in Self-deployable Mobile Networks

---

Devant le jury composé de :

CASETTI, Claudio	Maître de Conférences	Politecnico di Torino	Rapporteur
CHAOUCHI, Hakima	Professeure des Universités	Telecom Sud Paris	Rapporteure
LAGRANGE, Xavier	Professeur des Universités	IMT Atlantique	Rapporteur
CONAN, Vania	Habilité à Diriger des Recherches	Thales	Examinateur
FDIDA, Serge	Professeur des Universités	UPMC	Examinateur
PERROT, Nancy	Docteure	Orange Labs	Examinatrice
VALOIS, Fabrice	Professeur des Universités	INSA-LYON	Directeur de thèse
STANICA, Razvan	Maître de Conférences	INSA-LYON	Co-encadrant de thèse

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Département FEDORA – INSA Lyon - Ecoles Doctorales – Quinquennal 2016-2020

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage **secretariat@edchimie-lyon.fr** INSA : R. GOURDON	M. Stéphane DANIELE Institut de recherches sur la catalyse et l'environnement de Lyon IRCELYON-UMR 5256 Équipe CDFA 2 Avenue Albert EINSTEIN 69 626 Villeurbanne CEDEX **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **http://edeea.ec-lyon.fr** Sec. : M.C. HAVGOUDOUKIAN **ecole-doctorale.eea@ec-lyon.fr**	M. Gérard SCORLETTI École Centrale de Lyon 36 Avenue Guy DE COLLONGUE 69 134 Écully Tél : 04.72.18.60.97 Fax 04.78.43.37.17 **gerard.scorletti@ec-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 INSA : H. CHARLES **secretariat.e2m2@univ-lyon1.fr**	M. Philippe NORMAND UMR 5557 Lab. d'Ecologie Microbienne Université Claude Bernard Lyon 1 Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX **philippe.normand@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://www.ediss-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 INSA : M. LAGARDE **secretariat.ediss@univ-lyon1.fr**	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 Avenue Jean CAPELLE INSA de Lyon 69 621 Villeurbanne Tél : 04.72.68.49.09 Fax : 04.72.68.49.16 **emmanuelle.canet@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 Fax : 04.72.43.16.87 **infomaths@univ-lyon1.fr**	M. Luca ZAMBONI Bât. Braconnier 43 Boulevard du 11 novembre 1918 69 622 Villeurbanne CEDEX Tél : 04.26.23.45.52 **zamboni@maths.univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Marion COMBE Tél : 04.72.43.71.70 Fax : 04.72.43.87.12 Bât. Direction **ed.materiaux@insa-lyon.fr**	M. Jean-Yves BUFFIÈRE INSA de Lyon MATEIS - Bât. Saint-Exupéry 7 Avenue Jean CAPELLE 69 621 Villeurbanne CEDEX Tél : 04.72.43.71.70 Fax : 04.72.43.85.28 **jean-yves.buffiere@insa-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Marion COMBE Tél : 04.72.43.71.70 Fax : 04.72.43.87.12 Bât. Direction **mega@insa-lyon.fr**	M. Jocelyn BONJOUR INSA de Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69 621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **http://ed483.univ-lyon2.fr** Sec. : Viviane POLSINELLI Brigitte DUBOIS INSA : J.Y. TOUSSAINT Tél : 04.78.69.72.76 **viviane.polsinelli@univ-lyon2.fr**	M. Christian MONTES Université Lyon 2 86 Rue Pasteur 69 365 Lyon CEDEX 07 **christian.montes@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Abstract

Self-deployable mobile networks are a novel family of cellular networks, that can be rapidly deployed, easily installed, and operated on demand, anywhere, anytime. They target diverse use cases and provide network services when the classical network fails, is not suitable, or simply does not exist: for example, when the network saturates during crowded events, when first responders need private broadband communication in disaster-relief and mission-critical situations, or when there is no infrastructure in areas with low population density.

These networks are challenging a long-standing vision of cellular networks. Indeed, classical cellular networks are the result of careful planning and deployment strategies. Their fixed and hierarchical architecture is based on a clear physical separation between the radio access network (RAN) and the core network (CN), with an over-provisioned backhaul between them. On the contrary, the rapid deployment nature of self-deployable networks short-circuits the thorough planning phase. The network needs to self-configure and self-organize. Moreover, the split between the RAN and the CN is only functional. In fact, in addition to providing typical RAN functionalities, such as radio signal processing and radio resource management, a base station can also provide those of the CN, such as session management, routing, and authentication, in addition to housing application servers, based on virtualization technologies. As a result, a base station with no backhaul connection to a traditional CN is capable of providing local services to users in its vicinity. To cover larger areas, several base stations must interconnect. With the CN functions co-located with the RAN, the links interconnecting the BSs form the backhaul network. Being setup by the BSs, potentially in an ad hoc manner, the latter may have a limited bandwidth.

In this thesis, we build on the properties distinguishing self-deployable networks to revisit classical RAN problems but in the self-deployable context, and address the novel challenges created by the core network architecture. Starting with the RAN configuration, we propose an algorithm that sets a frequency and power allocation scheme. The latter outperforms conventional frequency reuse schemes in terms of the achieved user throughput, and is robust facing variations in the number of users and their distribution in the network. Once the RAN is configured, we move to the CN organization, and address both centralized and distributed CN functions placements. For the centralized placement, building on the shortages of state of the art metrics, we propose a novel centrality metric that places the functions in a way that maximizes the traffic that can be exchanged in the network. For the distributed placement, we evaluate the number of needed instances of the CN functions and their optimal placement, taking into account the impact on the backhaul bandwidth. We further highlight the advantages of distributing CN functions, from a backhaul point of view. Accordingly, we tackle the user attachment problem to determine the CN instances serving each user when the former are distributed. Finally, with the network ready to operate, and users starting to arrive, we tackle the user association problem. We propose a novel network-aware association policy adapted to the self-deployable network attributes, that outperforms a traditional RAN-based policy. It jointly accounts for the downlink, the uplink, the backhaul and the user throughput request, and mitigates both RAN and backhaul bottlenecks.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Résumé

Les réseaux mobiles auto-déployables sont des réseaux qui peuvent être rapidement déployés, facilement installés, sur demande, n'importe où, et n'importe quand. Ils visent divers cas d'utilisation pour fournir des services aux utilisateurs lorsque le réseau classique ne peut pas être utilisé, ou n'existe pas : par exemple, lorsque le réseau est saturé lors d'événements publics, lorsque les services de secours ont besoin d'un réseau qui leur est dédié dans les situations critiques, ou lorsqu'on a besoin de couverture dans les zones isolées.

Ces réseaux font évoluer l'architecture d'un réseau classique, en éliminant la séparation physique qui existe entre le réseau d'accès et le cœur de réseau. La séparation entre les deux est désormais uniquement fonctionnelle, vu qu'une station de base est colocalisée avec les fonctionnalités traditionnelles du réseau de cœur, telles que la gestion de session et le routage, en plus des serveurs d'applications, à travers la virtualisation de ces derniers. Une station de base, toute seule, sans connexion à un réseau de cœur externe, est capable de fournir des services aux utilisateurs dans sa zone de couverture. Lorsque plusieurs stations de base sont interconnectées, les liens entre elles forment un réseau d'interconnexion, qui risque d'avoir une capacité limitée.

Dans ce travail, nous nous appuyons sur les propriétés qui distinguent les réseaux mobiles auto-déployables pour revisiter des problèmes classiques du réseau d'accès dans ce nouvel contexte, mais aussi pour aborder de nouveaux défis créés par l'architecture du réseau de cœur et du réseau d'interconnexion. Tout d'abord, pour la configuration du réseau d'accès, nous proposons un algorithme qui retourne un schéma d'allocation de fréquences et de puissances pour les stations de base. Notre proposition augmente considérablement les débits des utilisateurs en comparaison avec des schémas classiques de réutilisation de fréquences. Ensuite, nous nous intéressons à l'organisation du cœur de réseau. Nous traitons le problème de placement des fonctionnalités de ce dernier, qu'elles soient centralisées ou distribuées. Pour le placement centralisé, nous proposons une nouvelle métrique de centralité qui permet de placer les fonctions de façon à maximiser le trafic pouvant être échangé dans le réseau. Pour le placement distribué, nous évaluons le nombre de fonctions nécessaires et leur placement optimal, en tenant compte de l'impact sur la capacité du réseau d'interconnexion entre les stations de base. En outre, nous démontrons les avantages d'un placement distribué par rapport à un placement centralisé en terme de consommation de ressources sur le réseau d'interconnexion. Dans le même contexte, nous abordons le problème d'attachement des utilisateurs, lorsque les fonctionnalités du cœur de réseau sont distribuées, pour déterminer par laquelle de ces fonctionnalités un utilisateur est-il servi. Enfin, avec le réseau d'accès configuré et le cœur de réseau organisé, les utilisateurs commencent à arriver. Alors, nous abordons le problème de l'association des utilisateurs. Nous proposons une nouvelle politique d'association adaptée aux propriétés des réseaux auto-déployables. Cette politique réduit la probabilité de blocage par rapport aux politiques classiques basées uniquement sur la qualité de la voie descendante, en tenant compte à la fois des ressources du réseau d'accès, des ressources sur le réseau d'interconnexion, et des demandes des utilisateurs en terme de débit.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Résumé Long

Le déploiement des réseaux cellulaires est généralement le résultat de stratégies de planification rigoureuses, mises en œuvre par les opérateurs mobiles afin de fournir des services aux utilisateurs. L'architecture de ces réseaux est standardisée, fixe, et hiérarchique.

Néanmoins, des propositions récentes sont entrain de revisiter ces notions sur l'architecture à travers l'introduction d'une nouvelle famille de réseaux mobiles qui peuvent être rapidement déployés, facilement installés, sur demande, n'importe où, et n'importe quand. Il s'agit des *réseaux mobiles auto-déployables*, qui constituent le cœur de cette thèse.

L'architecture d'un réseau mobile auto-déployable est particulière et diffère de celle d'un réseau classique. En fait, l'élément clé de cette architecture est la station de base (BS) compacte, facilement déplaçable et déployable. Conçue pour fonctionner en autonomie mais aussi au sein d'un réseau, cette BS dispose non seulement des fonctionnalités typiques du réseau d'accès (gestion de ressources radio), mais intègre également des fonctionnalités du cœur de réseau (routage, authentification). Par conséquent, la BS peut fournir des services aux utilisateurs sans avoir aucune interconnexion à un réseau de cœur externe. Pour cela, la BS est colocalisée avec une entité constituant un cœur de réseau local, qu'on appelle CN local (*local core network*). Cette entité est analogue au cœur de réseau classique, elle fournit les mêmes fonctionnalités de base que ce dernier, et supporte en plus les serveurs d'applications. Le CN local est colocalisé avec la BS en utilisant des technologies de virtualisation de fonctions. Pour élargir la couverture du réseau, plusieurs BSs doivent être interconnectées. Ces BSs ont besoin de se découvrir et de mettre en place le réseau d'interconnexion entre eux, où le trafic de données et de signalisation est routé. Un réseau auto-déployable peut également être connecté à n'importe quel réseau externe, via un réseau d'interconnexion dédié, que ce soit un réseau cellulaire, un cœur de réseau mobile, ou un serveur d'applications. Le réseau est dit auto-déployable, vu que le contrôle des opérateurs sur le déploiement de tels réseaux est réduit, car même le placement précis des BSs peut varier. En plus, le réseau est censé être autonome avec une intervention humaine minimale, nécessitant une auto-configuration et auto-organisation du réseau d'accès et du cœur de réseau.

L'utilité des réseaux auto-déployables réside dans la nécessité de fournir des services dans divers cas d'utilisation, lorsque le réseau cellulaire classique ne convient pas, ou n'existe pas. Par exemple, un défi majeur lors des situations d'urgence et suite à des catastrophes est de maintenir la communication, surtout entre les services de secours (les ambulanciers, les pompiers, les policiers), même lorsque l'infrastructure de communication est endommagée ou complètement détruite. Une des solutions pour rétablir rapidement la communication est d'avoir des réseaux auto-déployables, facilement et rapidement déployés dans ces zones.

Développer des réseaux auto-déployables pour les services d'urgences est l'un des objectifs du projet Fed4PMR, dans lequel cette thèse s'inscrit[1]. En fait, les agents de sécurité civile et les

[1]Fed4PMR fait partie des projets stratégiques de R&D pour la compétitivité (PSPC), dans le cadre du Programme d'Investissement d'Avenir (PIA). Opéré par Bpifrance, et piloté par Thales, le consortium regroupe un grand nombre d'industriels et d'académiques. Dans le cadre du projet, cette thèse a été menée en collaboration avec Thales.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

services de secours utilisent toujours les réseaux de radio professionnels (PMR) pour leurs communications. Les réseaux PMR, caractérisés par leur fiabilité, leur résilience et leur sécurité, sont des réseaux privés offrant une multitude de services vocaux conçus pour répondre aux besoins des communications critiques, mais qui ne fournissent pas des services de données vu leurs très faibles débits. Le consortium Fed4PMR vise à développer un réseau de communication PMR 4G très haut débit fédérateur, alignant les capacités des réseaux PMR aux celles des réseaux commerciaux 4G. Ce réseau doit intégrer au sein d'une même infrastructure plusieurs types de réseaux d'accès : réseaux PMR classiques, réseaux 4G dédiés, réseaux mobiles virtuels, et réseaux auto-déployables avec de bulles tactiques rapidement déployés, n'importe où, n'importe quand. Dans cette thèse, nous nous concentrons sur cette dernière plate-forme de déploiement.

Une autre utilisation des réseaux auto-déployables est d'assurer la couverture dans les zones rurales où aucune couverture ou infrastructure de réseau n'existe. Un exemple récent sur ce sujet est le projet Loon, développé par Google X, qui consiste en une flotte de ballons opérant dans la stratosphère, et offrant une couverture réseau avec des débits élevés aux zones rurales.

De plus, les réseaux auto-déployables peuvent jouer le rôle d'extension temporaire au réseau cellulaire classique. Ils permettent d'augmenter la capacité de ce dernier suite aux augmentations temporaires des demandes d'accès au réseau pour éviter la congestion ou la saturation. C'est le cas dans les grands événements, comme les concerts de musique, et les stades de sports, ou à la suite d'événements inattendus, tels que des catastrophes naturelles.

Dans cette thèse, nous partons des propriétés caractérisant les réseaux auto-déployables, et des différences les distinguant des réseaux classiques, pour définir les problèmes que nous abordons.

En fait, les réseaux auto-déployables sont relativement plus petits qu'un réseau cellulaire classique, avec moins de BSs et une zone de couverture plus restreinte. Ils ne comprennent que peu de BSs interconnectés et desservent un nombre limité d'utilisateurs. Du point de vue architecture, au contraire d'un réseau classique, la séparation entre le réseau d'accès et le cœur du réseau dans un réseau auto-déployable n'est pas physique, mais uniquement fonctionnelle. Toutes les fonctions du cœur de réseau sont colocalisées avec les BSs. L'élimination de cette séparation physique renforce davantage l'autonomie de la BS et lui permet de fonctionner même en état d'isolation, c'est à dire lorsqu'elle ne dispose pas d'une interconnexion dédiée vers un cœur de réseau externe. En outre, la topologie du réseau d'interconnexion est aussi différente de celle d'un réseau classique. Avec les BSs interconnectées, et les fonctions du CN local colocalisées avec elles, tout le trafic de données et de signalisation, habituellement échangé entre chaque BS et le cœur du réseau externe, est maintenant routé localement sur les liens entre les BS. Par conséquent, ces liens inter-BS forment le réseau d'interconnexion. Mis en place par les BSs, souvent de manière ad hoc, le réseau d'interconnexion peut potentiellement avoir une capacité limitée, en fonction de la technologie sans fil utilisée, la qualité des liens inter-BS, et la bande passante disponible. Dans un réseau auto-déployable, le routage sur le réseau d'interconnexion est critique puisque les liens inter-BS sont partagés entre plusieurs BSs, et peuvent avoir des capacités limitées. En ce qui concerne la topologie du réseau, le déploiement rapide de réseaux auto-déployables, souvent à la suite d'évènements inattendus, élimine la phase de planification qui est d'habitude recommandée, voire obligatoire, dans les réseaux cellulaires classiques. Par conséquent, dans un réseau auto-déployable, la topologie peut être aléatoire et présente des irrégularités. Finalement, comme le réseau auto-déployable est souvent dédié pour un usage professionnel par des utilisateurs appartenant à une même organisation, le trafic est principalement intra-réseau, c'est-à-dire que les utilisateurs impliqués dans une communication sont généralement rattachés au même réseau auto-déployable. Le trafic échangé, dans ce cas, est géré par le CN local et reste local au réseau. En plus, la communication avec des réseaux externes est également possible. Le trafic, dans ce cas, est inter-réseau et est acheminé vers le réseau externe via un réseau d'interconnexion. D'autre part, dans certains cas, comme par

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

exemple les réseaux dédiés aux agents de sécurité publique, le trafic a des exigences strictes en termes de débit. Les utilisateurs doivent disposer d'un débit minimal requis, en fonction de la mission en question et du service demandé.

D'une part, ces différences incitent à revisiter des problèmes classiques des réseaux cellulaires mais dans ce nouveau contexte, comme l'allocation des ressources et l'association des utilisateurs. D'autre part, elles soulèvent une série de défis, tel que le placement du CN local au sein du réseau.

Ce manuscrit est organisé comme suit. Après une introduction générale dans le chapitre 1, dans laquelle nous définissons le contexte et les motivations de cette thèse, nous fournissons dans le chapitre 2 les concepts techniques nécessaires, issus des réseaux cellulaires classiques, puis nous présentons un état de l'art sur les concepts les plus pertinents à notre travail. En premier lieu, nous discutons de la conception de BSs rapidement déployables, la virtualisation des fonctionnalités du cœur de réseau, et le développement des plates-formes de déploiement. Ensuite, nous présentons les technologies sans fils possibles pour interconnecter les BSs, et les différentes méthodes d'allocation de fréquences au réseau. Enfin, nous effectuons une revue de l'état de l'art sur les problèmes traités dans cette thèse, tels que : le problème d'allocation de fréquences et de puissances aux BSs, le problème de placement des fonctions dans différents types de réseaux et architectures, et le problème d'association des utilisateurs dans les réseaux cellulaires.

Dans le chapitre 3, nous proposons un algorithme d'allocation de fréquences et de puissances pour un réseau auto-déployable. L'algorithme retourne une carte de puissance qui détermine la puissance avec laquelle chaque BS transmet sur chaque canal. Cette carte de puissance est calculée uniquement à partir de la topologie du réseau, sans aucune information disponible sur la densité ou la distribution des utilisateurs à venir. Elle est calculée a priori, puis fixée et suivie par les BSs tout au long de leur fonctionnement. L'algorithme consiste à résoudre un problème d'optimisation complexe, non-convexe, non-linéaire, à travers une série de transformations, basées sur la programmation signomiale. Le problème est d'abord transformé en un programme géométrique complémentaire. Ce dernier est résolu itérativement en le transformant en un programme géométrique, qui est ensuite résolu en le transformant en un problème convexe via une transformation logarithmique. Le schéma d'allocation proposé donne des résultats supérieurs aux schémas d'allocation classiques basés sur la réutilisation des fréquences en ce qui concerne le débit des utilisateurs. Bien que la carte de puissance soit statique, elle assure un bon fonctionnement, et permet une augmentation des débits des utilisateurs, face aux variations du nombre d'utilisateurs concurrents dans le système et leur distribution, et face aux différentes politiques de réseau mises en œuvre.

Une fois le réseau d'accès est configuré, nous passons au réseau de cœur. Nous abordons dans le chapitre 4 le problème de placement du CN local dans un réseau auto-déployable, lorsque le CN local est centralisé. Nous déterminons avec quelle BS les fonctions du CN local doivent être colocalisées lorsqu'un réseau auto-déployable est créé. Nous proposons une nouvelle métrique de centralité dans un réseau, appelée centralité de flux. La centralité de flux d'un nœud du réseau est définie comme le trafic maximal pouvant être envoyé simultanément par tous les autres nœuds du réseau vers ce nœud, sous certaines contraintes de capacité et de répartition de charge. Nous montrons que, afin de maximiser le volume du trafic échangé entre les BS et le CN local, ce dernier doit être colocalisé avec la BS ayant la centralité de flux maximale. Nous comparons la métrique proposée à différentes métriques de centralité de l'état de l'art. Nous calculons la perte potentielle dans le trafic total reçu par le CN local, lorsque ce dernier n'est pas placé sur le nœud ayant la centralité de flux maximale. De plus, nous mettons en évidence certaines propriétés clés du placement du CN local, telles que la forte dépendance entre le placement optimal et la bande passante du réseau d'interconnexion. Le calcul de la centralité de flux repose sur la résolution d'un problème d'optimisation linéaire. De plus, nous proposons un certain nombre d'expressions analytiques permettant le calcul direct de la métrique, pour certaines topologies particulières.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Toujours sur le placement du CN local, nous comparons dans le chapitre 5 un placement centralisé avec un placement distribué du CN local. La comparaison se fait de point de vue consommation de bande passante sur le réseau d'interconnexion, en comptabilisant le trafic de données et de signalisation consommé dans chacun des cas. Nous montrons que la distribution des fonctions du routage du CN local sur toutes les BSs est moins coûteuse de point de vue consommation sur le réseau d'interconnexion en comparaison avec une fonction de routage centralisée. Les problèmes qui déterminent le placement optimal du CN local lorsque celui-ci est centralisé, et le nombre d'instances et leur placement lorsqu'il est distribué, sont formulés comme des programmes d'optimisation non-linéaire mixte en nombres entiers. L'objectif des problèmes est de minimiser la consommation sur le réseau d'interconnexion causée par le trafic de données et de signalisation échangé entre les BSs et le CN local.

Dans la suite du chapitre 5, toujours dans le même contexte de minimisation du trafic sur le réseau d'interconnexion, nous évaluons l'optimisation de l'association des utilisateurs (affecter un utilisateur à une BS) et l'optimisation de leur attachement (enregistrer un utilisateur à une instance du CN local). Nous montrons qu'une association qui prend en compte le réseau d'interconnexion permet de réduire considérablement la consommation sur le réseau d'interconnexion par rapport aux politiques classiques d'association basées uniquement sur le réseau d'accès. En outre, nous montrons qu'un attachement optimal des utilisateurs aux instances du CN local réduit la consommation sur le réseau d'interconnexion, notamment lorsque le trafic de signalisation est important. Sinon, le gain obtenu en optimisant l'attachement est marginal. Nous formulons les problèmes qui retournent l'association et/ou l'attachement avec l'objectif de minimiser la consommation sur le réseau d'interconnexion en se basant sur la programmation quadratique en nombres entiers.

Avec le réseau d'accès configuré, et les fonctionnalités du CN local placées dans le réseau, ce dernier est prêt à démarrer et accueillir les utilisateurs. Dans le chapitre 6, nous abordons le problème de l'association des utilisateurs. Nous proposons une nouvelle politique d'association adaptée aux réseaux auto-déployables. Cette politique prend en compte des paramètres clés du réseau, souvent ignorés dans les politiques d'association dans les réseaux classiques, tels que la capacité limitée du réseau d'interconnexion, les ressources disponibles sur le lien montant et les requêtes des utilisateurs, en plus des ressources disponibles sur le lien descendant. Nous proposons un algorithme exécuté à l'arrivée de chaque requête de flux initiée par un utilisateur. L'algorithme met en place une procédure d'admission de flux, suivie d'une procédure de décision sur l'association si le flux est accepté. L'objectif de l'association est de maximiser les ressources restantes dans le réseau en prenant en considération le lien descendant, le lien montant et le réseau d'interconnexion, pour maintenir la probabilité de refus d'un nouveau flux dû à l'insuffisance des ressources aussi bas que possible. L'évaluation de la politique est basée sur des simulations du réseau, en fonction d'un nombre de paramètres tels que la capacité du réseau d'interconnexion, le modèle de trafic de données et le nombre d'utilisateurs concurrents. Nous montrons que la politique d'association proposée réduit de façon significative la probabilité de blocage du flux par rapport à une politique d'association traditionnelle basée uniquement sur le lien descendant. Notre politique s'adapte aux différentes contraintes de réseau en atténuant les goulots d'étranglement, qu'ils soient causés par le réseau d'accès ou par le réseau d'interconnexion. Nous formulons le problème d'association sous la forme d'un programme d'optimisation non-linéaire mixte en nombres entiers, résolu à chaque requête de flux pour déterminer l'association des utilisateurs impliqués. Néanmoins, en raison de la taille limitée des réseaux auto-déployables, nous nous appuyons sur un calcul de force brute simple et rapide dans la phase d'admission du flux et dans la décision d'association, ce qui rend l'algorithme encore plus pratique.

Finalement, dans le chapitre 7, nous concluons le manuscrit en mettant en avance les voies possibles pour développer davantage le travail présenté, et les perspectives de recherche.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Contents

Abstract **v**

Résumé **vii**

Résumé Long **ix**

Table of Contents **xvi**

List of Figures **xvii**

List of Tables **xxi**

List of Acronyms **xxiii**

List of Notations **xxv**

1 Introduction **1**
1.1 Context 1
1.2 Motivation 4
1.3 Thesis outline and contributions 5

2 Challenges in Self-deployable Mobile Networks **8**
2.1 Technical overview 8
2.1.1 Overview of cellular networks evolution 8
2.1.2 4G cellular networks 8
2.1.3 Professional mobile radio networks 10
2.2 Self-deployable networks 12
2.3 Self-deployable base stations 13
2.4 Mobile core network virtualization 14
2.5 Deployment platforms 15
2.6 Wireless backhaul 16
2.6.1 In-band backhaul 17
2.6.2 Out-of-band backhaul 17
2.7 Spectrum allocation 18
2.7.1 Dedicated spectrum 18
2.7.2 Cognitive radio technology 18
2.8 Frequency and power allocation 18
2.8.1 Static reuse schemes 19
2.8.2 Dynamic coordination schemes 21

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

2.8.3 Addressed challenges . . . 22
2.9 Local core network functions placement . . . 22
2.9.1 Placement of virtualized network functions . . . 23
2.9.2 Placement in wireless mesh networks . . . 23
2.9.3 Placement in wireless sensor networks . . . 24
2.9.4 Addressed challenges . . . 24
2.10 User association . . . 25
2.10.1 User association in HetNets . . . 25
2.10.2 Uplink-aware user association . . . 26
2.10.3 Backhaul-aware user association . . . 26
2.10.4 Addressed challenges . . . 27

3 Frequency and Power Allocation Scheme 28
3.1 Introduction . . . 28
3.2 System model . . . 30
3.3 Selection of a robust power map . . . 31
3.3.1 Problem formulation . . . 32
3.3.2 Solving the problem . . . 32
3.4 Results on a toy scenario . . . 35
3.4.1 Finding the power map . . . 36
3.4.2 Evaluating the power map performance . . . 37
3.5 Algorithm parameterization . . . 39
3.5.1 Number of sub-bands . . . 40
3.5.2 Number of realizations . . . 41
3.5.3 Number of users . . . 42
3.5.4 Network topology . . . 43
3.6 Dynamic simulation . . . 45
3.6.1 Simulation setting . . . 45
3.6.2 Simulation results . . . 47
3.7 Global scheduling problem . . . 47
3.7.1 Problem formulation . . . 48
3.7.2 Numerical results . . . 49
3.8 Conclusion . . . 49

4 Core Network Functions Centralized Placement 51
4.1 Introduction . . . 51
4.2 System model . . . 53
4.2.1 Mathematical notation . . . 53
4.3 Local CN placement problem . . . 54
4.3.1 Centrality in networks . . . 54
4.3.2 Numerical example . . . 55
4.4 Flow centrality: a metric for local CN placement . . . 56
4.4.1 Local CN placement criteria . . . 56
4.4.2 Flow centrality . . . 56
4.4.3 Computing flow centrality . . . 57
4.5 Computing flow centrality in canonical topologies . . . 58
4.5.1 Path graph . . . 58
4.5.2 Balanced tree . . . 60

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

4.6 Flow centrality properties . . . 63
4.6.1 Position of the node with maximum flow centrality . . . 64
4.6.2 Link capacities distribution . . . 65
4.6.3 Link capacities average . . . 65
4.6.4 Link capacities range . . . 66
4.7 Benchmarking flow centrality . . . 67
4.8 Large scale networks . . . 69
4.9 Conclusion . . . 70

5 Core Network Functions Distribution 72
5.1 Introduction . . . 72
5.2 System model . . . 73
5.2.1 Core network and backhaul . . . 73
5.2.2 Traffic model . . . 74
5.2.3 Radio access network . . . 76
5.3 Problem overview . . . 77
5.4 Number and placement of S-GWs . . . 78
5.4.1 One default S-GW in the network . . . 79
5.4.2 All BSs co-located with S-GWs . . . 81
5.4.3 Optimized S-GW placement . . . 82
5.4.4 Numerical results . . . 83
5.5 Optimizing user association . . . 86
5.6 Attachment per flow . . . 88
5.6.1 Problem formulation . . . 89
5.6.2 Numerical results . . . 89
5.7 Impact of network topology . . . 90
5.8 Conclusion . . . 91

6 Network-Aware User Association Policy 92
6.1 Introduction . . . 92
6.2 System model . . . 93
6.2.1 Core network and backhaul . . . 93
6.2.2 Traffic model . . . 94
6.2.3 Radio access network . . . 95
6.2.4 Resource allocation . . . 96
6.3 Association policy overview . . . 97
6.4 Flow admission control . . . 98
6.4.1 Association feasibility on the RAN . . . 98
6.4.2 Association feasibility on the backhaul . . . 98
6.4.3 Association feasibility . . . 99
6.5 User association decision . . . 99
6.6 Problem formalization . . . 100
6.7 Numerical evaluation . . . 102
6.7.1 Simulation settings . . . 102
6.7.2 Simulation results . . . 102
6.8 Revisiting assumptions on user association . . . 110
6.8.1 Re-association . . . 110
6.8.2 Split DL/UL association . . . 112

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

6.9 SFR-based power map on the downlink 113
6.9.1 Frequency and power allocation model for the downlink 114
6.9.2 Numerical results 115
6.10 Conclusion 117

7 Conclusion and Perspectives **118**
7.1 Summary 118
7.2 Open perspectives 119

Bibliography **121**

List of Publications **131**

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

List of Figures

1.1 Self-deployable mobile network. 2
1.2 Comparison between self-deployable and classical cellular networks. 3

2.1 The EPS architecture. 9
2.2 LTE frames. 11
2.3 Equal power/reuse 1 scheme (EP-R1). 19
2.4 Equal power/reuse 3 scheme (EP-R3). 20
2.5 Partial frequency reuse scheme (PFR). 20
2.6 Soft frequency reuse scheme (SFR). 21

3.1 MCS-based step function with 15 discrete rates, and its continuous approximation $\tilde{\Psi}(\cdot)$, with $R_{max} = 963$ Kb/s, $\nu = 0.168$ and $\Delta = 0.43$. 36
3.2 A network topology with 5 BSs. 37
3.3 Power map for the studied network topology, obtained with $|\Omega| = 10$, $N_\omega = 100$ users, and $b = 5$ sub-bands. 37
3.4 Comparison of the throughput geometric mean and arithmetic mean with a soft frequency reuse scheme with the obtained power map (PM), a classical equal power with reuse 1 scheme (EP-R1), and a classical equal power with reuse 3 scheme (EP-R3). 39
3.5 The cumulative distribution function of the users' individual throughputs, with a soft frequency reuse scheme with the obtained power map (PM), a classical equal power with reuse 1 scheme (EP-R1), and a classical equal power with reuse 3 scheme (EP-R3). 39
3.6 Throughput geometric mean function of the number of sub-bands b, tested on $|\pi| = 100$ test realizations, with $N_\pi = 100$ users in each test realization, with a power map obtained for $|\Omega| = 10$ calibration realizations, and $N_\omega = 100$ users. . 40
3.7 Throughput geometric mean function of the number of calibration realizations $|\Omega|$ used to obtain the power map, for $b = 5$ sub-bands, and $N_\omega = 100$ users, tested on $|\Pi| = 100$ test realizations, with $N_\pi = 100$ users in each test realization. . . . 42
3.8 Throughput geometric mean function of the number of users in test realizations N_π, for different values of the number of users in calibration realizations N_ω, for $b = 5$ sub-bands, and $|\pi| = 100$ test realizations. 43
3.9 A network topology with 15 BSs. 44
3.10 Power map for the network topology with 15 BSs, obtained for $|\Omega| = 10$ calibration realizations, $N_\omega = 150$ users, and $b = 5$ sub-bands. 44
3.11 The geometric mean of the throughput of users associated to each BS, with a soft frequency reuse scheme with the obtained power map (PM), and a classical equal power with reuse 1 scheme (EP-R1). 45

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

3.12 Throughput geometric mean function of the number of sub-bands b, for $|\Omega| = 10$ calibration realizations, and $N_\omega = 150$ users, tested on $|\Pi| = 150$ test realizations, with $N_\pi = 150$ users in each test realization. 46
3.13 Average user sojourn time to download a file of 50 MB, as a function of the user arrival rate, with a classical equal power reuse 1 scheme and with the power map obtained for $|\Omega| = 10$ calibration realizations, $N_\omega = 100$, and $b = 5$ sub-bands, for different user association schemes (Best SINR and Load Aware). 47
3.14 Comparison of the average, minimum, and maximum throughput geometric mean in the global scheduling optimal problem (OPT), with an SFR-based power map (PM), and a classical equal power with reuse 1 (EP-R1) scheme. 50

4.1 An example of a self-deployable network topology with a centralized Local CN co-located with one BS and inter-BS backhaul links of capacity $c(u, v)$. 53
4.2 Limitations of state of the art centrality metrics. 55
4.3 Path graph with n nodes. 59
4.4 A balanced tree rooted at v_{00}, with height $h = 4$ and branching factor $\tau = 2$. . . . 60
4.5 A node v_{ij} at depth i within a balanced tree, branched into τ symmetrical subtrees, each having a height $h - i - 1$, and comprising $des(i)/\tau$ nodes. 61
4.6 Tree structure with node v_{10} set as unique destination in the tree for flow centrality computation. 62
4.7 A random geometric graph topology, with random link capacities $c(u, v) \in [0, 100]$. All nodes transmit $\rho_{max} = 18$ units of traffic towards the Local CN co-located with node 2. The values on the links represent $f(u, v)/c(u, v)$. 64
4.8 Variation of the average maximum flow centrality ρ_{max} function of the average link capacity c_{avg}, in random geometric graphs, for different values of the radius τ, with a constant capacity range Δc. 66
4.9 A random geometric graph topology, with link capacities $c(u, v) = 50$. All nodes transmit $\rho_{max} = 14.28$ units of traffic towards the Local CN co-located with node 2. The values on the links represent $f(u, v)/c(u, v)$. 67
4.10 The percentage of scenarios where the node with maximum flow centrality is identical to the nodes maximizing other centralities, in random geometric graphs, for different capacity ranges, with constant average capacity $c_{avg} = 50$. 68
4.11 Relative traffic loss when Local CN is placed on the node maximizing centrality metrics other than the flow centrality, in random geometric graphs, for $\Delta c = 100$. 69
4.12 The percentage of scenarios where the node with maximum flow centrality is identical to the nodes maximizing other centralities, function of the number of nodes. 69
4.13 The average relative traffic loss when the Local CN is placed on the node maximizing centrality metrics other than the flow centrality, function of the number of nodes. 70

5.1 An example on the different data and signaling traffic paths for a flow between two UEs, in scenarios with different Local CN functions placement. 79
5.2 Data and signaling traffic paths between BSs, and their corresponding bit rates d_f and Si_f, for a flow f, when there is one S-GW in the network ($\mathcal{P}_{1/g/g}$), or when attachment is optimized per flow ($\mathcal{P}_{o/g/of}$, and $\mathcal{P}_{o/o/of}$) 80
5.3 Data and signaling traffic paths between BSs, and their corresponding bit rates d_f and Si_f, for a flow f, when there are $|\mathcal{J}|$ S-GWs ($\mathcal{P}_{\mathcal{J}/g/g}$ and $\mathcal{P}_{\mathcal{J}/o/g}$). 82

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

5.4 Data and signaling traffic paths between BSs, and their corresponding bit rates d_f and Si_f, for a flow f, when the number of S-GWs is optimized ($\mathcal{P}_{o/g/o}$ and $\mathcal{P}_{o/o/o}$). 83
5.5 A network topology. 84
5.6 The total backhaul bandwidth consumption in $\mathcal{P}_{1/g/g}$, $\mathcal{P}_{\mathcal{J}/g/g}$, and $\mathcal{P}_{o/g/o}$, function of σ. 85
5.7 User attachment distribution in $\mathcal{P}_{1/g/g}$, $\mathcal{P}_{\mathcal{J}/g/g}$, and $\mathcal{P}_{o/g/o}$, for the studied topology and user distribution, with signaling traffic at $\sigma = 6\%$. 86
5.8 The relative backhaul bandwidth consumption reduction δ function of σ, in $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, and $\mathcal{P}_{o/o/o}$. 87
5.9 User association distribution in $\mathcal{P}_{\mathcal{J}/g/g}$, $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, and $\mathcal{P}_{o/o/o}$, for the studied topology and user distribution, with signaling traffic at $\sigma = 6\%$. 88
5.10 The relative backhaul bandwidth consumption reduction δ function of σ, in $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, $\mathcal{P}_{o/o/o}$, $\mathcal{P}_{o/g/of}$, and $\mathcal{P}_{o/o/of}$. 90
5.11 The average relative backhaul bandwidth consumption reduction δ function of σ, in $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, $\mathcal{P}_{o/o/o}$, $\mathcal{P}_{o/g/of}$, and $\mathcal{P}_{o/o/of}$, for different network topologies. 91

6.1 Intra-network and inter-network flows. 94
6.2 A network topology. 103
6.3 Blocking probability function of the percentage of intra-network flows out of all flow requests β, for different inter-BS link capacities C_l. 103
6.4 Blocking probability function of the average flow arrival rate λ_f, for different inter-BS link capacities C_l. 104
6.5 Blocking causes for best SINR and NAS, for $C_l = 5$ Mb/s. 105
6.6 Blocking causes for best SINR and NAS, for $C_l = 1$ Mb/s. 106
6.7 Blocking probability and blocking causes function of the UL traffic ratio out of the total total traffic α. *To improve readability, please note that the figures have different scales.* 107
6.8 Blocking probability for different random topologies. 107
6.9 Blocking probability function of the inter-BS link capacities C_l, for different average arrival rates λ_f. 108
6.10 An example with two UEs, u and v, having the possibility of associating to one of two BS pairs, $(0, 0)$ or $(0, 1)$. 109
6.11 Re-association of a UE with an ongoing flow and a new flow request (NAS-Re). . 111
6.12 Blocking probability function of the average flow arrival rate λ_f, when re-association is allowed (NAS-Re), for different number of UEs in the network. 111
6.13 Example of split DL/UL association (NAS-Split). 112
6.14 Blocking probability function of the average flow arrival rate λ_f, when split DL/UL association is allowed (NAS-Split). 113
6.15 Blocking causes for NAS with joint DL/UL association, and NAS-Split with split DL/UL association. 114
6.16 Computed power map for the studied topology. 116
6.17 Blocking probability for best SINR and NAS, under two allocation schemes: (a) EP-R3 on DL and UL, (b) PM on DL and EP-R3 on UL. 116

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

List of Tables

3.1 Solving time, in minutes, function of number of sub-bands b, the number of calibration realizations $|\Omega|$, and the number of users in each calibration realization N_ω, for $\epsilon = 10^{-6}$. 41

5.1 Optimization problems nomenclature & summary. 78
5.2 SINR thresholds and the corresponding data rates based on the MCS. 84

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

List of Acronyms

2G	Second generation mobile telecommunications
3G	Third generation mobile telecommunications
3GPP	Third Generation Partnership Project
4G	Fourth generation mobile telecommunications
5G	Fifth generation mobile telecommunications
BS	Base station
CAPEX	Capital expenditure
CDF	Cumulative distribution function
CN	Core network
DL	Downlink
EDGE	Enhanced Data rates for GSM Evolution
eNodeB	Evolved NodeB
EPC	Evolved packet core
EP-R$_r$	Equal power, reuse r
EPS	Evolved packet system
E-UTRAN	Evolved UMTS radio access network
FFR	Fractional frequency reuse
GM	Geometric mean
GP	Geometric programming
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HAPs	High altitude platform
HetNet	Heterogeneous network
HSPA	High Speed Packet Access
HSS	Home subscriber server
IAB	Integrated access and backhaul
ICI	Inter-cell interference
ICIC	Inter-cell interference coordination
IMS	IP multimedia subsystem
IOPS	Isolated E-UTRAN operation for public safety
LAPs	Low altitude platform
LTE	Long Term Evolution
MCS	Modulation and coding scheme
MILP	Mixed integer linear programming
MINLP	Mixed-integer non-linear programming
MIQP	Mixed integer quadratic programming
MME	Mobility management entity

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

mmWave	Millimeter-wave
NAS	Network-aware user association
NFV	Network function virtualization
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
OPEX	Operational expenditure
PCRF	Policy control and charging function
PDN	Packet data network
PFR	Partial frequency reuse
P-GW	Packet data network gateway
PMR	Professional mobile radio
PRB	Physical resource block
QoE	Quality of experience
QoS	Quality of service
RAN	Radio access network
SAE	System architecture evolution
SFR	Soft frequency reuse
S-GW	Serving gateway
SINR	Signal to interference and noise ratio
S-MVNO	Secured-mobile virtual network operator
TDD	Time division duplex
UAVs	Unmanned aerial vehicle
UDN	Ultra-dense network
UE	User equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
VNF	Virtualized network function
WMN	Wireless mesh network
WSN	Wireless sensor network

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

List of Notations

The following list is non-exhaustive. It consists of the common notations used throughout the thesis, in addition to the notations specific to each chapter.

$\mathcal{J}$	Set of BSs
$\mathcal{U}$	Set of UEs
$\mathcal{U}_j$	Set of UEs associated to BS j
$\mathcal{L}$	Set of backhaul inter-BS links
M	Total number of orthogonal channels
T	Number of time-slots in a frame
$\mathcal{K}_j^{DL}$	Number of channels allocated to BS j on the DL
$\mathcal{K}_j^{UL}$	Number of channels allocated to BS j on the UL
P_{BS}	BS maximum transmission power
P_{UE}	UE maximum transmission power
$\mathcal{I}_j^{DL}$	Set of BSs interfering with BS j on the DL
$\mathcal{I}_j^{UL}$	Set of BSs interfering with BS j on the UL
I_j^{UL}	Estimate of the average per channel UL interference on BS j
r	Frequency reuse factor
$\gamma_{u,j}^{DL}$	Per channel SINR between UE u and BS j on the DL
$\gamma_{u,j}^{UL}$	Per channel SINR between UE u and BS j on the UL
$R_{u,j}^{DL}$	Per channel rate between UE u and BS j on the DL
$R_{u,j}^{UL}$	Per channel rate between UE u and BS j on the UL
Ψ	Step function mapping SINR to rate based on the modulation and coding scheme
$\Gamma_{u,j}$	Path loss between UE u and BS j
$G_{u,j}$	Channel gain between UE u and BS j
$D_{u,j}$	Distance between UE u and BS j
G^a	Antenna gain
E	Equipment losses
$\mathcal{N}_0$	Guassian noise
η	Radius of a random geometric graph

Chapter 3	
$\mathcal{S}$	Set of sub-bands
b	Total number of sub-bands
k	Number of channels per sub-band
P_j^s	Transmission power of BS j on the channels of sub-band s
$\gamma_{u,j}^s$	SINR per channel between UE u and BS j on sub-band s
$R_{u,j}^s$	Rate seen by UE u from BS j on sub-band s
$\alpha_{u,j}^s$	Proportion of time during which all sub-band s is allocated to UE u on BS j

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

ϕ_u	Throughput of UE u
Φ	Throughput geometric mean
Ω	Set of calibration realizations ω
N_ω	Number of users in a calibration realization ω
Π	Set of test snapshots π
N_Π	Number of users in test snapshots π
λ	User arrival rate
$x_u^{c,t}$	Boolean indicating if UE u is assigned the PRB at channel c and time slot t
$P_{u,j}^{c,t}$	BS j transmission power on the PRB at channel c and time slot t
$\gamma_{u,j}^{c,t}$	SINR between UE u and BS j on the PRB at channel c and time slot t
$R_{u,j}^{c,t}$	Rate seen by UE u from BS j on the PRB at channel c and time slot t

Chapter 4

n	Number of nodes
$\mathcal{O}$	Set of source nodes
$\mathcal{D}$	Set of destination nodes
$c(u,v)$	Capacity of an edge between nodes u and v
c_{min}	Minimum link capacity
c_{max}	Maximum link capacity
c_{avg}	Average link capacity
Δc	Link capacity range
$f(u,v)$	Flow on an edge between nodes u and v
$z(u,d)$	Supply function between a source node u and destination node d
$\bar{\rho}(d)$	Flow centrality of node d
ρ_{max}	Maximum flow centrality
h	Height of a balanced tree
τ	Degree of a balanced tree
$des(u)$	Descendants of a node u in a balanced tree
$\epsilon_\lambda(u)$	Relative traffic loss on node u

Chapter 5

$\mathcal{F}$	Set of flows f
p	Probability of a flow between two users
d_f	Data rate of flow f
Si_f	Bit rate of signaling traffic S_i of a flow f
σ	Percentage of signaling rate with respect to data rate
$Z_{j,j'}^l$	Boolean indicating if link l belongs to the routing path from BSs j and j'
$X_{u,j}$	Boolean indicating if UE u is associated to BS j
W_j	Boolean indicating if the MME is co-located with BS j
G_j	Boolean indicating if the S-GW is co-located with BS j
$Y_{u,j}$	Boolean indicating if UE u is attached to S-GW j
$A_{f,j}$	Boolean indicating if flow f is attached to S-GW j
C_l^d	Total bandwidth consumed by data traffic of all the flows on a link l
C_l^{Si}	Total bandwidth consumed by the signaling traffic Si of all flows on a link l
C_l	Total bandwidth consumed by all the flows (data and signaling) on a link l
δ	Relative backhaul bandwidth consumption reduction

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 6	
N_j	Number of users associated to BS j
λ_f	Average flow arrival rate
μ_f	Average flow duration
$d_{u,v}$	Data rate of a flow between UEs u and v
β	Percentage of intra-network flows
α	Percentage of UL traffic
$\mathbb{F}$	Set of candidate BS pairs for association
$P(i,j)$	A routing path from BS i to BS j
$\mathcal{P}_{i,j}$	Set of feasible routing paths from BS i to j
C_l	Capacity of a link l
A_l	Remaining capacity of a link l
$m_{u,j}^{DL}$	Number of channels needed by UE u from BS j on the DL
$m_{u,j}^{UL}$	Number of channels needed by UE u from BS j on the UL
M_j^{DL}	Remaining number of DL channels on BS j
M_j^{UL}	Remaining number of UL channels on BS j
$L_{P(i,j)}^{BH}$	Normalized remaining capacity of the most charged link of $P(i,j)$
$L_{(i,j)}^{DL}$	Normalized remaining DL channels of the most charged BS in a pair (i,j)
$L_{(i,j)}^{UL}$	Normalized remaining UL channels of the most charged BS in a pair (i,j)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 1

Introduction

1.1 Context

Cellular networks have been continuously reshaping communication as we know it, through the rapid evolution of standards, products, and use cases. Since the early 1990s, when the first analog generation of wireless cellular technology was replaced by digital communication, cellular networks have not ceased evolving: from voice-centric circuit-switched networks in the second generation (2G), to packet-switched third generation networks (3G), followed by an all-IP fourth generation (4G), culminating today in the paradigm shifting fifth generation (5G) [1].

With a fixed, hierarchical, and standardized architecture, cellular networks are generally the result of careful planning and deployment strategies, implemented by mobile operators in order to provide coverage and data services to users [2]. Nevertheless, recent proposals are challenging this vision through mobile networks that can be rapidly deployed, easily installed, and operated on demand, anywhere, anytime. We denominate such networks as *self-deployable networks*, and they constitute the subject of this thesis.

The introduction of self-deployable networks emanates from the need of providing users with coverage and data services in a variety of uses cases, when a classical cellular network fails or is not suitable. For instance, a major challenge in the aftermath of natural and man-made disasters is maintaining communication [3]. Ensuring communication, notably between first responders (e.g., paramedics, firefighters, police officers), is crucial for saving people's lives. However, communication failures are common in emergency situations. For example, the tropical storm Harvey, that hit the United States of America in 2017, caused service outage in up to 90% of the cell sites in some of the affected regions [4]. Self-deployable networks are one of the foreseen solutions for having easily and rapidly deployed networks, when the existing communication infrastructure is damaged or completely destroyed. Such networks allow a quick recovery of communication in disaster-hit areas. Being easily transported and installed, first responders can arrive on site carrying base stations (BSs), on vehicles or in backpacks [5, 6], for example, install them, and have an operating network in a short amount of time, providing them with the mission-critical services necessary for their intervention. The network can also be connected to a control room overseeing the mission [5].

Developing self-deployable networks for first responders is one of the objectives of Fed4PMR, the project in which this thesis takes part [7]. In fact, first responders still rely today on professional mobile radio (PMR) networks for their communications. PMR networks, characterized by their reliability, resiliency, and security, are private networks that offer a multitude of voice-centric services designed to meet first responders mission critical requirements, but fall short when it comes to data services due to their low data rates [8]. Substantial efforts are being invested to bring PMR

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

up to date with modern commercial cellular networks [8]. In this context, the Fed4PMR project aims at developing high data rate 4G-based PMR networks for security agencies and emergency services. One of the envisioned 4G PMR deployment platforms, on which we particularly focus, is self-deployable networks.

On the other hand, a connectivity problem exists in rural areas around the world, where no coverage or network infrastructure exists. Indeed, in 2016, 31% of the world's population did not have 3G coverage, most of them living in rural areas [9]. Telecommunication companies and operators opt out from investing in cellular networks or any other infrastructure in such areas, mainly because of the low population density and low returns. Having self-deployable networks is one way to ensure easy and cheap coverage and internet connectivity in remote areas, by relying on innovative deployment platforms. A recent example is the Loon project, developed by Google X, with a fleet of balloons operating at the stratosphere, and providing network coverage with high bit rates to large rural areas [10]. In the same context, in some isolated geographical areas, with no existing infrastructure, communication may be imperative. This is the case, for example, of workers on offshore drilling platforms, or first responders during mission-critical operations in isolated areas, such as search and rescue missions in valleys or undergrounds. Self-deployable networks are a way to provide local services and reliable communication in such situations.

Moreover, temporary surges in demand represent one of the leading causes of cellular network congestion and saturation [11]. This is common in crowded events, such as music concerts, festivals, and football matches, where accessing network services is usually difficult. The increase in demand is due to a large number of users simultaneously trying to share videos and pictures, live-stream the event, receive and place phone calls, and download media content [11]. This is also the case following unexpected events, such as natural disasters, or terrorist attacks, where people tend to communicate with family and friends. With the main infrastructure failing to handle this demand overload, self-deployable networks can act as network extensions to quickly increase the network capacity. Besides, they can relief the network congestion by handling local communication among users within the same covered area. Furthermore, self-deployable networks can serve as private networks for crowd management in such public events. They can provide a dedicated platform for reliable communication among security officers, in order to maintain security and protect people at crowded events.

The architecture of a self-deployable mobile network is presented in Fig. 1.1.

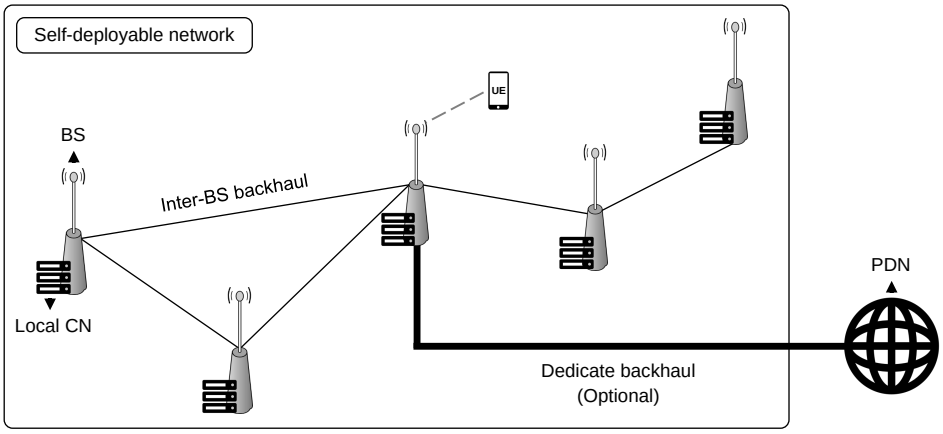


Figure 1.1: Self-deployable mobile network.

The key element in this architecture is the rapidly-deployable, easily movable BS. Conceived

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

to function both as self-contained and within a network, this BS does not only have typical radio access network (RAN) capabilities (e.g., radio signal processing, radio resource management), but also integrates core network (CN) functions (e.g., routing, authentication, session management) [12, 13]. Hence, the BS can provide network coverage to users in its vicinity without any backhaul connection to an external CN. The BS is co-located with, or at least has access to, an entity called Local CN. This entity is analogous to the classical mobile core network, providing the same basic functionalities as the latter, in addition to housing the application servers. The Local CN can be co-located with the BS using function virtualization technologies [13]. To cover larger areas, several BSs must interconnect. Those BSs need to discover each other and set up a backhaul network interconnecting them, where data and signaling traffic can be exchanged. A self-deployable network can also have connectivity via a dedicated backhaul to an external packet data network (PDN), such as a cellular network, a core network, an application server, or the Internet.

The control of network operators over the deployment of such networks is reduced, as even the precise placement of the BSs can vary. Moreover, the network is supposed to be as autonomous as possible, involving minimal human intervention: the RAN infrastructure is required to self-configure and self-organize, and the backhaul is set up in an ad-hoc manner, hence the *self-deployable* nomenclature.

Self-deployable mobile networks present a number of characteristics that distinguish them from classical cellular networks, scale-wise, architecture-wise, and even from a data usage perspective. We summarize in the following the main differences.

- **Scale:** self-deployable networks are relatively smaller, with fewer BSs and smaller coverage areas. They comprise only few interconnected BSs (generally less than 10), and serve a limited number of users. The required coverage area and the number of served users depend on the deployment cause and the scale of the situation in question. For example, the number of users of a network deployed in a football stadium is much larger than the one in a network deployed by first responders, in a disaster-hit area. In contrast, the coverage area in the first scenario is generally smaller than in the second.

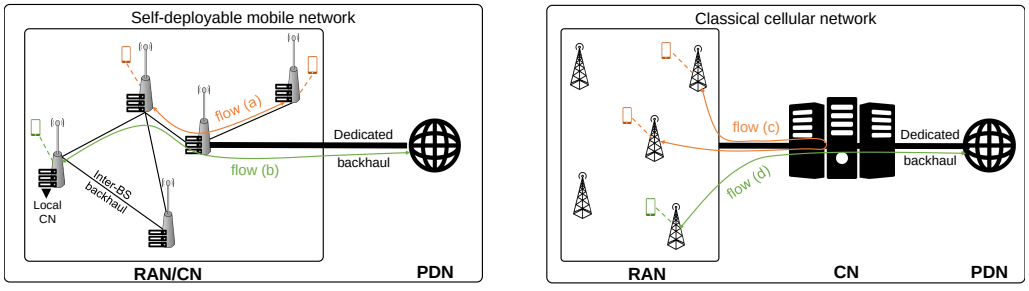


Figure 1.2: Comparison between self-deployable and classical cellular networks.

- **Architecture:** a comparison between the two network architectures is depicted in Fig. 1.2. In classical networks, the BSs in the RAN are dependent on several separate entities (physical or logical), usually located in the CN. Each BS has a dedicated backhaul link towards the CN, usually with an over-provisioned bandwidth. The split between the RAN and the CN is both functional and physical. However, in a self-deployable network, with the Local CN concept, all the core network functions are co-located with the BSs. The split, in this case, is only functional. The elimination of the physical split between the RAN and the CN further

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

enforces the BS autonomy and allows it to operate even when isolated, i.e., when there is no dedicated backhaul to an external CN [13].

- **Backhaul:** the backhaul network topology, as can be observed in Fig. 1.2, is inherently different in the two networks. With interconnected BSs, and the Local CN functions co-located with them, all data and signaling traffic, usually exchanged between each BS and the external CN, is now routed locally on the links interconnecting the BSs. Hence, those inter-BS links form the backhaul network. The backhaul is set up by the nodes in an ad-hoc manner, and can potentially represent a bottleneck, depending on the used technology, the available bandwidth and the inter-BS links quality [14].

- **Topology:** the rapid deployment of self-deployable networks, which in some cases is also unexpected, entails the skipping of the planning phase, which is recommended, or even mandatory, in classical networks. This time-consuming phase usually includes, among others, the BSs placement and the backhaul dimensioning. Therefore, in a self-deployable network, the topology is seemingly random, and can present several irregularities. Indeed, the BSs deployment can face some difficulties, depending on the deployment area, such as geographical obstacles (e.g., irregular terrains, large hills, water features), population dispersion (e.g., highly concentrated population centers, isolated locations), and site acquisition difficulties (i.e., unauthorized areas, toxic or dangerous environment) [15]. Furthermore, being easily transportable, the BSs in a self-deployable network can be mobile, even after deployment. This makes the network topology dynamic and prone to changes [14].

- **Traffic model:** self-deployable networks are usually deployed temporarily or following unexpected events. In most of the use cases cited earlier, the network is dedicated for professional usage by users in the same area, belonging to the same organization (e.g., first responders on site, security officers in public events). In such cases, the traffic transported by the network is mostly intra-network, i.e., users involved in a communication are generally attached to the same self-deployable network. The exchanged traffic, in this case, is handled by the Local CN. It remains local, and does not have to leave the network. An example is shown in Fig. 1.2, with the intra-network flow, denoted (a). Additionally, communication with external networks is also possible. The traffic, in this case, is inter-network and is routed towards the PDN via the dedicated backhaul (Fig. 1.2, flow (b)). In classical cellular networks, traffic is always routed to the external core network, whether the communication is between users that belong to the same network (Fig. 1.2, flow (c)) or not (Fig. 1.2, flow (d)). Thus, the routing on the backhaul, mostly irrelevant in classical networks with over-provisioned backhaul, is critical in self-deployable networks, since inter-BS links are shared among multiple flows.

  On the other hand, in some use cases, like public safety-oriented networks, the traffic has stringent requirements in terms of bit rate constraints. Users must be provided with a minimum required throughput, depending on the mission in question and the requested service. For example, for a live video streaming during a rescue operation, a certain quality of service (QoS) threshold must be respected.

1.2 Motivation

All of the aforementioned differences between classical cellular networks and self-deployable networks make the latter a worthy study item. We build on these differences to motivate and define the problems we address in this thesis. Indeed, on the one hand, the different characteristics

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

incite revisiting classical problems in cellular networks within the self-deployable context, such as resource allocation and user association. On the other hand, they raise a series of novel challenges, such as the Local CN placement within the network, the backhaul set-up, and BSs mobility. We summarize below the key properties we retain for self-deployable networks in the remainder of this study:

1. a limited network size (number of BSs and number of users);
2. BS(s) co-located with Local CN functions;
3. a potentially limited inter-BS backhaul forwarding data and signaling among BSs;
4. a random topology with fixed BSs (BS mobility is out of the scope of this work);
5. an abundance of intra-network communication;
6. a minimum guaranteed throughput requirement.

These properties constitute the driving idea behind the raised challenges, and the proposed contributions in this thesis. Indeed, we try to bring answers to the following unavoidable questions when configuring a self-deployable network. Starting with the RAN, for a random network topology, and a temporary (or unexpected) rapid deployment with no forecast of upcoming user density nor distribution, how is an efficient frequency and power allocation scheme set among the BSs in RAN? Once the RAN is configured, we move our focus to the Local CN. With an inter-BS backhaul of limited bandwidth, responsible of forwarding all the BSs traffic to the Local CN and vice-versa, how can backhaul saturation be avoided? How does the limited backhaul impact the Local CN functions placement, and where should the latter be placed to help relief the backhaul congestion? Once the RAN and the CN are configured, the network is operational, and the users start connecting. By which Local CN instance must each user be served? And, most importantly, to which BS each user must associate if all of the aforementioned properties of the architecture, the backhaul and the traffic model are to be taken into account?

1.3 Thesis outline and contributions

This thesis is organized as follows.

In Chapter 2, we present an overview of the most relevant concepts to this work. Following a technical summary on classical cellular networks, we present a literature review of the key elements that constitute self-deployable networks, and the related problems that have been mostly studied in the context of cellular networks. First, we discuss the achieved advancements and the remaining challenges concerning the conception of rapidly-deployable BSs, the virtualization of core network functions, and the development of deployment platforms. Then, we present the possible inter-BS wireless backhaul technologies, and the different spectrum allocation possibilities for self-deployable networks. Finally, we conduct a detailed review of the state of the art literature directly related to this thesis contributions, by covering: frequency and power allocation schemes among the BSs in cellular networks, the function placement problem in different network types and architectures, and the user association problem in cellular networks.

Afterwards, the thesis contributions on the configuration of self-deployable networks unfold as follows.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

- **Proposal of a frequency and power allocation algorithm**

  In Chapter 3, we propose an algorithm that outputs a frequency and power allocation scheme for a self-deployable network. The algorithm results in a power map, indicating with which power each BS transmits on each channel. This power map is computed offline, based solely on the network topology, and prior to the network operation, when no sufficient information on the user density or distribution is available. Then, it is fixed and followed by the BSs throughout their operation.

  We show that our proposed scheme significantly outperforms conventional equal power frequency reuse schemes, in terms of the achieved user throughput. Despite being static, the power map performs well, and is robust facing the variations in the number of concurrent users in the system and their distribution, and in the implemented network policies, such as the user association scheme.

  The algorithm consists of solving a non-convex, non-linear optimization problem, through multiple but simple transformations, based on signomial programming. The problem is first transformed into a complementary geometric programming problem. The complementary geometric program is then solved iteratively by turning it into a series of geometric programs, solved by further transforming them into convex problems. The same computation process is also used to solve the system-wide optimal frequency and power allocation problem, to which we compare our scheme.

- **Definition of a centrality metric for centralized Local CN placement**

  In Chapter 4, we tackle the Local CN placement problem in a self-deployable network, when the Local CN is centralized. In other words, we determine with which BS the Local CN functions must be co-located when a network of self-deployable BSs is created. To that end, we propose a novel centrality metric in a network, denoted as flow centrality. The flow centrality of a network node is defined as the maximum traffic that can be sent simultaneously by all the other nodes in the network towards this node, under certain capacity and load distribution constraints.

  We show that, in order to maximize the amount of exchanged traffic between the BSs and the Local CN, the latter should be co-located with the BS having the maximum flow centrality. We benchmark the proposed metric against different state of the art centrality metrics. We compute the potential loss in the total traffic received by the Local CN when the latter is not placed on the node with the maximum flow centrality. Moreover, we highlight some key properties of the Local CN placement, such as the tight dependence between the optimal placement and the backhaul bandwidth.

  The computation of the flow centrality is based on solving a linear optimization problem. Moreover, we propose a number of analytical expressions allowing the direct computation of the metric, in some particular network topologies.

- **Comparison between centralized and distributed Local CN placements**

  In Chapter 5, we compare a Local CN centralized placement with a distributed Local CN, from a backhaul bandwidth consumption perspective, by accounting for both data and signaling traffic.

  We show that distributing the Local CN routing functions on all the BSs is less costly from a backhaul point of view.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

We formulate the problems as mixed integer linear programming problems (MILP), that determine the optimal placement of the Local CN when the latter is centralized, and the number of instances and their placement when it is distributed. The objective is minimizing the backhaul bandwidth consumption caused by data and signaling traffic exchanged between the BSs and the Local CN.

- **Evaluation of backhaul-aware optimized user association and user attachment**

  In Chapter 5, we evaluate optimized user association (assigning a user to a BS) and user attachment (assigning a user to a Local CN instance), from a backhaul bandwidth consumption perspective.

  We show that a backhaul-aware association significantly reduces bandwidth consumption on the backhaul in comparison with classical RAN-based association schemes. Furthermore, we show that an optimal attachment of users to Local CN instances helps reduce backhaul bandwidth consumption if the signaling traffic is significant. Otherwise, the gain achieved by optimizing attachment is marginal.

  We formulate and solve the corresponding mixed integer quadratic programming problems (MIQP), returning the association and/or attachment with the objective of minimizing the backhaul bandwidth consumption.

- **Conception of a network-aware user association policy**

  In Chapter 6, we propose a novel user association policy adapted to self-deployable networks. This policy takes into account key network parameters, usually ignored in classical networks, such as the backhaul capacity, the available resources on the uplink, and users' requests, in addition to the downlink available resources. We propose an algorithm executed at each flow request initiated by a user. It first implements a flow admission procedure, followed by an association decision if the flow is accepted. The association objective is to maximize the remaining resources in the network by jointly accounting for the downlink, the uplink, and the backhaul. The goal is to keep the blocking due to insufficient resources as low as possible.

  We show that the proposed scheme significantly reduces the flow blocking probability in comparison to a traditional downlink-based association scheme. It adapts to different network constraints by mitigating bottlenecks in RAN-limited and/or backhaul-limited scenarios.

  The association problem can be formulated as a mixed-integer non-linear program (MINLP), solved at each flow request to return the corresponding user association. Nevertheless, due to the limited size of self-deployable networks, we rely on a quick and simple brute force computation in both the flow admission phase and the association decision, making the algorithm more practical. The evaluation is based on dynamic simulations under different parameter settings concerning the backhaul bandwidth, the data traffic model, and the number of concurrent users.

Finally, in Chapter 7, we draw the relevant conclusions from the above-mentioned contributions. We emphasize the possible directions to further develop the presented work, and shed light on the perspectives for future research.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 2

Challenges in Self-deployable Mobile Networks

2.1 Technical overview

2.1.1 Overview of cellular networks evolution

Throughout the different cellular network generations, the overall structure of cellular networks remained practically the same. It comprises two main elements: a radio access network and a core network. The RAN implements a radio access technology that allows users to get radio resources to connect to the cellular network. The CN handles network management and configuration, service provisioning, and ensures communication with third-party networks (e.g., the Internet). The RAN and the CN are connected via a reliable, cautiously dimensioned, backhaul. The radio access technology, the CN organization, the entities names, their specific functions, and the corresponding interfaces linking them differ from one generation to another.

The initial 2G networks, based on the Global System for Mobile communications (GSM), are voice-centric, and designed as a circuit-switched network, where a direct connection (circuit) between two communicating parties is established and reserved exclusively for their communication. The first step toward mobile Internet was the introduction of the General Packet Radio Service (GPRS), and the Enhanced Data rates for GSM Evolution (EDGE) standards, as a packet-switched addition to GSM networks. Mobile internet was pushed further when 3G was launched with the Universal Mobile Telecommunications System (UMTS), combining a circuit-switched voice network with a packet-switched data network. Data rate enhancements followed, referred to as High Speed Packet Access (HSPA). 4G started with the Long Term Evolution of UMTS (LTE). LTE is data-driven and pushed the data services expansion beyond GPRS, EDGE and UMTS. Indeed, by design, all services in LTE are packet-switched, even voice, completely letting go of the circuit-switched model [1].

2.1.2 4G cellular networks

We present below a summary of the 4G LTE standard, highlighting the key concepts we use in the upcoming chapters. The term LTE designates the evolution of the RAN, referred to as the evolved UMTS radio access network (E-UTRAN). Alongside LTE, the evolution of the network non-radio aspects is referred to as system architecture evolution (SAE). SAE encompasses the CN evolution into an all-IP network, referred to as the evolved packet core (EPC). The evolved packet system (EPS) is the umbrella that covers both LTE and SAE. The mobile device is referred to as the user

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

equipment (UE). Fig. 2.1 shows the EPS main network elements and interfaces.

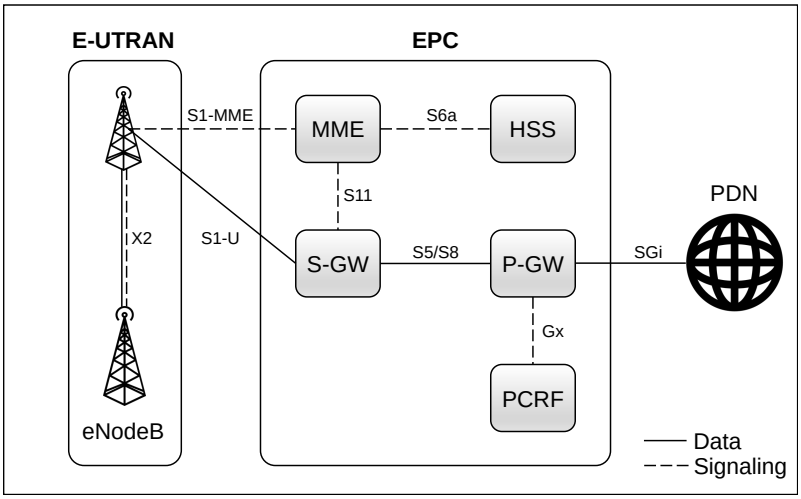


Figure 2.1: The EPS architecture.

2.1.2.1 The E-UTRAN

The E-UTRAN comprises a single equipment, the evolved NodeB (eNodeB), which is an autonomous unit that integrates functionalities previously ensured by centralized controllers managing the BSs. The eNodeB is responsible for the air interface, i.e., radio channel modulation/demodulation, channel coding/decoding, multiplexing/demultiplexing. Moreover, the eNodeB responsibilities include radio resource management, such as radio bearer control, radio admission control, scheduling, resource allocation, interference management, and security functions, such as data encryption over the air interface [16]. The eNodeB also ensures connectivity to the EPC via high-speed backhaul. The X2 interface provides communication between two eNodeBs for mobility management and interference coordination purposes [17].

2.1.2.2 The EPC

The all-IP EPC comprises several entities, namely the mobility management entity (MME), the serving gateway (S-GW), the packet data network gateway (P-GW), the home subscriber server (HSS), and the policy control and charging function (PCRF). We give a brief overview below of the main functions ensured by those entities [16].

- **MME:** the MME is the control node that manages the network and handles all the signaling between the UE and the EPC. Among the many services provided by the MME, we cite: S-GW selection, bearer management (establishment, maintenance, and release of data bearers), UE authentication when connecting to the network, mobility management, and inter-working with other networks.
- **S-GW:** all user IP packets are transferred through a S-GW. The S-GW acts as a router for user data, routed in data bearers between the eNodeB and the P-GW. In the case of an inter-eNodeB handover (i.e., a UE moves from one eNodeB to another), the S-GW acts as a mobility anchor of the connection and remains the same while the signaling and data paths are switched to the new eNodeB. Moreover, S-GW buffers downlink data destined to a UE in idle state until a successful paging and bearer re-establishment is completed by the MME.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

- **P-GW:** the P-GW is the network gateway to external PDNs, e.g., application servers or the Internet. It terminates data bearers started by the eNodeB. The P-GW allocates IP addresses to users. It also enforces QoS parameters, such as priority and throughput requirement, and handles flow-based charging according to rules that are dictated by the PCRF.
- **HSS:** the HSS comprises the database containing the UEs complete subscription information, such as their QoS profile, access restrictions, and the PDNs to which they can connect. Furthermore, it comprises the authentication center with the subscribers security keys, that generates the security vectors used for authentication.

Several data and signaling interfaces interconnect the different EPC entities. The main data interfaces are S1-U, between the eNodeB and the S-GW, and S5, between the S-GW and the P-GW. The main signaling interfaces are S1-MME, between the eNodeB and the MME, S11, between the MME and the S-GW, and S6a, between the MME and the HSS.

2.1.2.3 The air interface

The LTE air interface (Uu) uses orthogonal frequency division multiplexing (OFDM) as transmission scheme. OFDM splits the available bandwidth into multiple, narrower sub-carriers, and data is transported in parallel streams on the different sub-carriers, simultaneously, instead of spreading one stream over the complete carrier bandwidth. The sub-carriers are orthogonal to each other, which prevents interference among them [18].

LTE implements orthogonal frequency division multiple access (OFDMA), an extension of OFDM to multiuser communication systems, such that the different sub-carriers are used by different users, simultaneously [18]. The assignment of the sub-carriers to the users is the user scheduling procedure, and is handled by the eNodeB. As shown in Fig. 2.2, the OFDM symbols are grouped into physical resource blocks (PRB). The number of bits transmitted per OFDM symbol depends on the modulation and coding scheme (MCS). A PRB is a time-frequency resource, having a total size of 180 kHz in the frequency domain, and 0.5 ms in the time domain. The number of parallel PRBs (i.e., on different frequencies) depends on the system total bandwidth. The PRB is the smallest user assignment resource unit. Two PRBs form a sub-frame with a duration of 1 ms. A sub-frame represents the LTE scheduling time interval, which means that, each 1 ms, the eNodeB decides which users are to be scheduled and which PRBs are assigned to which user. The more PRBs a user gets and the more complex the modulation it uses, the higher the data rate the user receives. An LTE radio frame consists of 10 sub-frames.

2.1.3 Professional mobile radio networks

The aforementioned cellular networks are generally destined for commercial use by mass population. In parallel, there are PMR networks, which are professional mobile radio networks, also known as private mobile radio. These are private networks that serve a closed group of users, usually belonging to the same organization.

2.1.3.1 Legacy professional mobile radio

PMR networks consist of one or more interconnected base stations serving a number of special user terminals (commonly referred to as two-way hand-held radios). Numerous PMR standards exist, such as TETRAPOL [19] and TETRA [20], mostly used in Europe, and APCO P25 [21], mostly used in North America.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

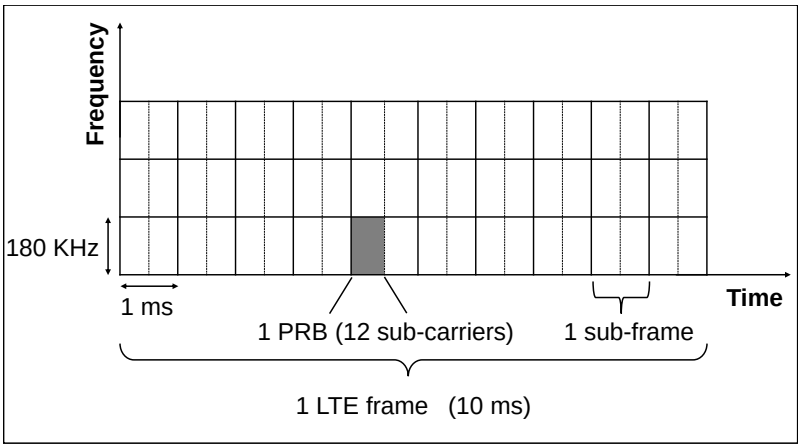


Figure 2.2: LTE frames.

PMR networks are of local or regional scale, destined for professional use, by organizations from a variety of sectors, such as: transportation (e.g., railways, bus companies, taxis, airport services), construction and public works, hospitals and medical facilities, electricity distribution companies, guarding and security services, military, public safety organizations (e.g., law enforcement, fire and emergency services).

The reason PMR is adopted by the above-mentioned sectors is the multitude of services and facilities provided by these networks, not always available in classical networks [22]. For instance, group communication is a key attribute of PMR, such that users can belong to closed groups and communicate within these groups. PMR allows point to multi-point communications, conversely to classical cellular networks which are limited to point to point. Moreover, push-to-talk is among the key services of PMR, in which pressing a single button opens communication on a radio frequency channel, and releasing the button releases the channel. This makes PMR call set-up times much shorter in comparison with classical cellular networks. Furthermore, PMR allows direct mode operation, in which terminals can communicate with one another directly when they are out of the coverage area of a network. Another feature is priority and preemption, such as the possibility of transmitting emergency signals, overriding any other communication taking place at the same time. In addition, communication within PMR is secure due to end-to-end encryption and air interface encryption [23].

On the downside, PMR networks are based on legacy standards implementing 2G technologies. They are mostly limited to voice-centric services, in addition to short messaging, and GPS, in some cases. Despite the devoted efforts to enhance the PMR standards by adding data-centric services, data rates remained low by modern standards. Indeed, recent enhanced TETRA standards only support data rates in the order of hundreds of kb/s [20], in comparison with 100 Mb/s in LTE.

2.1.3.2 Bringing PMR to LTE

Bringing PMR networks up to speed with LTE networks became a necessity, in order to provide a wider range of data-centric services (e.g., multimedia-based applications, video calls, image sharing) with high data rates, while preserving the PMR-specific features [23]. The choice of LTE as the next generation PMR resides in the need to align PMR networks with the capabilities of modern cellular networks. LTE supports high bandwidth, low latency, and high security data services, as well as real-time communication. Moreover, 4G traffic accounted for 69% of mobile

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

traffic in 2016 [24], with over 300 commercially launched networks globally by 2015 [25]. Using LTE allows benefiting from this large-scale deployment and the existing ecosystem, in order to reduce capital expenditure (CAPEX), i.e., infrastructure costs, as well as operational expenditures (OPEX) [25]. Furthermore, as the market for PMR systems is much smaller than that of commercial cellular, a significant gap has always existed in the investment and research that goes in PMR, which further contributed in slowing their advancement. By using the LTE architecture, PMR benefits from its massive innovation ecosystem. On the other hand, LTE commercial networks also benefit by expanding their markets to include new sectors, such as public safety agencies. In this context, notable efforts were made by industrials and standardization organizations to push for the adoption of LTE as the key technology of broadband public safety networks, which have been historically based on legacy PMR [22, 26]. The Third Generation Partnership Project (3GPP) is proposing several enhancements of LTE to include key PMR features, such as group communication [27], push-to-talk [28], and device to device communication [29], under the label "LTE for public safety".

2.1.3.3 The Fed4PMR project

Having a federated LTE-based PMR network is the goal of the Fed4PMR project, in which this thesis takes part [7][1]. Funded by the French government, and gathering a number of industrial and academic partners, the Fed4PMR consortium aims at putting in place a high data rate PMR/LTE system. The proposed solution will integrate, within the same infrastructure, different types of access networks:

- dedicated PMR/LTE Networks, independent of the commercial LTE network;
- a secured mobile virtual network operator (S-MVNO) integrating PMR services on existing public LTE infrastructure;
- legacy PMR networks;
- rapidly-deployable networks on demand.

In this work, we particularly focus on the last network type, consisting of rapidly deployable networks, to which we refer as self-deployable mobile networks. However, we do not limit the study to specific uses cases of PMR or LTE applications, since self-deployable networks impose themselves as a new family of cellular networks with a multitude of use cases. Instead, we tackle critical aspects of the configuration and the deployment of such networks, in general. In accordance with the Fed4PMR project, we adopt LTE as the base technology for self-deployable networks in the remainder of this thesis.

2.2 Self-deployable networks

As stated earlier, cellular networks have always relied on a hierarchical architecture, with a clear physical split between the RAN and the CN. Careful planning is usually needed before the network deployment. However, even when pre-planned, a network can sometimes fail to provide services

[1]The Fed4PMR project is part of the telecommunication sovereignty component of France's new industrial regeneration policy. It is partly funded by the French authorities through the Public Investment Bank (BPI) and the PIA investments for the future program. The consortium is lead by Thales, and comprises Air-Lynx, Archos, Expway, Ibelem, Silicom, Sysoco, and Pierre et Marie Curie University [30]. As part of the project, this thesis was conducted in close collaboration with Thales Communications & Security.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

to users. For example, the network can saturate due to a surge in demand in crowded events or emergency situations. The backhaul connectivity between the RAN and the CN can be lost. In a worst case scenario, in disaster-hit areas, the network infrastructure can be completely destroyed. In all of these cases, self-deployable networks, a novel concept in cellular networks, are promising and innovative solutions to guarantee service continuity for users.

Self-deployable networks can be rapidly-deployed and operated on demand, anywhere, anytime, in a zone where no network infrastructure exists in advance, or as a complement to an existing infrastructure (Fig. 1.1, p. 2). Such networks are autonomous, with minimum human intervention. They can be private, similarly to PMR, or public, depending on the use case.

The key element in a self-deployable network is the BS itself, which we call a self-deployable BS. Even with no backhaul connection to a classical CN, a self-deployable BS is capable of providing local services to users in its vicinity thanks to a Local CN. The Local CN is the set of virtualized CN functions, that provides the same functionalities as a traditional CN, and is co-located with the BS. The self-deployable BS can also have a dedicated backhaul to an external network. We overview self-deployable BSs and CN functions virtualization in the literature in Section 2.3. To cover larger areas, several BSs must interconnect, via a backhaul, to form a network. We shed light on the different deployment platforms of self-deployable networks, the achieved advancements, and the ensuing challenges, in Section 2.5. To establish interconnection among BSs, a multitude of wireless technologies have been reviewed in the literature as candidate solutions. We further discuss them in Section 2.6. Like any other network, self-deployable networks need a frequency band to operate on. The different possibilities in this regard are presented in Section 2.7. Once frequency resources are allocated to the network as a whole, the problem of resource allocation among the BSs in the RAN unfolds. We review in Section 2.8 resource allocation schemes in cellular networks, as a preliminary step to presenting a novel frequency and power allocation scheme for self-deployable networks in Chapter 3. With the RAN all setup, and the frequencies allocated, we move our focus to the Local CN. When multiple BSs are interconnected, they are all served by the Local CN functions. Hence, the Local CN placement problem arises. Before proposing a novel placement metric for a centralized Local CN in Chapter 4, and evaluating the distribution of the Local CN functions in Chapter 5, we survey the placement problems in different networks in Section 2.9. Finally, when the Local CN functions are placed, the self-deployable network is ready to operate. Users can start connecting to the network. Thus, users must select the BS to which they associate. We present a state of the art on the user association problem in cellular networks in Section 2.10, prior to our proposition of a novel association policy for self-deployable networks in Chapter 6.

2.3 Self-deployable base stations

Recent technology advances in microelectronics allowed reducing the size and weight of wireless network equipment [31]. On the other hand, the advances in virtualization techniques allowed easier virtualization of different network functions, including the mobile CN functions, such as the MME, S-GW, and P-GW [32]. The combination of those two concepts instigated the development of lightweight, easy to move, rapidly-deployable BSs, that can be co-located with virtualized CN functions, and application servers. The result is a stand-alone self-contained BS, with both RAN and CN capabilities. Such a BS is capable of autonomously providing network coverage and local services to users in its vicinity, depending on the hosted application servers, even without connectivity to external networks. Moreover, the BS can establish a backhaul connection to an external network, whether cellular (e.g., an existing LTE network) or not (e.g., Internet).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

In this context, the authors in [13] designed a software architecture and a set of protocols to co-locate a BS with virtualized EPC entities. This virtualization entails the deployment of the main EPC functions (MME, P-GW, S-GW), customized services, and resource management solutions locally at the BS, which reduces or completely eliminates its dependence on physical entities. This BS model can be fully standalone and provide local services to users within its coverage. Additionally, it can support physical core infrastructures.

The concept of a standalone BS was also defined as a requirement within the 3GPP standardization of LTE for public safety and, more specifically, in the isolated E-UTRAN operation for public safety (IOPS) standard [8, 12]. IOPS, that first appeared in 3GPP Release 13, envisions providing LTE connectivity and local mission-critical services to first responders in out-of-coverage areas and in emergency situations, when the network infrastructure is not accessible (e.g., saturated, damaged, or destroyed) [8]. An E-UTRAN is referred to as isolated when the eNodeB does not have a backhaul connection to an EPC. In this case, in order to provide services to users, IOPS requires that eNodeBs must be co-located with a Local EPC, offering the same basic functionalities as a traditional EPC.

2.4 Mobile core network virtualization

Virtualizing the CN functions has been an emerging topic in cellular networks in the past few years. Indeed, with the development of network function virtualization (NFV), mobile operators are increasingly interested in exploring the benefits of this technology to enhance their networks [33]. In a classical network architecture, the BSs depend on several functions of the CN. Typically, these functions are deployed on customized hardware, with proprietary OS, designed to meet their specific requirements. Such equipments are usually costly, which entails high deployment and operational costs, and increases time-to-market delays [32]. Moreover, it leads to a rigid and hardware-dependent infrastructure [13]. On the other hand, resiliency measures are needed to ensure the CN reliability, in order to avoid network failure if one of the CN entities fails. Installing, operating, monitoring, and replacing hardware entities are costly operations [34]. By virtualizing the CN functions, hardware and software are decoupled. Instead of running on dedicated specialized hardware, virtualization allows hosting CN functions as virtualized network functions (VNF), deployed on low-cost commodity servers. This decreases OPEX and CAPEX for operators, while accelerating service delivery, and adding scalability and flexibility to the network infrastructure [34].

While the concept of Local CN in self-deployable networks is also based on function virtualization, several differences exist with respect to an operator virtualized core. Indeed, as opposed to running them on off-the-shelf servers in distant data centers, the VNFs in a self-deployable network are co-located with the BSs in the Local CN. The idea behind the Local CN is to move the functions closer to the RAN, and lose the physical dependency. For instance, in case communication is lost between a BS and the CN functions at the core server, the former would not be able to operate properly, nor provide the needed services. Embedding the CN functions at the BS overcomes this setback, and improves the network resiliency [13]. On the other hand, with VNFs co-located with BSs, the network interconnecting these BSs constitutes the backhaul. Hence, different types of constraints are dealt with for the placement of these functions and their interconnection, mostly related to the backhaul network dimensioning, and its potentially limited bandwidth. Such problems are not necessarily relevant in an operator network. Furthermore, the served network scale is significantly different in the two cases. An operator virtualized core is usually designed at the scale of a country to serve a large number of customers, while in a self-

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

deployable network, both the coverage area and the number of users are limited. Consequently, the management and orchestration tasks of the VNFs in the former case are more challenging and crucial for an efficient network operation [32].

2.5 Deployment platforms

The need for self-deployable networks in a variety of use cases has led to the development of innovative deployment platforms, both terrestrial and aerial.

Indeed, self-deployable networks attracted industry players in the past few years, and several commercial products are already available on the market. For example, several telecommunication companies developed compact lightweight BSs, easily installed and configured. The end result is vehicle-mounted mobile BSs, co-located with a Local CN, and capable of providing LTE services on demand [5, 6, 35, 36]. Moreover, ultra-compact BSs that can be hand carried or worn as backpacks were also developed [5, 6]. Being easily transportable, such BSs are advertised as rapidly-deployable solutions to establish networks, anywhere, anytime, in disaster relief situations, emergency scenarios, and cruise ships. Destined for public safety personnel, military, or even civilians, they allow a quick and efficient establishment of network services.

Aerial platforms providing wireless communication are also under development. They consist of a single or multiple unmanned aerial vehicle (UAVs), which are entities that fly in an autonomous manner, or while operated remotely, with no human personnel on board [37]. UAVs include, among others, drones, aircrafts, and helikites. Mounting self-deployable BSs on interconnected UAVs is one of the envisioned deployment platforms of self-deployable networks. Being rapidly deployable, easily movable, and having the ability to self-organize and to support direct line of sight communications with terrestrial equipments, aerial platforms garnered attention in both industry and academia [38]. Depending on the altitude of UAVs, they are classified as high altitude platforms (HAPs) (17-25 km) or low altitude platforms (LAPs) (0.3-4 km) [38].

HAPs are typically deployed for long-term applications, such as supporting broadband mobile access for a long duration. The most recent HAP example is the Loon project, developed by Google X, aiming to provide internet coverage to remote geographical areas [10]. The main components of a BS were redesigned to be light enough to be carried by a balloon, powered entirely by renewable energy, and launched 20 km up in the stratosphere. The BSs carried by the autonomous balloons are able to connect to ground BSs, and relay the signal down to users in remote areas. Each balloon can have a coverage area of 5000 km^2, and a lifetime of several months. A network of interconnected balloons is set-up via free-space optical communication.

On the other hand, LAPs are usually associated with temporary deployments, such as during disaster and emergency situations, to extend or replace an existing communication infrastructure, or during crowded events, to avoid network saturation [38]. Drones equipped with BSs facilitate network deployment in scenarios where a terrestrial solution is not possible. For example, when the deployment area is hard to reach or deemed dangerous for humans due to toxic materials or severe environmental conditions. Drones hovering over the targeted area overcome such hurdles. Furthermore, terrestrial deployments are limited by building heights and line of sight obstacles, which reduces their coverage areas, and prompts the need for denser deployments. A drone, due to its higher position, can provide much larger coverage and line of sight is mostly assured.

Nonetheless, drone networks create design, operation, and management challenges, all still a work in progress [39]. The energy challenge is a critical one for battery-powered drones. Long-time hovering, data processing, wireless transmission, and long endurance are some of the main power consuming attributes of a drone, besides flying [40]. In this context, propositions include

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

parking drones at charging stations on top of buildings or lamp-posts, when needed [40], and developing gas or solar powered drones allowing extended lifetime. Alternative UAV designs based on helikites or balloons consume relatively lesser energy due to simpler flying mechanics [41]. Another critical challenge is maintaining a wireless backhaul interconnecting the drones. Rapid channel variations due to mobility, maintaining interconnectivity despite potential topology changes, and the impact of weather conditions are all crucial factors to take into account [39]. The positioning of the individual drones and of the network as a whole is another challenging problem for drone networks [40]. Indeed, the positioning can have a significant impact on the user throughput [40]. The location, the altitude, and the velocity of the drones must take into consideration the location, the demand, and the velocity of the users.

To date, all UAV systems have some kind of connection to one or several ground entities, whether other BSs, a ground control system, or a CN. They are equipped with on-board communication gateways, such as WiFi access points or cellular femtocells [38]. To the best of our knowledge, standalone UAV systems capable of providing services to users while fully autonomous, similarly to terrestrial self-deployable networks, are yet to be developed. The additional challenge, in this case, is to add the Local CN functions and the needed application servers to the already burdened drone payload. This necessitates further investigation of the drone size, the take-off weight, and the design of light-weight and low-energy suitable equipments.

2.6 Wireless backhaul

The backhaul network in self-deployable networks has a particular architecture: it consists of the wireless links interconnecting the BSs. Indeed, not all BSs in a self-deployable network have a dedicated backhaul connection to a traditional CN. However, in order to be able to serve users, all the BSs must be able to reach the Local CN, or reach the BS with the dedicated backhaul link towards an external PDN. This is made possible by the links interconnecting the BSs. Those inter-BS links are responsible of forwarding all the data and signaling traffic exchanged between the BSs and the Local CN.

Wirelessly interconnecting BSs is not new to mobile networks. In LTE, eNodeBs can communicate directly with each other via the X2 interface (Fig. 2.1). However, this interface is conceived and standardized for two purposes: mobility management and interference coordination. It does not mimic or replace the backhaul [17].

In heterogeneous networks (HetNets), in order to reach the CN, small cells are usually connected to a macro BS that has a dedicated backhaul link towards the CN (e.g., fiber) [42]. That is, each small cell has a backhaul link to the macro BS (star topology), or small cells are interconnected, at least one having a link to the macro BS (tree topology) [14]. For outdoor small cells, the backhaul is usually wireless. A number of wireless technologies, whether in-band or out-of-band, have been studied for small cells backhauling (e.g., microwave, millimeter-wave, WiFi) [42], all of which we discuss in the following as potential backhaul technologies for self-deployable networks.

The support for wireless backhaul and relay links is one of the attributes envisioned for the 5G architecture [43]. Referred to as integrated access and backhaul (IAB), the goal of this architecture is to enable flexible and dense deployments without having to proportionately densify the network. IAB targets several deployment scenarios, such as indoors, outdoor small cells, and mobile relays (e.g., trains) [43]. Although the wireless backhaul technology used to interconnect BSs is yet to be determined, the standard supports both in-band and out-of-band backhauling. IAB is still in its early standardization phase, where design guidelines and requirements are developed [43].

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

2.6.1 In-band backhaul

With in-band backhauling, the access and backhaul links are multiplexed on the same frequency band, and may overlap. In this case, tighter coordination is required between the access and backhaul links, focused on interference avoidance or mitigation schemes, and efficient resource scheduling.

With many GHz of spectrum to offer, millimeter wave (mmWave) is the front runner wireless technology for radio access and backhauling in 5G networks [44]. The availability of large bandwidth at mmWave frequencies makes it particularly attractive: it allows providing high data rates to users, on the one hand, and backhaul capacity to operators, on the other hand. Recent studies focused on proving the feasibility of in-band mmWave-based backhaul [44, 45], and its advantages in comparison with the typical out-of-band wireless backhaul based on microwave links [45]. Scheduling mechanisms for inter-BS communication, and the multiplexing between radio access and backhaul links are also being investigated [44]. On the down side, mmWave has a reduced communication range [45].

Prior to the 5G mmWave trend, LTE in-band solutions were proposed. Among those solutions, the authors in [46] proposed connecting neighboring eNodeBs via special enhanced UEs capable of associating with multiple eNodeBs simultaneously. A UE acts as an intermediate node between the two eNodeBs it is associated to, forwarding traffic between them. However, this solution requires new types of UEs with the aforementioned capabilities, and a high UE density in the network to ensure connectivity among the different eNodeBs. Furthermore, this scheme raises questions on the energy efficiency of the UEs burdened with the task of interconnecting the eNodeBs. In another in-band solution, the authors in [14] introduced an enhanced eNodeB design (e2NB), consisting of a physical eNodeB extended with several UE stacks. By depicting the behavior of a physical UE, the UE stack allows the e2NB to act as a virtual UE. Hence, it discovers neighboring e2NBs and associates with them, similarly to a standard UE, establishing in-band inter-eNodeB links. Nevertheless, since the procedure is identical to a typical UE-eNodeB association, the two e2NBs must be at proximity. This further restricts the approach to nearby e2NBs, and requires investigation of the resulting interference.

2.6.2 Out-of-band backhaul

With out-of-band backhauling, access and backhaul resources are independent. The in-band bandwidth sharing constraints and limitations are relaxed. Having inter-BS WiFi connectivity is one possibility. However, this requires additional dedicated equipment and antennas, which increases the deployment costs [14]. Moreover, the industrial, scientific, and medical bands, on which WiFi works, may cause more interference in comparison with the cellular licensed bands. WiFi was also shown to increase delay in comparison with LTE in commercial networks [47]. Using satellite links to establish the backhaul is another possibility. Nevertheless, satellite links are costly, and could suffer from high latency [48]. Satellite is preferably used to connect the network to an external PDN, rather than to establish inter-BS wireless connectivity. Having point-to-point or point-to-multipoint radio frequency links are amongst the best solutions for inter-BS connectivity, and the most adopted today to establish wireless backhaul [14]. While they do not cause high latency problems, they do require careful network planning, being based on line-of-sight wireless connectivity [14].

The design of inter-BS links is out of the scope of our work. To avoid the critical implications of in-band backhauling solutions, requiring tight coordination between access and backhaul networks, we adopt, henceforth, an out-of-band backhauling. That is, the backhaul and the radio access links resources are independent. Furthermore, we do not limit our study to a specific wireless

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

backhaul technology. Without loss of generality, we consider that, regardless of the wireless technology used, there is no contention among the inter-BS backhaul links for resource utilization. We assume that interfering wireless links are operating on distinct channels, allowing parallel transmissions on the different links, with no interference [49]. Moreover, we suppose that the wireless backhaul bandwidth can be limited. In that case, the amount of traffic that can be routed on the inter-BS links is limited.

2.7 Spectrum allocation

The first and mandatory requirement to operate a self-deployable network, like any other wireless network, is to determine the spectrum bands it is allowed to use. While we do not adopt a particular band in this work, we present in the following the different possibilities.

2.7.1 Dedicated spectrum

Having dedicated spectrum, reserved for self-deployable networks, is the desired solution to avoid interference with other networks in the vicinity. In this context, the adoption of LTE as the broadband technology for public safety applications warranted regulatory agencies to allocate dedicated spectrum bands for LTE-based public safety networks within the 400 MHz and 700 MHz bands (e.g., United States [50], France [51, 52], Europe [53]). Thus, the networks deployed temporarily in disaster-hit areas and emergency situations, by public safety agencies, can benefit from this dedicated spectrum.

On the other hand, as mentioned in Section 2.1.3.2, self-deployable networks are one of the envisioned deployment strategies of broadband PMR networks. Dedicated spectrum bands are usually allocated for PMR networks. With the development of broadband PMR networks, new spectrum bands, suitable for broadband services, are needed. In France, for example, 40 MHz of the 2.6 GHz band were recently allocated to broadband PMR [54], from which self-deployable networks can benefit.

2.7.2 Cognitive radio technology

In case dedicated spectrum is not available, cognitive radio technology is as a viable candidate, in accordance with the needed flexibility in self-deployable networks [55, 56]. The concept behind cognitive radio is to exploit under-utilized spectral resources in an opportunistic manner, by detecting unused spectrum and making use of it. In their work on aerial self-deployable networks, the authors in [41] recommended the use of dynamic spectrum sharing, in the absence of dedicated spectrum. BSs learn about the radio environment, and create a radio map of their surroundings. This task involves spectrum sensing and localization of nearby radio users. Cognitive extensions are added to the BSs with sensing functionalities to obtain spectrum occupancy thresholds. On the other hand, a radio environment map is output by a database storing and processing information about the radio environment status, based on spectrum sensing, predicted propagation models, and knowledge of previous spectrum allocation, for example [55]. The outcomes of both techniques can be combined to determine the unoccupied and ready-to-use radio resources [41].

2.8 Frequency and power allocation

Due to the scarcity of available spectrum in licensed bands, the frequency allocation problem is a recurrent one in cellular networks [57], that extends to self-deployable networks. The challenge is

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

to use (and reuse) the available frequencies in a way that guarantees service to the largest number of users, while avoiding the harmful effects of the resulting interference [58].

In general, there are two major classes of interference: intra-cell and inter-cell. Intra-cell interference takes place between frequency channels within the same cell, due to their adjacency [59]. Inter-cell interference (ICI) is caused by the simultaneous use of one frequency channel in multiple cells [59]. The use of OFDMA on the downlink radio interface, such as in LTE, eliminates intra-cell interference, since data is transmitted over orthogonal sub-carriers [18]. In this case, the primary source of interference in a network is ICI. It affects the signal to interference and noise ratio (SINR) of active users, especially those near the edge of a cell, causing significant degradation in their throughput [58].

Inter-cell interference coordination (ICIC) mechanisms aim at reducing interference, in order to increase the overall data rates of the cell users [60]. There are two types of ICIC techniques: mitigation and avoidance.

Interference mitigation techniques reduce the impact of interference during transmission or after the signal reception. Such techniques include, among others, as cited in [59]: (i) interference cancellation by detecting and subtracting interference signals from the received signal, (ii) selecting the signal with the best quality when various signals are received in multiple antenna systems, (iii) dynamically changing the antenna radiation pattern based on interference levels and (iv) frequency hopping.

Interference avoidance techniques rely on a resource allocation planning, which determines how frequency and time resources, and power levels are allocated among the different BSs and users [60]. A multitude of interference avoidance techniques on the downlink (DL) have been proposed and evaluated in the literature. They can be classified into two branches: static reuse schemes, and dynamic coordination schemes. A survey on uplink (UL) resource allocation schemes is presented in [61]. Since our contribution in Chapter 3 only concerns resource allocation on the DL, we exclusively focus, in the following, on the DL schemes.

2.8.1 Static reuse schemes

The most basic approach is conventional frequency reuse, with a reuse factor r. It consists of partitioning the available bandwidth into r non-overlapping sets of channels, such that the channels of each set are re-used by sufficiently-distanced BSs, and transmission power is equal on all channels. Full frequency reuse is the trivial case with a reuse factor $r = 1$, also referred to as equal power/reuse 1 scheme (EP-R1). In this case, all the spectrum is used by all the BSs, as shown in Fig. 2.3.

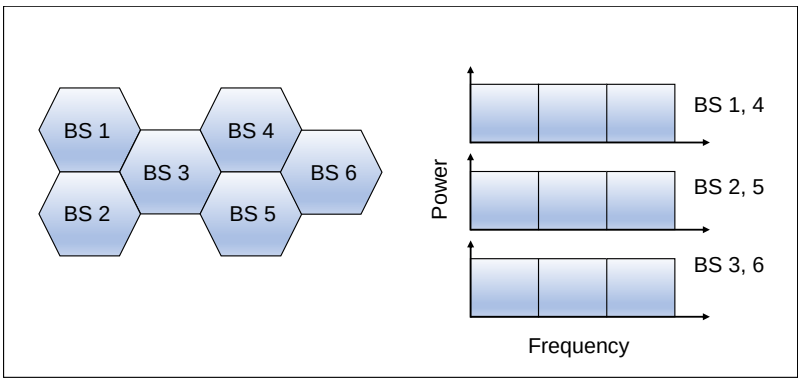


Figure 2.3: Equal power/reuse 1 scheme (EP-R1).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

While EP-R1 achieves a better overall throughput in the network, it leads to severe interference, especially for cell-edge users [62]. With a reuse factor $r > 1$, interference on cell edge is reduced. Nevertheless, spectrum efficiency is also reduced, since the usage of each BS is limited to $1/r$ of the available spectrum. This causes significant limitations in the maximum achievable data rate [63]. A typical scheme in cellular networks is to have $r = 3$, referred to as equal power/reuse 3 (EP-R3), shown in Fig. 2.4.

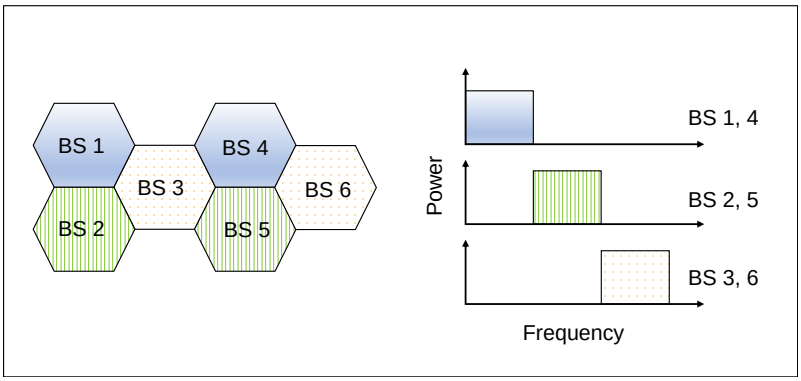


Figure 2.4: Equal power/reuse 3 scheme (EP-R3).

To avoid the limitations of EP-R1 and EP-R3, fractional frequency reuse (FFR) schemes were introduced. There are two FFR deployment modes: partial frequency reuse (PFR) and soft frequency reuse (SFR).

In PFR [64], a set of channels are re-used by all BSs and allocated to users in the cell-center area only (i.e., with a frequency reuse factor $r = 1$). The remaining channels, allocated to users in cell-edge areas, are partitioned across BSs based on a reuse factor r, similarly to a conventional frequency reuse scheme. An example is shown for $r = 3$ in Fig. 2.5. PFR lowers ICI while slightly improving spectrum efficiency with respect to a conventional reuse [65]. Nevertheless, with a strict no spectrum sharing between edge-users and center-users, the spectrum can still be under-utilized in some cases [66].

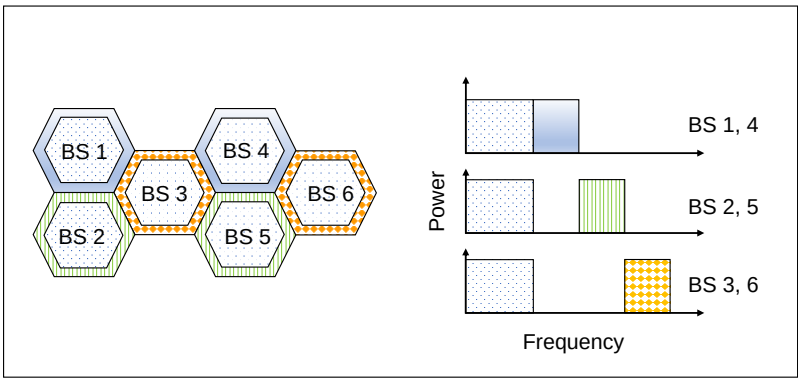


Figure 2.5: Partial frequency reuse scheme (PFR).

SFR has been proposed in order to add flexibility to the PFR scheme [67]. SFR introduces BS transmission power levels tuning as a mean to reduce ICI. All BSs use all the available spectrum, but with different power levels on different partitions of the spectrum. While transmission power is high on some channels (usually reserved to cell-edge users), it is lower on the others (usually

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

reserved to cell-center users). An example of SFR is shown in Fig. 2.6. The fact that each BS utilizes the entire spectrum eliminates the loss in spectral efficiency, while improving the system performance [68]. The advantages of SFR in achieving better throughput have been already demonstrated, mostly in hexagonal-grid networks [63, 67–70]. Moreover, the gains expected from SFR in large-scale networks with irregular cell patterns were studied in [71], an analytical framework to evaluate the coverage of SFR systems in a random network deployment was proposed in [72], and a modified version of SFR adapted to HetNets was presented in [73].

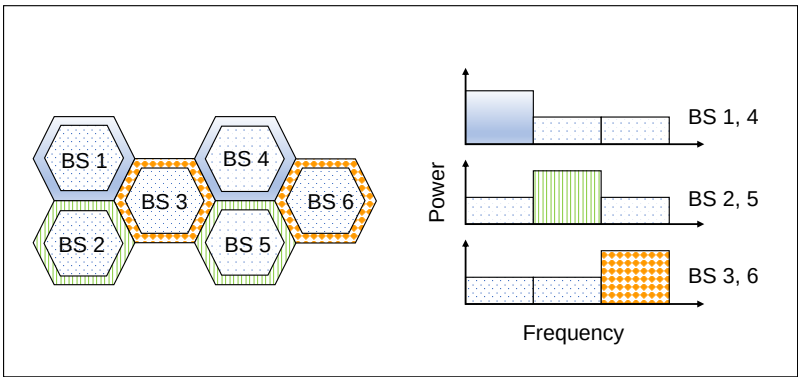


Figure 2.6: Soft frequency reuse scheme (SFR).

All the above-mentioned schemes are static, with the set of channels and the power levels allocated to each BS fixed in advance. The planning is done a priori, with no regard to the upcoming, and usually varying, user distribution and BS traffic load. It has been argued that the lack of adapting to such dynamic changes can lead to degraded performance [62, 74]. In this context, a load balancing algorithm that can be executed by each BS autonomously, and that does not require reconfiguration of the initial frequency reuse scheme, was proposed in [75]. In this scheme, users are redistributed between overloaded sub-bands and lightly loaded ones to balance the loads among cell-center and cell-edge, in an effort to adapt to the varying user distributions. On the other hand, dynamic interference avoidance techniques were developed.

2.8.2 Dynamic coordination schemes

Various dynamic resource allocation schemes were proposed in the literature, where BSs are required to coordinate and exchange information. Those schemes can be classified as centralized, semi-distributed, or fully distributed, depending on the level of needed coordination [59].

In centralized dynamic schemes, a central unit collects the channel state information of all UEs in the network from each BS, determines the channel allocation based on a defined objective function, then sends the allocation information back to the BSs. Such a process generates significant communication overhead, and risks causing severe delays [76]. Indeed, the optimization problem, which is a global scheduling problem that determines which user transmits on which specific PRB and with which power, is very difficult to solve [76].

In an effort to reduce the amount of exchanged information, semi-distributed schemes were proposed [77, 78]. In these schemes, a central entity also controls a number of BSs, but it only allocates resources in bulks to each BS, without determining the allocation scheme of each user. Each BS is then responsible of allocating the channels to its users on each frame. With the allocation problem distributed between the central entity and the BSs, its computational complexity is reduced. Nevertheless, the communication overhead between a BS and the central entity can still

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

be considerable [59].

Several fully distributed schemes were proposed, in which a central entity is not required [74, 79]. In this case, coordination is needed only among the BSs. Using local information collected from its users, and the exchanged information from other BSs, resource allocation is done by the BS itself. For example, the authors in [63] proposed an adaptive SFR scheme that dynamically adapts power levels to changing traffic load and user distribution, based on exchanged information among neighboring eNodeBs via the X2 interface. Fully distributed schemes are more suitable for practical implementation, due to the reduced computation complexity and communication overhead. However, inter-BS communication latency may still cause notable delays, questioning the reactivity of such schemes and their ability to adapt to varying conditions in real time [80].

2.8.3 Addressed challenges

To summarize, on one hand, static allocation schemes are easier to implement in practice, and do not require any costly coordination among BSs. On the other hand, dynamic schemes adapt to dynamic changes in the network, which is crucial to avoid degraded performance. Therefore, we propose, in Chapter 3, a practical approach that combines the advantages of both schemes. Our proposed scheme is static, with no coordination among BSs, nor real-time computation. Frequency allocation is based on a modified version of SFR, in which all BSs use the entirety of the available spectrum, divided into multiple sub-bands, and each BS can transmit on each sub-band with a different power level. We propose an algorithm that determines the different power levels on the different sub-bands for each BS, prior to the network operation, given solely a network topology, and with no a priori information on the traffic load and user distribution. Nevertheless, the resulting allocation scheme is robust facing the varying network conditions, without needing continuous adjustments as in dynamic schemes. Furthermore, it adapts well to random network topologies with irregular BS placement, making it suitable for self-deployable networks.

2.9 Local core network functions placement

As stated earlier, the Local CN is the entity that provides the basic functionalities of a traditional CN. It must at least include the functionalities of the MME, the S-GW, the P-GW, and the HSS, if we adopt the 4G EPC nomenclature. In an EPC, each function is usually deployed on a separate dedicated hardware. In a Local CN, these functions are virtualized and deployed as software on the BS itself. The Local CN supports all bearer services and can also host application servers.

Saying that a self-deployable network is served by only one Local CN is another way of saying that the network is operated by only one operator. The concept of a *single* Local CN here is an analogy to the single core network serving a classical operator network. It does not entail that the Local CN cannot comprise multiple instances of the same function, similarly to how an EPC, for example, comprises a pool of MMEs and a pool of S-GWs, all belonging to the same EPC. Hence, one or several instances of the same function can exist in the same network.

This raises two correlated problems: (1) the Local CN dimensioning problem, i.e., how many instances of each function do we need, and (2) the Local CN placement problem, i.e., where do we place each function.

In the first problem, the Local CN dimensioning depends on several factors, including the network size, the network load, the processing/buffering capacities of the hardware hosting the functions, and their ability to provide the necessary services to all users in a timely manner. Considering the self-deployable network limited size, these factors are not necessarily relevant. Another

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

key factor is the impact of the number of functions on the resulting signaling traffic exchanged among these functions. A detailed assessment in this regard is presented in Chapter 5.

For the placement problem, two options unfold: a centralized Local CN, and a distributed Local CN. Indeed, it is possible to co-locate all the function with only one BS, or to have them distributed in the network, such that functions are co-located with different BSs. In both cases, the BS(s) selected to host the Local CN functions should be determined. All traffic (data and signaling) exchanged between a BS and the Local CN functions is routed locally on the backhaul links interconnecting the BSs. When their bandwidth is limited, these links can represent a bottleneck, by limiting the amount of traffic that can be exchanged between the BS and the Local CN. Hence, the backhaul must be carefully considered when placing the Local CN functions.

The placement problem appears in the literature under different forms and in different networks, such as: placing VNFs [81], placing one or several gateways in wireless mesh networks (WMN) [82], and placing the sink in a wireless sensor network (WSN) [83]. Despite some similarities with the Local CN functions placement, each of those problems is distinct. Indeed, the function to be placed differs from one problem to another. Having a defined set of tasks and different constraints to meet impacts the problem objectives in each case. In the following, we give an overview of those placement problems, and highlight the main similarities, or the lack thereof, with the Local CN placement.

2.9.1 Placement of virtualized network functions

As discussed earlier, network operators are opting for running core network functions as VNFs on commodity hardware. For a successful mobile core network virtualization, an optimal placement of the VNFs is crucial, whether across federated clouds (i.e., geographically distributed and interconnected data centers) [84] or within the same data center [32]. In addition to the significantly different scale with respect to a local CN, the objectives of the placement problem in such networks are also different. In fact, with a virtualized core network on a federated cloud, operators can create VNFs on-demand, decide on where to place them during runtime, and dynamically dimension and re-plan the network according to users behavior, the offered services, or network performance. The placement problem in this case is more like a runtime management problem. For example, the authors in [84] proposed a VNF placement, for virtual S-GWs in particular, focused on minimizing the frequency of S-GWs relocations. Placing the VNFs within the same data center has also been studied. The focus, however, is on minimizing computing and networking resources consumption within the data center [32].

Placing VNFs is a general problem, and not particularly limited to core network virtualization. In any network where traffic may be required to pass through a certain VNF, the placement of the VNF can be analogous to the Local CN placement problem [81]. In this context, the authors in [85] formulated the VNF placement problem as a cost minimization problem, and proposed a placement metric inspired from centrality metrics in networks in general, and the betweenness centrality in particular [86]. Their approach was on a per-flow basis, placing the VNF as a middle point between a source node and a destination of each flow. While a Local CN is either a source or a destination node for all the network flows, using centrality metrics to place the Local CN in the network is a noteworthy study item, that we look into in details in Chapter 4.

2.9.2 Placement in wireless mesh networks

In WMNs, each node has the ability to forward traffic of other nodes [82]. In other words, each node can operate as both a host and a router. Gateway functionalities allow the integration of

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

WMNs with other networks, such as cellular, WiFi, or a wired backbone. Hence, all the traffic in WMNs flows to or from a gateway. One or several gateways may exist in the network. For example, in a wireless neighborhood network, houses are equipped with antennas that form a multi-hop WMN that routes traffic to the Internet. To that end, a few Internet Transit Access Points, serving as gateways to the Internet, are deployed across the neighborhood, and their placement must be determined [87].

The efficient placement of gateways in WMNs is critical, seeing its impact on the network management, performance and resource usage [49]. There have been numerous studies addressing this problem in multi-hop WMNs [49, 87–89]. Most of the studies consider that multiple gateways need to be placed in the network. Different optimization problems are formulated, with the main objective of minimizing the number of deployed gateways, while optimizing their placement. Constraints are set on the envisioned QoS metrics, such as: guaranteeing nodes bandwidth requirements (e.g., [49,87]), or satisfying delay constraints (e.g., [49,88,89]). A use case closer to the placement of a centralized Local CN appears in [90], in which the placement of a single gateway in a WMN is optimized, with the objective of maximizing the minimum network throughput. A heuristic, hereafter referred to as the weighted path, was proposed to achieve this. It uses a shortest path algorithm to compute the minimum weighted path from any node to one destination node. The link weight represents the likelihood of transmission failure on that link. Then, the gateway is placed on the destination node that minimizes the sum of the weights of all the shortest paths from all the nodes to that destination. Since it focuses on the links interconnecting the nodes, a key criteria in self-deployable networks, this weighted path heuristic seems well adapted to determine the placement of a centralized Local CN, and will be evaluated afterwards and compared to our proposed metric in Chapter 4.

2.9.3 Placement in wireless sensor networks

In WSNs, sensors collect data and transmit it to a destination node, referred to as the sink [83]. The placement problem in WSNs focuses on placing the sink within the network [91]. An analogy could be made between the sink, receiving all the data generated by sensor nodes in a WSN, and the Local CN, receiving all the traffic generated by BSs in a mobile network. However, in most cases, the two problems scale differently. Indeed, the number of sensor nodes in a WSN significantly exceeds the number of BSs in a self-deployable mobile network, and the nodes usually generate less data, with lower frequency. Moreover, due to tight energy constraints in WSNs, sink placement aims at minimizing the energy consumption, and extending network lifetime [91]. These constraints are not a priority in our Local CN placement problem. Furthermore, while our model is based on the orthogonality of resources used by the inter-BS links, the same does not apply in a WSN, resulting in a different problem model.

2.9.4 Addressed challenges

In our work, we study both centralized and distributed Local CN placements, in Chapter 4 and Chapter 5, respectively. For both approaches, the driving idea is the limited backhaul bandwidth. For the centralized placement, building on the shortages of state of the art metrics, we propose a novel placement metric adapted to self-deployable networks. For the distributed placement, we study the number of needed functions and their optimal placement, taking into account the data and signaling traffic exchanged within the network, and the backhaul bandwidth consumption.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

2.10 User association

User association, i.e., determining which BS serves each UE, is a prevalent problem in cellular networks. Selecting the serving BS can have significant impact on the network performance and the user quality of experience (QoE). Indeed, the maximum achievable rate at the physical layer, and hence, the user throughput, depend on the serving BS and on the number of users associated to it [92]. Moreover, user association sets the load distribution in the network, which has consequences on efficient radio resource utilization, and can risk degrading performance, or even BS saturation, when overlooked [93].

In classical networks, the most adopted user association policy is based on the received power on the DL, such that a user associates to the BS providing the strongest signal strength [94]. The association rule mainly adopted in classical networks is to select the BS from which the received SINR is the highest, referred to as best SINR. This simple, yet widely accepted rule, suffers from several shortages [93–95], and has been generally deemed ineffective for several network architectures. Indeed, it was mainly criticized for its poor load balancing [93], and for the fact that it solely accounts for the DL while ignoring the UL [96, 97]. Moreover, best SINR was ruled out as a suitable association policy in HetNets [98], backhaul-limited networks [99], and 5G networks [94]. To overcome these limitations, numerous association policies have been proposed in the literature, as discussed in the following.

In an effort to define the challenges of user association in 5G networks, the problem was extensively surveyed in [94], with a focus on 5G driving technologies, such as massive MIMO, mmWave, and device-to-device communication [100]. The survey highlights the substantial research efforts needed in those areas to design association schemes adapted to 5G networks. Another important trend in 5G making user association more challenging is having ultra-dense networks (UDN), where the number of available small cells is very high [101]. In this context, the authors in [102] studied joint user association and a time division duplex (TDD) allocation in UDN. They focused on the benefits of dynamic TDD systems, where the percentage of resources used for DL and UL is variable.

Based on our review of the user association problem in the literature, we observe that most of the works formulate an offline optimization problem, with a utility-based objective function targeting a network performance parameter, such as network throughput, spectrum efficiency or energy efficiency. Then, the core contribution of these works focuses on *how* the formulated problem is solved, i.e., the required transformations and the proposed heuristics that try to solve the rather complicated problem. In contrast, for our contribution in Chapter 6, we evaluate the performance of our proposed association scheme in dynamic network scenarios.

2.10.1 User association in HetNets

A considerable amount of the research efforts on user association were dedicated to HetNets. The reason why best SINR is not suitable for HetNets is the significant difference in transmission powers between macro BSs and small cells, which causes more users to associate to the macro and creates a strong load imbalance among BSs [103].

As a solution, an association referred to as range expansion was proposed [104], such that a bias is added to the reference signal power received from small cells to "virtually" extend their coverage. It was proved that such a scheme improves throughput and load balancing, with respect to best SINR, since more users associate to less loaded BSs in comparison to populated BSs with higher transmission power [98]. Several other objectives were targeted by user association rules for HetNets. With the aim of achieving both load balancing and user fairness, the authors in [105]

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

proposed a distributed algorithm that maximizes a rate-based utility function, defined as a logarithmic function of the user data rate. Focusing on spectrum efficiency, several studies considered problems with joint objectives besides user association. For example, a joint user association and channel allocation optimization was proposed in [92], with the objective of maximizing the minimum data rate. This work was extended in [106] to include transmission coordination, in addition to the previous objectives, with the objective of maximizing a rate-based utility function. Another targeted objective in HetNets is optimizing energy efficiency. In [107], for example, a user association optimization problem was proposed, aiming to maximize the network energy efficiency. The latter is represented by the total number of successfully transmitted bits by all the UEs divided by the total energy consumption.

The authors in [93] proposed a distributed association framework focused on load balancing under inhomogeneous traffic distributions. The defined optimization problem supports different objectives, such as optimizing throughput, delay, or load-balancing.

2.10.2 Uplink-aware user association

Most studies on user association are mainly focused on the DL. The UL has been usually overlooked, since network traffic used to be mostly asymmetric, such that the required throughput on the DL significantly surpassed what is required on the UL. Nevertheless, UL traffic is gaining importance due to trending symmetric traffic applications (e.g., video calls, real-time video gaming). Furthermore, on a system level, there is an inherent asymmetry between DL and UL, notably in the SINR levels, and the coverage. While the interference source in the DL are the BSs transmitting at high powers, the interference in the UL is caused by UEs with lower transmission powers and different geographical distribution. Thus, an association policy designed for the DL is not necessarily suitable for the UL, and vice versa [97]. Some studies considered a joint DL and UL user association. In [108], authors proposed an optimization problem that associates users such that a utility function, consisting of a ratio between the DL data rate and the UL power consumption, is maximized. In [97], separate QoS constraints are considered for DL and UL (e.g., a minimum rate), and user association maximizes the sum of weighted utility of long term data rates in DL and UL, while maintaining users QoS.

To date, in all mobile communication standards, a UE associates to one BS at a time, handling its DL and UL traffic. However, with the increase of UL traffic, recent studies are advocating for a decoupled DL and UL user association, where a UE can associate to two distinct BSs, one for the DL and one for the UL [109]. However, implementing such a scheme requires tight synchronization, in addition to high-speed and low-delay connectivity among BSs [110]. Design guidelines on the required changes in current mobile networks to enable this decoupling are discussed in [110]. Preliminary results reported in [109] and [111] were promising. They showed that a decoupled association is advantageous, achieving a very high gain in UL throughput, and a dramatic decrease in the outage rate from 90% to below 10% in networks with minimum throughput requirements.

2.10.3 Backhaul-aware user association

In classical networks, where the backhaul connecting BSs to the core network is usually over-provisioned, the backhaul is understandably ignored in user association mechanisms. Nonetheless, in some network architectures, the specific backhaul implementation should be carefully considered, whether it is a backhaul connecting small and macro cells, or the backhaul between the BSs and a network controller, in HetNets, or inter-BS backhaul links in self-deployable networks. The assumption of over-provisioned backhaul does not hold in these cases, since the backhaul band-

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

width is potentially limited, depending on the used wireless technology, and may create a bottleneck. Ignoring the backhaul load in the user association decision may have a severe impact on the network performance [106]. Therefore, several backhaul-aware user association policies, taking the backhaul bandwidth constraints into consideration, were proposed. Under the same constraints of a limited backhaul, different optimization problems with distinct objective functions were formulated, such as maximizing user data rate [112], maximizing a function reflecting proportional fairness [113], and maximizing the overall network capacity [114]. The authors in [99] proposed a centralized iterative algorithm to optimize association, taking into consideration spectral efficiency, BS load, and backhaul constraints. They also showed that their proposition fits well with decoupled DL/UL association.

2.10.4 Addressed challenges

In our work, we show the inadequacy of a DL-based association policy in self-deployable networks. We argue that user association must be completely rethought, to include key parameters other than the DL signal strength, such as the backhaul capacity, the UL available resources, and users requests. We propose in Chapter 6 an elaborated policy, that adapts to the network architecture and mitigates bottlenecks on the RAN and on the backhaul. Furthermore, while the majority of the aforementioned works formulate optimization problems, with either theoretical analysis or performance evaluation on given independent network snapshots, we propose a user association policy and evaluate it in dynamic network scenarios. Finally, what also distinguishes our work from most of the above-mentioned contributions is the fact that we evaluate our scheme based on the flow blocking probability, rather than the traditional utility-based performance metrics.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 3

Frequency and Power Allocation Scheme

3.1 Introduction

One of the first tasks when configuring a self-deployable mobile network is to determine the frequency allocation scheme, i.e., which frequency channels are used by which BSs, and the power allocation of each BS on each of its channels. This task is usually confronted by one of the most prominent challenges in cellular networks: coping with the scarce available spectrum [57]. This problem has been lingering throughout cellular networks history. Like in any other network with scarce spectrum, it imposes itself in self-deployable mobile networks, with a twist: the network topology is irregular, and the rapid deployment nature bypasses a thorough planning phase.

The efficient exploitation of spectral resources is of vital importance for cellular networks. The goal of spectrum efficiency is to make the most out of a limited available spectrum in order to serve the largest number of users, while satisfying their requirements in terms of throughput and QoE. The reuse of available frequencies by multiple BSs seems to solve the limited spectrum availability problem, but it comes at the costly expense of increasing the interference in the network [58]. Hence, the interference problem comes into view as the limiting factor in the network, and its mitigation becomes a necessity [115]. The primary source of interference in OFDMA networks is ICI, due to the simultaneous use of a frequency channel by multiple BSs. ICI degrades the SINR of users, notably those near the cell boundary. A degrade in the SINR translates into a drop in user throughput [58].

In this chapter, we focus on the downlink of an OFDMA-based self-deployable mobile network. Our goal is to avoid ICI in the network, by proposing an efficient frequency reuse and power allocation scheme, and the algorithm to put it in place. As detailed in Section 2.8, the most basic interference avoidance approach is to have a conventional frequency reuse scheme, referred to as equal power/reuse r (EP-R$_r$). The bandwidth is partitioned into r non-overlapping set of channels, each used by multiple BSs transmitting with the same power on the different channels of a set. EP-R3 is a common approach, with the reuse factor $r = 3$ emanating from the hexagonal grid architecture properties [59] (Fig. 2.4, p. 20). Even though a scheme with $r > 1$ improves the SINR values and minimizes cell-edge interference, it is not spectrum efficient since it reduces the usage of each BS to $1/r$ of the available spectrum. EP-R1, with a reuse factor $r = 1$, such that all the spectrum is used by all the BSs, achieves a better overall throughput, even though it causes severe interference on cell-edge users [62] (Fig. 2.3, p. 19). In an effort to improve performance, several schemes were proposed with the aim of minimizing interference by efficiently allocating

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

resources, while also maximizing spectrum efficiency (see Section 2.8, p. 18). One example is SFR, which performs both radio resource management and power allocation [67]. In SFR, the spectrum is partitioned into sub-bands. Each cell-edge zone uses one sub-band with high power, and the rest are used by the cell-center, with lower power. This means all BSs are allowed to use all the available spectrum, but with different power levels (Fig. 2.6 , p. 21).

Up to now, most studies on frequency allocation in general, and SFR in particular, have been based on the theoretical cellular architecture, with BSs placed in a uniform-size hexagonal grid [67, 69]. One of the particularities of our work is that we study self-deployable networks whose topologies consist of irregular BS placement. In such networks, the number of interfering BSs and the amount of interference are highly variable from one BS to the other [71]. Hence, given a network layout, finding a straightforward and applicable frequency reuse pattern is intuitively not a simple task. The properties based on the hexagonal-grid, such as three sub-bands and two power levels in SFR [67], are not necessarily suitable for an irregular cell pattern. In this chapter, we focus on the SFR scheme, and its adequate parametrization in such networks. We loosen the constraints on the number of sub-bands and the number of power levels. Indeed, we adopt a modified version of SFR, that divides the spectrum into multiple sub-bands, and allow each BS of the network to transmit on all the sub-bands, possibly with a different transmission power level on each [76]. The challenge, in our case, is to find a suitable number of sub-bands, and to assign the corresponding power levels on each sub-band, keeping in mind two key properties of the network: (i) the irregular cell pattern, and (ii) the possible lack of knowledge of the users' distributions and densities and their possible evolutions. Hence, we address the frequency and power allocation problem. One of our concerns is robustness, i.e., to verify that the proposed scheme can be generalized to any given network topology, and that it allows good performance, independently of the user distribution and/or density in the network. While SFR with dynamic power management may be an option, such that power levels are regularly adapted to the actual state of the network [62, 63], questions arise on whether such approaches are feasible, easily manageable in real-time, and sufficiently reactive [80]. Instead, we propose a more practical approach to parameterize SFR, that is offline and easily manageable, and then, we verify its robustness.

The contributions of this chapter can be summarized as follows. First, we propose an offline algorithm that outputs an SFR-based power map to be followed by the BSs, for a given network topology. In other words, the algorithm determines with which power should each BS transmit on each sub-band. The power map is computed prior to the actual network operation, based solely on the network topology, with no accurate information on the user density or distribution to come. Then, this power map is followed by the BSs throughout the network operation. We verify that the power map is robust, in the sense that it performs well, independently of the variations in the network, such as the number of users and their distribution. The algorithm consists of solving a non-convex, non-linear optimization problem, through multiple transformations, based on signomial and geometric programming.

Second, since the proposed algorithm that outputs the power map depends on several input parameters, we assess its performance with respect to those parameters. That is, we study the significance of each of the parameters, and their impact on both the algorithm and the obtained power map performance. Eventually, we give some insights on how to set those parameters, notably the number of sub-bands. This study is done on given snapshots of the network with different numbers of users already associated to BSs, and for different network sizes and topologies.

Third, we evaluate the parametrized SFR on dynamic scenarios, by simulating user arrivals, association, and departures from the network over a period of time. We set as benchmark the classical EP-R1 scheme along with a round robin scheduler. Results show that the parametrized SFR, obtained through our algorithm by solely knowing the network topology, significantly outperforms

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

the benchmark in terms of the achieved throughput, with a gain of up to 40%.

Finally, we compare our scheme with the optimal network performance, obtained when all BSs coordinate while scheduling their users. We solve the system-wide global user scheduling problem to determine how far is our proposition from the optimal case. Results show that our scheme is on average at 70% of this optimal case. We consider this performance acceptable given the practicality of our scheme that allows a per-BS scheduling, in comparison to the optimal case, which is based on solving a difficult to solve optimization problem by a central entity with full knowledge of the network state.

3.2 System model

In this section, we describe the system model on which this work is based. We denote by $\mathcal{J}$ the set of BSs, and by $\mathcal{U}$ the set of UEs in the network. Given a user association strategy, let $\mathcal{U}_j$ be the set of users associated to BS j. We consider an OFDMA-based system with a total of M orthogonal channels allocated to the set of BSs and a time frame made of T time-slots. We assume that all BSs are identical in terms of maximum transmission power P_{BS}, and antenna gain G^a. Channels are flat within a frame, i.e., the gains across different channels between a pair of nodes are equal. We consider an SFR-based scheme, where the frequency band is divided into sets of sub-channels, referred to as sub-bands. Let $\mathcal{S}$ be the set consisting of b sub-bands, where each sub-band has $k = M/b$ sub-channels. All BSs can transmit on all sub-bands, with a different transmission power on each sub-band. We denote by P_j^s the transmission power of a BS $j \in \mathcal{J}$ on sub-band $s \in \mathcal{S}$. P_j^s is the same on all the k sub-channels of sub-band s. We define as a ***power map*** the set of all P_j^s values, indicating with which power each BS transmits on each sub-band. The per channel SINR between UE u and BS j, on a sub-band $s \in \mathcal{S}$, is defined as:

$$\gamma_{u,j}^s = \frac{\frac{P_j^s}{k} \cdot G_{u,j}}{\mathcal{N}_0 + \sum_{h \in \mathcal{J}, h \neq j} \frac{P_h^s}{k} \cdot G_{u,h}}, \tag{3.1}$$

where $\mathcal{N}_0$ is the additive white Gaussian noise power, and $G_{u,j}$ is the channel gain between UE u and BS j, computed in Eq. 3.2. It accounts for the antenna gain G^a, equipment losses E, the path loss $\Gamma_{u,j}$, and shadow fading, modeled with a normal distribution of zero mean and standard deviation sd, all expressed in dB.

$$G_{u,j} = 10^{\left(G^a - E - \Gamma_{u,j} - \mathcal{N}(0,sd)\right)/10} \tag{3.2}$$

The path loss $\Gamma_{u,j}$ between u and j is such that $\Gamma_{u,j} = a + b \cdot \log(D_{u,j})$, where a and b are empirically-determined coefficients that depend on the path loss model, and $D_{u,j}$ is the distance between UE u and BS j [116].

We denote by $R_{u,j}^s$ the rate seen by UE u from BS j on sub-band s. Let $R_{u,j}^s = k \cdot \Psi(\gamma_{u,j}^s)$, with $\Psi(\cdot)$ a function mapping the per channel SINR to the per channel data rate. Function $\Psi(.)$ is a discrete step function, with its values determined by the MCS, such that:

$$R_{u,j}^s/k = \mathfrak{r}_\theta \text{ if } \gamma_{u,j}^s \in [\theta, \theta + 1) \tag{3.3}$$

Due to the discrete nature of $\Psi(.)$, it further complicates our subsequent computations, by adding integer variables to the problem formulation in Section 3.3.1, and necessitating additional transformations [76]. Therefore, to reduce the problem's complexity, we use in this chapter a

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

continuous function $\tilde{\Psi}(\cdot)$ to approximate $\Psi(\cdot)$, with R_{max} the maximum achievable rate set by the MCS [117], and ν and Δ constant parameters, such that:

$$R^s_{u,j} = k \cdot \min\left(\nu \left(\gamma^s_{u,j}\right)^{\Delta}, R_{max}\right) \tag{3.4}$$

Given a power map, the user scheduling (i.e., which users are using which channels and at what time) is done by each BS locally. The smallest scheduling unit in OFDMA-based systems is the PRB, consisting of one time slot and one channel (Fig. 2.2, p. 11). For now, we consider an abstraction of the PRB-based scheduling, by determining the proportion of time during which a user occupies the channels of a sub-band, without specifying the implicated PRBs. We revisit the PRB-based scheduling later in Section 3.7. Specifically, let $\alpha^s_{u,j}$ be the proportion of time during which the channels in sub-band s are allocated to user u in BS j. In this case, the actual throughput ϕ_u of a user u is:

$$\phi_u = \sum_{s \in \mathcal{S}} \alpha^s_{u,j} \cdot R^s_{u,j} \tag{3.5}$$

Our goal is to improve the network throughput, and more specifically, the users' individual throughputs. Fairness among users is one of the criteria we are trying to meet. That is, we do not want the overall gain in the throughput to be caused by only few users with very high throughputs while the others have low throughputs. The geometric mean (GM) of the throughput (Eq. 3.6) is a metric that measures both efficiency and fairness at the same time [92]. Being based on a product, maximizing the GM throughput tries to maximize the individual users' throughputs. Indeed, having a user with a relatively very low throughput has a severe impact on the GM throughput. In comparison, the arithmetic mean of the throughput, based on a sum, loses this fairness criterion. When maximizing the arithmetic mean, one user with a very high throughput can bias the result, causing a high arithmetic mean despite having other users with negligible throughputs. Therefore, under a proportional fairness criteria, we seek to maximize the GM throughput, and adopt it as the evaluation metric. It is written as:

$$\Phi = \left(\prod_{u \in \mathcal{U}} \phi_u\right)^{1/|\mathcal{U}|} \tag{3.6}$$

We remind that, unlike an SFR-based scheme, a conventional EP-R$_r$ scheme divides the frequency band into r sets of channels, and the BSs into r sets. Then, all the BSs of a particular set transmit on only one of the channel sets, dividing their transmit power budget equally between the channels in the set. Note that EP-R1 is a particular case of our adopted SFR-based scheme, with $b = 1$ sub-band, and all the BSs transmitting with equal power $P^s_j = P_{BS}$, $\forall j \in \mathcal{J}$.

3.3 Selection of a robust power map

Our goal is to find a robust SFR-based power map for a given network topology. We refer to it as SFR-based, since each BS can transmit on each sub-band, with a different power level. Such a power map can be computed at the beginning of the deployment, with only knowledge of the network topology. Then, it is fixed and followed by the BSs throughout the network operation. This power map is computed with no prior information on the actual user density and/or distribution in the network. However, we want it to be robust, in the sense that it should work properly, independently of the actual network state (e.g., the varying number of concurrent users) and the implemented network policies (e.g., the adopted user association technique).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

3.3.1 Problem formulation

In order to determine the power map for a given network characterized by the set of BSs $\mathcal{J}$ and M, we generate a priori a set of realizations, referred to hereafter as the set of calibration realizations, denoted Ω. Each realization $\omega \in \Omega$ corresponds to a random deployment of UEs in the network, and is characterized by the number of users N_ω, the sets $\mathcal{U}_j$ for all $j \in \mathcal{J}$, and the channel gains for each pair of BS and UE.

In the following, we formulate the problem $\mathcal{P}_1$ that takes a given topology and a set of calibration realizations as input, and returns one power map as output. Aiming at a proportional fair objective function, the objective function in $\mathcal{P}_1$ (Eq. 3.7) maximizes the arithmetic mean, over all the calibration realizations ω, of the corresponding GM throughput $\Phi(\omega)$ (Eq. 3.8). In constraint 3.9 and constraint 3.10, the SINR $\gamma^s_{u,j}(\omega)$ and the rate $R^s_{u,j}(\omega)$ on sub-band s between BS j and user $u \in \mathcal{U}_j$ are computed, respectively, for each realization w. In constraint 3.12, the throughput $\phi_u(\omega)$ of each user in realization ω is computed. Constraint 3.13 ensures that the total power used by a BS on all the sub-bands does not exceed the maximum transmission power of a BS, P_{BS}. Constraint 3.14 represents the scheduling constraint, where the sum of the time proportions allocated for all the users of a BS on a single sub-band cannot exceed 1, where $\alpha^s_{u,j}(\omega)$ is the proportion of time allocated to user $u \in \mathcal{U}_j$ in sub-band s in realization ω. Finally, constraint 3.15 states that all the variables in $\mathcal{P}_1$ are continuous and positive.

$$\mathcal{P}_1 : \max_{P^s_j, \alpha^s_{u,j}(\omega)} \mathcal{Z} \tag{3.7}$$

$$\mathcal{Z} = \frac{\sum_{\omega\in\Omega} \left(\prod_{u\in\mathcal{U}(\omega)} \phi_u(\omega)\right)^{\frac{1}{|\mathcal{U}(\omega)|}}}{|\Omega|} \tag{3.8}$$

$$\gamma^s_{u,j}(\omega) = \frac{P^s_j \cdot G_{u,j}(\omega)}{k\left(\mathcal{N}_0 + \sum_{h\in\mathcal{J}, h\neq j} \frac{P^s_h}{k} \cdot G_{u,h}(\omega)\right)} \ , \forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall s \in S,\ \forall \omega \in \Omega \tag{3.9}$$

$$R^s_{u,j}(\omega) = k \cdot \nu \cdot \left(\gamma^s_{u,j}(\omega)\right)^{\Delta},\ \forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall s \in \mathcal{S},\ \forall \omega \in \Omega \tag{3.10}$$

$$R^s_{u,j}(\omega) \leq k \cdot R_{max}\ , \forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall s \in \mathcal{S},\ \forall \omega \in \Omega \tag{3.11}$$

$$\phi_u(\omega) = \sum_{s\in\mathcal{S}} \alpha^s_{u,j}(\omega) \cdot R^s_{u,j}(\omega)\ , \forall u \in \mathcal{U}_j(\omega), \forall j \in \mathcal{J},\ \forall \omega \in \Omega \tag{3.12}$$

$$\sum_{s\in\mathcal{S}} P^s_j \leq P_{BS}\ ,\ \forall j \in \mathcal{J} \tag{3.13}$$

$$\sum_{u\in\mathcal{U}_j(\omega)} \alpha^s_{u,j}(\omega) \leq 1\ ,\ \forall s \in \mathcal{S},\ \forall j \in \mathcal{J},\ \forall \omega \in \Omega \tag{3.14}$$

$$P^s_j \geq 0,\ \phi_u(\omega) \geq 0,\ \gamma^s_{u,j}(\omega) \geq 0,\ R^s_{u,j}(\omega) \geq 0,\ \alpha^s_{u,j}(\omega) \geq 0 \tag{3.15}$$

3.3.2 Solving the problem

Problem $\mathcal{P}1$, as formulated in Section 3.3.1, is a non-convex problem with non-linear constraints. In order to solve it, $\mathcal{P}1$ is first transformed to belong to the class of nonlinear optimization called signomial programming problems, and more specifically to complementary geometric program-

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

ming (GP) problems [118, 119]. GP is a class of nonlinear, non-convex optimization problems, characterized by an objective and constraints that have a special form, with theoretical and computational properties that allow its resolution efficiently and reliably [118]. A GP can be easily turned into a convex optimization problem through a logarithmic change of variable, and a global optimum can always be efficiently computed. In a GP problem, all inequality constraints are of the form $g(x) < 1$, with $g(x)$ a *posynomial* [118]. A posynomial is a sum of monomials. A monomial is a function of the form $h_k(x) = d_k x_1^{a_k^{(1)}} x_2^{a_k^{(2)}} ... x_n^{a_k^{(n)}}$, with a constant $d_k \geq 0$ and $a_k^{(j)} \in \mathbb{R}$. Thus, a posynomial is of the form:

$$g(x) = \sum_k h_k(x) = \sum_k d_k x_1^{a_k^{(1)}} x_2^{a_k^{(2)}} ... x_n^{a_k^{(n)}} \tag{3.16}$$

In a complementary GP, a constraint can be of the form $g_1(x)/g_2(x) < 1$, with $g_1(x)$ and $g_2(x)$ posynomials, even though the ratio of two posynomials is not a posynomial. A complementary GP can be turned to a GP using an iterative algorithm [119].

3.3.2.1 Transformation into a complementary geometric program

Problem $\mathcal{P}_1$ can be easily transformed into a complementary GP, by re-writing the constraints to match the typical structure of a complementary GP [119]. First, since we are solving a maximization problem, we replace the equality constraints with inequality constraints, which does not affect the optimal point of the problem. Then, we re-write the constraints as upper-bounded posynomials, or, when needed, ratios between two posynomials. The complementary GP problem, denoted $\mathcal{P}_2$, is formulated as follows:

$$\mathcal{P}_{\mathbf{2}} : \max_{P_j^s, \alpha_{u,j}^s(\omega)} \mathcal{Z} \tag{3.17}$$

$$\frac{\mathcal{Z} \cdot |\Omega|}{\sum_{\omega \in \Omega} \left(\prod_{u \in \mathcal{U}(\omega)} \phi_u(\omega) \right)^{\frac{1}{|\mathcal{U}(\omega)|}}} \leq 1 \tag{3.18}$$

$$\frac{\gamma_{u,j}^s(\omega) \cdot k \cdot \mathcal{N}_0 + \gamma_{u,j}^s(\omega) \cdot k \cdot \sum_{h \in \mathcal{J}, h \neq j} P_h^s \cdot G_{u,h}^s}{P_j^s \cdot G_{u,j}(\omega)} \leq 1 \ ,$$

$$\forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall s \in \mathcal{S},\ \forall \omega \in \Omega \tag{3.19}$$

$$\frac{R_{u,j}^s(\omega)}{k \cdot \nu \cdot \left(\gamma_{u,j}^s(\omega) \right)^{\Delta}} \leq 1 \ , \forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall s \in \mathcal{S},\ \forall \omega \in \Omega \tag{3.20}$$

$$\frac{R_{u,j}^s(\omega)}{k \cdot R_{max}} \leq 1 \ , \forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall s \in \mathcal{S},\ \forall \omega \in \Omega \tag{3.21}$$

$$\frac{\phi_u(\omega)}{\sum_{s \in \mathcal{S}} \alpha_{u,j}^s(\omega) \cdot R_{u,j}^s(\omega)} \leq 1,\ \forall j \in \mathcal{J},\ \forall u \in \mathcal{U}_j(\omega),\ \forall \omega \in \Omega \tag{3.22}$$

$$\frac{\sum_{s \in \mathcal{S}} P_j^s}{P_{BS}} \leq 1 \ ,\ \forall j \in \mathcal{J} \tag{3.23}$$

$$\sum_{u \in \mathcal{U}_j(\omega)} \alpha_{u,j}^s(\omega) \leq 1 \ ,\ \forall s \in \mathcal{S},\ \forall j \in \mathcal{J},\ \forall \omega \in \Omega \tag{3.24}$$

$$P_j^s \geq 0,\ \phi_u(\omega) \geq 0,\ \gamma_{u,j}^s(\omega) \geq 0,\ R_{u,j}^s(\omega) \geq 0,\ \alpha_{u,j}^s(\omega) \geq 0 \tag{3.25}$$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

3.3.2.2 Single condensation method for GP

Complementary GP problems can be solved with the single condensation method, in which they are transformed into a GP problem using an iterative algorithm [118]. To transform the complementary GP $\mathcal{P}_2$ into a GP, the constraints that are not posynomials (i.e., a ratio of posynomials) should be transformed into posynomials to match the structure of a GP. This is done by approximating the denominator with a monomial, while leaving the numerator as a posynomial[1]. For a posynomial $g(x) = \sum_k h_k(x)$, at a given point $x = x_0$, the monomial approximation $\widetilde{g}(x)$ is written as:

$$\widetilde{g}(x) = \prod_k \Big(\frac{h_k(x)}{\beta_k(x_0)} \Big)^{\beta_k(x_0)}, \tag{3.26}$$

where:

$$\beta_k(x_0) = \frac{h_k(x_0)}{g(x_0)} \tag{3.27}$$

In order to find x_0, the point around which $g(x) \approx \widetilde{g}(x)$, an iterative approach is used.

Algorithm 1 Single condensation method for GP

1: Find an initial feasible point $\mathbf{x^{(0)}}$
2: At the i^{th} iteration, approximate $g(x)$ with $\widetilde{g}(x)$ around the point $\mathbf{x^{(i-1)}}$
3: Form the i^{th} approximated GP problem using the approximated $\widetilde{g}(x)$
4: Solve the i^{th} approximated GP problem (by turning it into a convex problem through a logarithmic change of variables) to obtain $\mathbf{x^{(i)}}$.
5: **if** $\left|\left|\mathbf{x^{(i)}} - \mathbf{x^{(i-1)}}\right|\right| < \epsilon$ **then**
6: End
7: **else**
8: Go to step 2
9: **end if**

Specifically, Algorithm 1 is used to solve $\mathcal{P}_2$ iteratively. At each iteration i, at step 2, we approximate the denominator of the left hand side of Eq. 3.18, which is a posynomial, by a monomial. Thus, Eq. 3.18 is replaced by Eq. 3.28.

$$\frac{\mathcal{Z} \cdot |\Omega|}{\prod_{w\in\Omega} \left(\frac{\left(\prod_{u\in\mathcal{U}(\omega)} \phi_u(\omega)\right)^{\frac{1}{|\mathcal{U}(\omega)|}}}{q^i(\omega)} \right)^{q^i(\omega)}} \leq 1, \ \forall j \in \mathcal{J}, \ \forall u \in \mathcal{U}_j(\omega), \tag{3.28}$$

where $q^i(\omega)$ is computed as shown in Eq. 3.29, from the values of $\phi_u^{i-1}(\omega)$ obtained from the solution at iteration $i-1$.

$$q^i(\omega) = \frac{\left(\prod_{u\in\mathcal{U}(\omega)} \phi_u^{i-1}(\omega)\right)^{\frac{1}{|\mathcal{U}(\omega)|}}}{\sum_{y\in\Omega} \left(\prod_{u\in\mathcal{U}(y)} \phi_u^{i-1}(y)\right)^{\frac{1}{|\mathcal{U}(y)|}}} \tag{3.29}$$

Similarly, at each iteration i, at step 2 in Algorithm 1, we approximate the posynomial denominator of the left hand side of Eq. 3.22 by a monomial, i.e., Eq. 3.22 is replaced by Eq. 3.30, where we

[1]The ratio between a posynomial and a monomial is a posynomial.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

omit the variable ω for brevity.

$$\frac{\phi_u}{\prod_{s\in\mathcal{S}}\left(\frac{\alpha^s_{u,j}\cdot R^s_{u,j}}{\theta^{s,i}_{u,j}}\right)^{\theta^{s,i}_{u,j}}} \leq 1\ ,\ \forall j\in\mathcal{J},\ \forall u\in\mathcal{U}_j \tag{3.30}$$

The new exponent $\theta^{s,i}_{u,j}$ is computed as shown in Eq. 3.31, from the values of $\alpha^{s,i-1}_{u,j}$ and $R^{s,i-1}_{u,j}$, obtained from the solution at iteration $i-1$.

$$\theta^{s,i}_{u,j} = \frac{\alpha^{s,i-1}_{u,j}\cdot R^{s,i-1}_{u,j}}{\sum_{l\in\mathcal{S}}\alpha^{l,i-1}_{u,j}\cdot R^{l,i-1}_{u,j}} \tag{3.31}$$

Then, at step 4 of each iteration i in Algorithm 1, the obtained problem is a non-linear, non-convex GP. This problem can be solved efficiently by simply turning it to the following convex problem, denoted $\mathcal{P}_3$, through a logarithmic change of variables. Let $l_u(\omega) = \log\phi_u(\omega)$, $z = \log(\mathcal{Z})$, $p^s_j = \log P^s_j$, $a^s_{u,j}(\omega) = \log\alpha^s_{u,j}(\omega)$, $y^s_{u,j}(\omega) = \log\gamma^s_{u,j}(\omega)$, $r^s_{u,j}(\omega) = \log R^s_{u,j}(\omega)$, and $o^s_{u,j}(\omega) = \log\theta^s_{u,j}(\omega)$.

$$\mathcal{P}_{\mathbf{3}}: \max_{p^s_j(\omega),a^s_{u,j}(\omega)} z \tag{3.32}$$

$$z + \log|\Omega| - \sum_{w\in\Omega} q(\omega)\left(\sum_{u\in\mathcal{U}(\omega)}\frac{l_u(\omega)}{|\mathcal{U}(\omega)|} + \log q(\omega)\right) \leq 0 \tag{3.33}$$

$$\log\left(\frac{k\cdot e^{\left(y^s_{u,j}(\omega)-p^s_j\right)}\left(\mathcal{N}_0 + \sum_{\substack{h\in\mathcal{J}\\ h\neq j}} G_{u,h}(\omega)\cdot e^{p^s_h}\right)}{G_{u,j}(\omega)}\right) \leq 0\ ,$$

$$\forall j\in\mathcal{J},\ \forall u\in\mathcal{U}_j(\omega),\ \forall s\in\mathcal{S},\ \forall w\in\Omega \tag{3.34}$$

$$r^s_{u,j}(\omega) - \Delta\cdot y^s_{u,j}(\omega) - \log k - \log x \leq 0\ ,\forall j\in\mathcal{J},\ \forall u\in\mathcal{U}_j(\omega),\ \forall s\in\mathcal{S},\ \forall w\in\Omega \tag{3.35}$$

$$r^s_{u,j}(\omega) - \log k - \log R_{max} \leq 0\ ,\forall j\in\mathcal{J},\ \forall u\in\mathcal{U}_j(\omega),\ \forall s\in\mathcal{S},\ \forall w\in\Omega \tag{3.36}$$

$$l_u(\omega) + \sum_{s\in\mathcal{S}}\theta^s_{u,j}(\omega)\left(\log\theta^s_{u,j}(\omega) - a^s_{u,j}(\omega) - r^s_{u,j}(\omega)\right)\ ,\forall u\in\mathcal{U}_j(\omega),\ \forall j\in\mathcal{J},\ \forall w\in\Omega \tag{3.37}$$

$$\log\Big(\sum_{s\in\mathcal{S}} e^{p^s_j}\Big) - \log P_{BS} \leq 0\ ,\ \forall j\in\mathcal{J} \tag{3.38}$$

$$\log\Big(\sum_{u\in\mathcal{U}_j(\omega)} e^{a^s_{u,j}(\omega)}\Big) \leq 0\ ,\ \forall s\in\mathcal{S},\ \forall j\in\mathcal{J},\ \forall w\in\Omega \tag{3.39}$$

$$p^s_j \geq \epsilon,\ t\geq\epsilon,\ l_u(\omega)\geq\epsilon,\ y^s_{u,j}(\omega)\geq\epsilon,\ r^s_{u,j}(\omega)\geq\epsilon,\ a^s_{u,j}(\omega)\geq\epsilon \tag{3.40}$$

3.4 Results on a toy scenario

We consider a system with $M = 120$ channels, and a distance-based path loss model for an urban setting, such that $\Gamma_{u,j} = 128.1 + 37.6\cdot\log(D_{u,j}/1000)$ dB ($D_{u,j}$ expressed in m) [116]. We set

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

the BS power to $P_{BS} = 46$ dBm, the antenna gain of the BS to $G^a = 10$ dB, equipment losses to $E = 20$ dB, noise power to $\mathcal{N}_0 = -121$ dBm, and we model shadow fading through the normal distribution with zero mean and standard deviation of 8 dB. For the rate function in Eq. 3.4, we set the parameters $\nu = 0.168$, $\Delta = 0.43$, and $R_{max} = 932.4$ Kb/s, approximating the rate function in a LTE system with 15 discrete rates, as shown in Fig. 3.1.

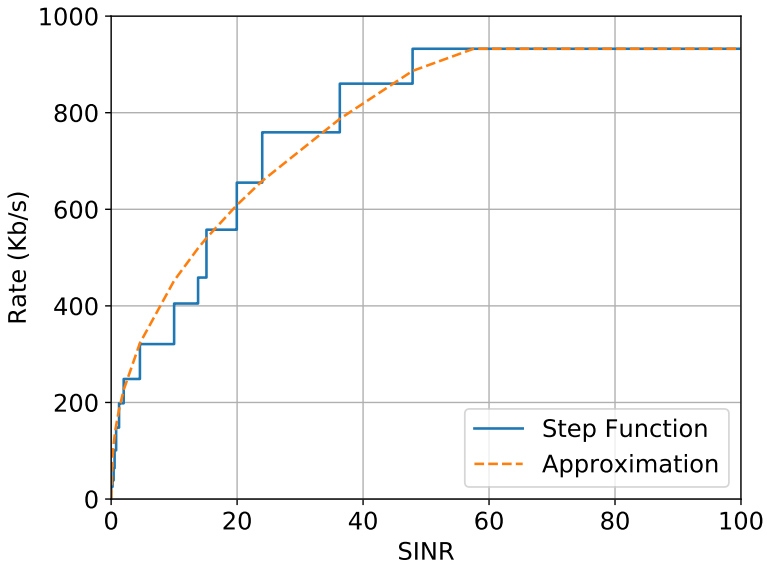


Figure 3.1: MCS-based step function with 15 discrete rates, and its continuous approximation $\tilde{\Psi}(\cdot)$, with $R_{max} = 963$ Kb/s, $\nu = 0.168$ and $\Delta = 0.43$.

In the following, we show a numerical example of an SFR-based power map computation, for a given set of parameters. We first consider the network topology shown in Fig. 3.2, with $|\mathcal{J}| = 5$ BSs, randomly deployed in an area of 1 unit square. Once found, we evaluate the power map performance on network snapshots by computing the achieved GM throughput and comparing it to the one obtained with a classical EP-R$_r$ scheme. The impact of all the input parameters in the power map computation are studied in detail in the next section. Furthermore, we evaluate the power map performance in a dynamic network setting, rather than independent snapshots, in Section 3.6. Throughout the chapter, all optimization problems are solved using the open-source solver "Bonmin" [120], on a server with Intel Xeon CPU E5-2697 v2 @ 2.7 GHz.

3.4.1 Finding the power map

Following the steps presented in Section 3.3 allows us to solve problem $\mathcal{P}_1$, and get a power map with b sub-bands, for a given topology, based on a set of calibration realizations Ω. For example, we consider a set Ω of 10 calibration realizations ($|\Omega| = 10$). In each realization ω, we randomly distribute $N_\omega = 100$ users, and assume that a user associates to the BS that yields the highest channel gain. We set $b = 5$ sub-bands. Note that the impact of b is assessed in the next section. For these values, the obtained power map from Algorithm 1 is shown in Fig. 3.3.

We notice that, for each BS, there is at least one sub-band on which the power is relatively higher than on the other sub-bands. This sub-band is generally different for each BS, meaning two BSs rarely transmit with a high power on the same channels. On the other hand, while a BS transmits with high power on one sub-band, it also transmits on almost all the other sub-bands,

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

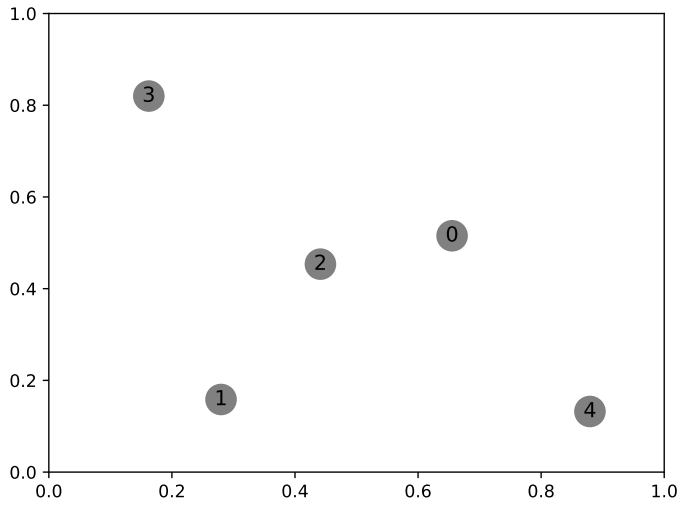


Figure 3.2: A network topology with 5 BSs.

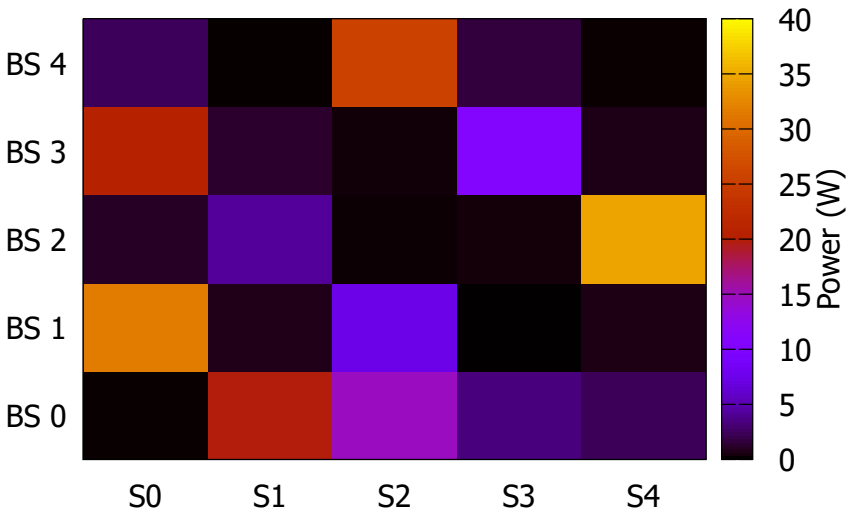


Figure 3.3: Power map for the studied network topology, obtained with $|\Omega| = 10$, $N_\omega = 100$ users, and $b = 5$ sub-bands.

but with relatively lower powers. What we observe is the typical behavior in a SFR scheme: all BSs use all the channels, and allocate few channels with high power (low interference) to users with relatively bad channel conditions (e.g., farthest from BS), and plenty of channels with lower power to users with better channel conditions (e.g., closest to the BS).

3.4.2 Evaluating the power map performance

In this section, we evaluate the performance of the obtained power map on different network snapshots. In order to achieve that, we consider different snapshots capturing different user dis-

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

tributions in the network. For now, we assume that a user associates with the BS offering the best channel gain. For each snapshot, since the transmit powers of the BSs on each sub-band P_j^s are known, the corresponding ICI interference, SINR, and rate can be directly computed. The scheduling is done locally at each BS, for all the users associated to it, independently of the other BSs, but respecting the computed power map. The scheduling assigns, for each user u associated to BS j, the proportion of time, $\alpha_{u,j}^s$, during which all the channels of sub-band s are allocated to u. We consider that users of each BS are optimally scheduled, with the objective of maximizing a proportional fair objective function (Eq. 3.41). The local scheduler in BS j is formulated with the following problem, where the $R_{u,j}^s$ are computed beforehand, given the power map:

$$\max_{\alpha_{u,j}^s} \sum_{u \in \mathcal{U}_j} \log(\phi_u) \tag{3.41}$$

$$\phi_u = \sum_{s \in \mathcal{S}} \alpha_{u,j}^s \cdot R_{u,j}^s \ , \ \forall u \in \mathcal{U}_j \tag{3.42}$$

$$\sum_{u \in \mathcal{U}_j} \alpha_{u,j}^s \leq 1 \ , \ \forall s \in \mathcal{S} \tag{3.43}$$

$$\phi_u \geq 0, \ \alpha_{u,j}^s \geq 0 \tag{3.44}$$

We note that replacing the above problem with another local scheduling algorithm is possible. Depending on the scheduling, i.e., the values of $\alpha_{u,j}^s$, the throughput of each user can be computed according to Eq. 3.5. Eventually, the GM throughput of the snapshot is computed according to Eq. 3.6.

We denote by Π a set of snapshots, with each snapshot $\pi \in \Pi$ corresponding to a random deployment of UEs in the network, $\mathcal{U}(\pi)$, all with the same number of UEs N_Π. For each of the snapshots in Π, we compute the throughput geometric mean, Φ_{PM}, obtained when the computed SFR-based power map is used by the BSs. We remind the reader that the power map was computed independently of Π, using a different set of realizations Ω. For comparison, we also compute for the same snapshots the GM throughput, Φ_{EP-R_r}, obtained when a EP-R$_r$ scheme is used.

Fig. 3.4 shows the values of the computed throughput geometric means, Φ_{PM}, Φ_{EP-R1}, and Φ_{EP-R3}, obtained with the power map in Fig. 3.3, an EP-R1 scheme, and an EP-R3 scheme, respectively. It also shows, for comparison, the corresponding throughput arithmetic means. The results are averaged on $|\Pi| = 100$ snapshots, with $N_\pi = 100$ users in each. The confidence intervals are at 95%. Our power map outperforms the two classical equal power schemes. Indeed, we notice a gain of 42% in GM throughput with respect to EP-R1, and a gain of 64% with respect to EP-R3.

As explained earlier, we adopted the GM throughput as evaluation metric to satisfy a proportional fairness criteria. To verify that maximizing the GM throughput behaves as expected, such that the observed gain entails a gain in the users' individual throughputs, we compute their cumulative distribution function (CDF). For the same snapshots in Π, we show in Fig. 3.5 the CDF of the throughputs of individual users, i.e., the CDF of $\phi_u, \ \forall u \in \mathcal{U}(\pi) \ , \ \forall \pi \in \Pi$. We observe in Fig. 3.5 that an overall increase in users' throughputs is achieved. Indeed, the users' throughputs values are higher with our power map, and the range to which they belong is wider (i.e., the minimum and the maximum throughputs are both higher). This explains that the gain in the GM throughput, seen in Fig. 3.4, is due to an increase in all the users' throughputs in the network when the BSs follow our power map, and does not only concern few users.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

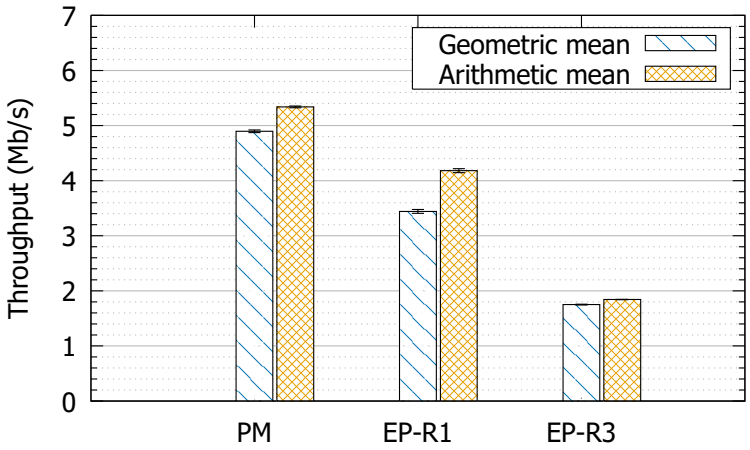


Figure 3.4: Comparison of the throughput geometric mean and arithmetic mean with a soft frequency reuse scheme with the obtained power map (PM), a classical equal power with reuse 1 scheme (EP-R1), and a classical equal power with reuse 3 scheme (EP-R3).

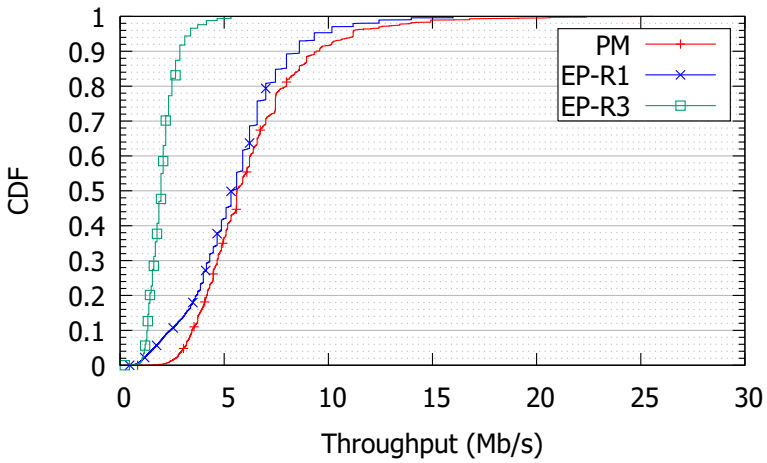


Figure 3.5: The cumulative distribution function of the users' individual throughputs, with a soft frequency reuse scheme with the obtained power map (PM), a classical equal power with reuse 1 scheme (EP-R1), and a classical equal power with reuse 3 scheme (EP-R3).

3.5 Algorithm parameterization

Computing the power map depends on several tunable parameters, raising some questions on its performance with respect to those varying parameters. First, the power map corresponds to a given number of sub-bands b, which determines the number of sub-channel sets the frequency band is divided into. How many sub-bands are sufficient for the power map to perform well? Then, finding the power map is based on two given parameters: the number of calibration realizations $|\Omega|$, and the number of users in each calibration realization N_ω. How many calibration realizations are enough to obtain a robust power map? How does the obtained power map with given N_ω perform in networks with a varying number of users? Finally, each power map depends on a particular network, i.e., a given topology of a precise number of BSs $|\mathcal{J}|$. Does the procedure shown in

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Section 3.3 still hold regardless of the number of BSs in the network and their topology? In the following, we answer these questions by varying each of these parameters, and studying their effect on the power map performance, and give useful insights on how to set those parameters.

3.5.1 Number of sub-bands

As detailed earlier, the frequency band is divided into b sub-bands, and each BS transmits on the channels of each sub-band with a possibly different power level. In other words, the number of sub-bands b also corresponds to the number of possibly different power levels of a BS. In the following, we test different values of b to observe its effect on the power map performance, and to estimate how many sub-bands (or power levels) are considered to be "enough".

For the network topology in Fig. 3.2 with $|\mathcal{J}| = 5$ BSs, we fix $|\Omega| = 10$ calibration realizations, and $N_\omega = 100$ users in all calibration realizations ω, and we vary the number of sub-bands b. For each value of b, we first find the corresponding power map. Then, we compute the GM throughput, $\Phi_{PM}(b)$, achieved with the resulting power map in $|\Pi| = 100$ test snapshots, with $N_\pi = 100$ randomly distributed users in each snapshot. For the same snapshots, we compute the GM throughput, Φ_{EP-R1}, achieved with a classical EP-R1 scheme (we omit hereafter comparisons with the EP-R3 scheme since it performs worse than EP-R1).

We show the results in Fig. 3.6, demonstrating that the relative gain in the achieved GM throughput increases with the number of sub-bands b. We notice that, by dividing the channels into $b = 3$ sub-bands, similarly to a classical SFR scheme, a gain of 37% in GM throughput can be achieved with respect to classical EP-R1 scheme. This gain goes up to 49% with $b = 10$ sub-bands. Nevertheless, increasing the number of sub-bands b also increases the size of problem $\mathcal{P}_1$, which increases the corresponding solving time, due to the convergence time of Algorithm 1. Besides, the increase in the gain becomes slower the more b increases, eliminating the need for higher values of b. A suitable value of b can be determined based on a compromise between the achieved gain and the solving time.

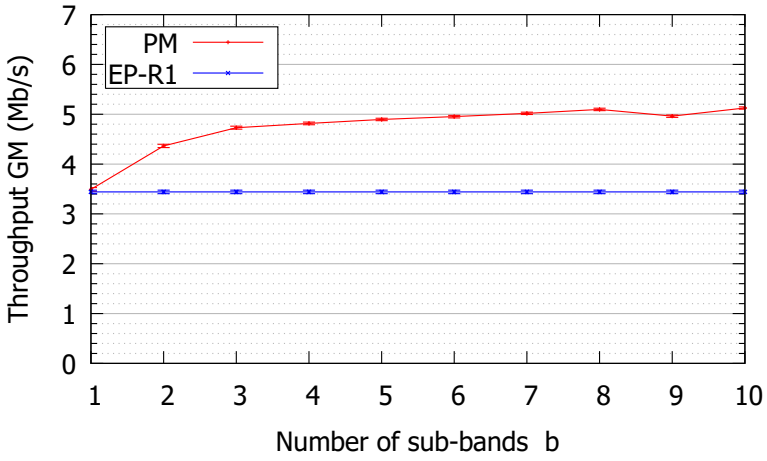


Figure 3.6: Throughput geometric mean function of the number of sub-bands b, tested on $|\pi| = 100$ test realizations, with $N_\pi = 100$ users in each test realization, with a power map obtained for $|\Omega| = 10$ calibration realizations, and $N_\omega = 100$ users.

Table 3.1 shows the computation time as a function of the number of sub-bands b, with $|\Omega| = 10$ and $N_\omega = 100$ users, for all ω. As expected, the convergence time increases with b. Having

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

b	**Time** (min) $\lvert\Omega\rvert = 10$ $N_\omega = 100$	$\lvert\mathbf{\Omega}\rvert$	**Time** (min) $b = 5$ $N_\omega = 100$	$\mathbf{N}_\omega$	**Time** (min) $b = 5$ $\lvert\Omega\rvert = 10$
1	0.03	**5**	46.4	**25**	15.89
2	20.39	**10**	97.33	**50**	41.49
3	68.51	**20**	289.08	**100**	97.33
4	72.41	**30**	469.88	**200**	287.08
5	97.33	**40**	959.12		
6	217.01				
7	214.48				
8	253.16				
9	335.83				
10	342.45				

Table 3.1: Solving time, in minutes, function of number of sub-bands b, the number of calibration realizations $|\Omega|$, and the number of users in each calibration realization N_ω, for $\epsilon = 10^{-6}$.

$b = 5$ sub-bands seems to be a good compromise, achieving a relatively high gain of 42%, with a manageable 5 power levels per BS, and an acceptable solving time.

We note that the convergence time depends on the algorithm stopping criterion, i.e., the value of ϵ in Step 5 in Algorithm 1. The results in Table 3.1 are obtained for $\epsilon = 10^{-6}$. We verified that even higher values of ϵ (e.g., up to 10^{-4}), achieving faster solving times, can be considered without significantly affecting the precision of the solution.

While $b = 5$ sub-bands seems reasonable for this topology of 5 BSs, we show in Section 3.5.4 that the number of BSs and the number of sub-bands are not necessarily correlated. Indeed, we show afterwards that $b = 5$ sub-bands, for example, would still be a reasonable choice for a topology of 15 BSs, generalizing the aforementioned results concerning the number of sub-bands.

3.5.2 Number of realizations

Another critical parameter when finding the power map is the number of calibration realizations $|\Omega|$. Indeed, the objective function in $\mathcal{P}_1$ (Section 3.3.1) is a maximization of the average of the throughput geometric means over all the calibration realizations. We evaluate in the following the impact of having more calibration realizations, and if this improves the power map performance, by testing different values of $|\Omega|$.

For the network topology in Fig. 3.2, we vary the number of calibration realizations $|\Omega|$, while fixing $N_\omega = 100$ users in each calibration realization ω, and $b = 5$ sub-bands. For $|\Pi| = 100$ test snapshots, with $N_\pi = 100$ users in each, we compute the GM throughput $\Phi_{PM}(|\Omega|)$ achieved with the resulting power map for each value of $|\Omega|$, and compare it to the GM throughput Φ_{EP-R1} achieved with a classical EP-R1 scheme (with a performance that does not depend on $|\Omega|$).

Results in Fig. 3.7 show that only a few calibration realizations are enough in order to get a robust power map. Indeed, with only $|\Omega| = 5$ calibration realizations, the gain in throughput geometric mean, with respect to the classical EP-R1 scheme, is already at 42%.

Solving times are shown in Table 3.1 function of $|\Omega|$, with $b = 5$ sub-bands, and $N_\omega = 100$ users. Increasing the number of calibration realizations beyond $|\Omega| = 5$ slightly increases the achieved gain, while significantly increasing the problem size, and consequently, the solving time.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

For example, the gain only goes up to 44% with 40 calibration realizations, even though it takes up to 3 more times to compute in comparison with 5 calibration realizations.

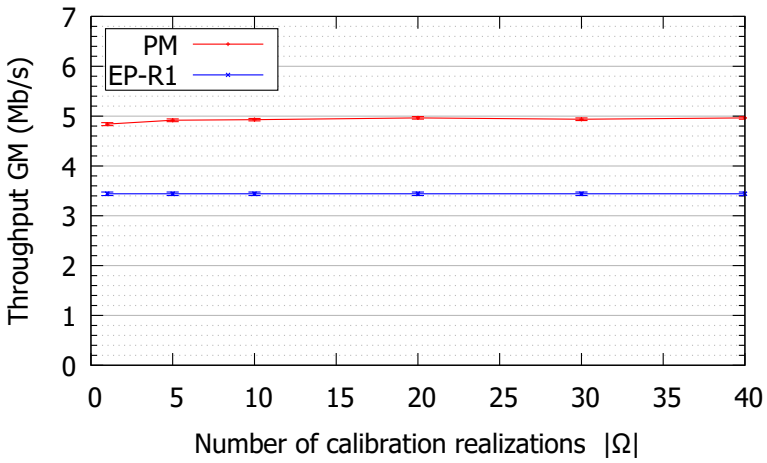


Figure 3.7: Throughput geometric mean function of the number of calibration realizations $|\Omega|$ used to obtain the power map, for $b = 5$ sub-bands, and $N_\omega = 100$ users, tested on $|\Pi| = 100$ test realizations, with $N_\pi = 100$ users in each test realization.

3.5.3 Number of users

The power map is obtained based on a set of calibration realizations Ω, each of them with a random distribution of N_ω users. Then, we are able to test the power map on a set of different snapshots Π, each having a distribution of N_Π users. In the following, we check whether a power map obtained based on calibration realizations with N_ω users can still be used in networks with N_Π users, even when $N_\Pi \neq N_\omega$. In other words, we verify the robustness of the power map with respect to the number of users in the network.

For the network topology in Fig. 3.2, we fix $|\Omega| = 10$ calibration realizations, and $b = 5$ sub-bands. We vary the number of users N_ω in the calibration realizations, and find the corresponding power map for each N_ω. Then, we take four sets of $|\Pi| = 100$ snapshots each. Each of the sets has a different number of users, such that $N_\Pi \in \{25, 50, 100, 200\}$. We test the power maps obtained for each value of N_ω, on the different snapshot sets, i.e., we compute the GM throughput $\Phi_{PM}(N_\omega, N_\Pi)$ for all the combinations (N_ω, N_Π). We compare them to the GM throughput $\Phi_{EP-R1}(N_\Pi)$ achieved with a classical EP-R1 scheme for the different values of N_Π.

Results in Fig. 3.8 show that even when $N_\Pi \neq N_\omega$, the gain in the GM throughput is not really affected. This is why the different curves corresponding to different values of N_ω are actually indistinguishable, independently of the value of N_Π. Indeed, there is practically no difference in the outcome when a power map obtained with calibration realizations with N_ω users is tested on snapshots with $N_\Pi \neq N_\omega$ users. This means that, even with a lack of information on the number of users in a network during the deployment phase, the initial power map can be obtained based on calibration realizations with any value of N_ω, and still be performing well, regardless of the actual number of users in the network. Moreover, this is an indicator of the power map robustness facing drops or increases in the number of users throughout the network deployment.

In fact, the power map returned by the algorithm performs well in all cases since it is constructed with both low-power and high-power sub-bands. The ones with low power are serving the users close to the BS, the ones with high power are serving users in the cell-edge. Therefore,

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

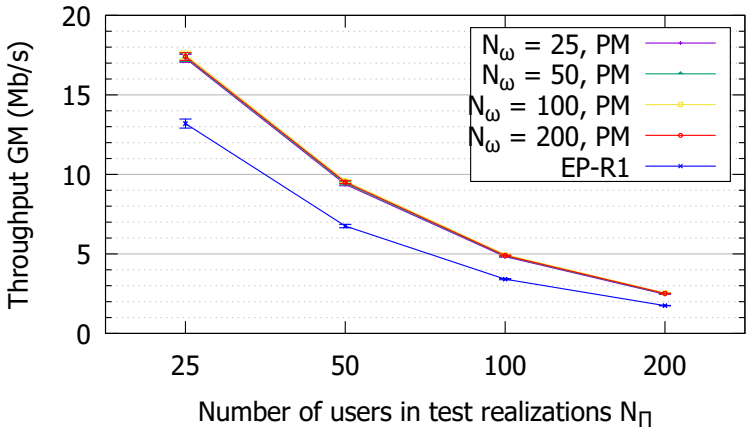


Figure 3.8: Throughput geometric mean function of the number of users in test realizations N_π, for different values of the number of users in calibration realizations N_ω, for $b = 5$ sub-bands, and $|\pi| = 100$ test realizations.

the number of users used in the calibration realizations N_ω is not relevant, as long as the user distribution in ω has users in both areas (cell-center and cell-edge). This is also the case for the number of calibration realizations $|\Omega|$, as seen in the previous section. On a side note, while the gain in GM throughput is not affected by N_ω, the GM throughput itself is affected by the actual number of users in the test snapshots N_Π. As expected, the higher the number of users, the lower the GM throughput, which explains the decreasing curve shape in Fig. 3.8.

On the other hand, increasing the number of users in the calibration realization, N_ω, increases the problem size and the convergence time. The corresponding solving times are shown in Table 3.1 function of N_ω, with $b = 5$ sub-bands, and $|\Omega = 10|$ calibration realizations.

3.5.4 Network topology

All the previous results were obtained for the particular topology of 5 BSs shown in Fig. 3.2. In order to verify and generalize the results, we repeated the previous study on a multitude of topologies, with different numbers of BSs. The same evaluation on the number of sub-bands, the number of realizations and the number of users, and their impact was conducted as in Section 3.5. All the results are consistent with the previous ones, and all the conclusions hold, regardless of the network topology or the number of BSs in question.

As an example, we show in Fig. 3.9 of a network topology with $|\mathcal{J}| = 15$ BSs randomly deployed. Fig. 3.10 shows the power map computed for this topology, with the following parameters: $b = 5$ sub-bands, $|\Omega| = 10$ calibration realizations, and $N_\omega = 150$ users in each calibration realization. We notice that, on each sub-band, there is at least one BS that transmits with a significantly higher power than the others. Most of the BSs, however, do not use their full power. Moreover, even though most of the BSs transmit on several sub-bands, there is at least one sub-band **per** BS where transmission power is negligible. The distribution of transmission powers on the sub-bands can differ significantly from one BS to another. For example, while BS 5 has transmission powers ranging from 1 to 32 W, on 4 sub-bands, BS 7 transmits with powers lower than 1 W on all the sub-bands.

Nevertheless, we show in Fig. 3.11 that all BSs are operational and serving users, even those with seemingly low powers. Indeed, Fig. 3.11 shows the GM throughput achieved per BS, i.e.,

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

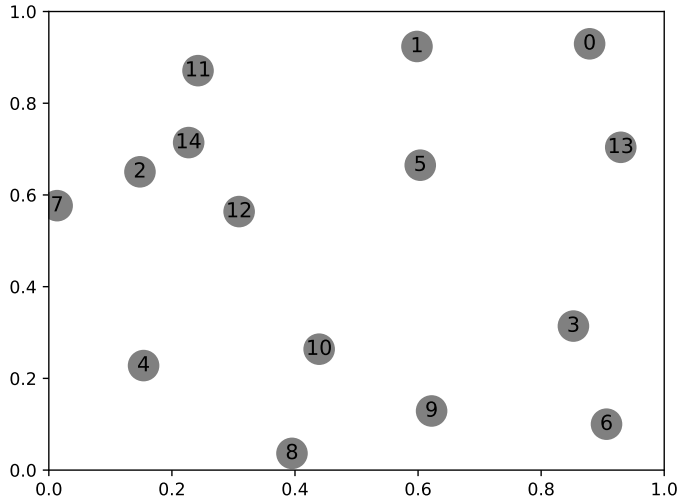


Figure 3.9: A network topology with 15 BSs.

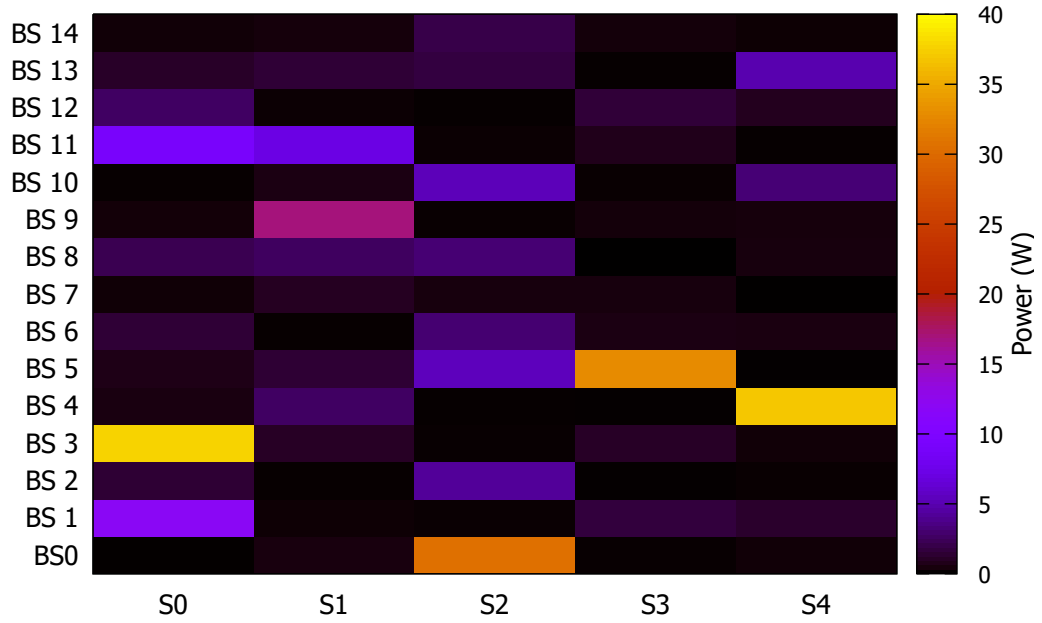


Figure 3.10: Power map for the network topology with 15 BSs, obtained for $|\Omega| = 10$ calibration realizations, $N_\omega = 150$ users, and $b = 5$ sub-bands.

the GM of the throughputs of the users associated to each BS. These results are obtained for $|\Pi| = 100$ test snapshots, and $N_\pi = 150$ users in each snapshot. We notice that, for all BSs, the per BS GM throughput achieved with the power map in Fig. 3.10 outperforms the one achieved with an EP-R1 scheme. Moreover, even BSs with lower powers on most sub-bands, such as BS 7, have a significant value of GM throughput. Those BSs are mostly serving a small number of users that are close to them, i.e., with good channel conditions, and hence, do not need high transmission power. We note that the achieved GM throughput per BS depends on the number of

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

users associated to that BS. The more there are users sharing the BS, the less the throughput per user is.

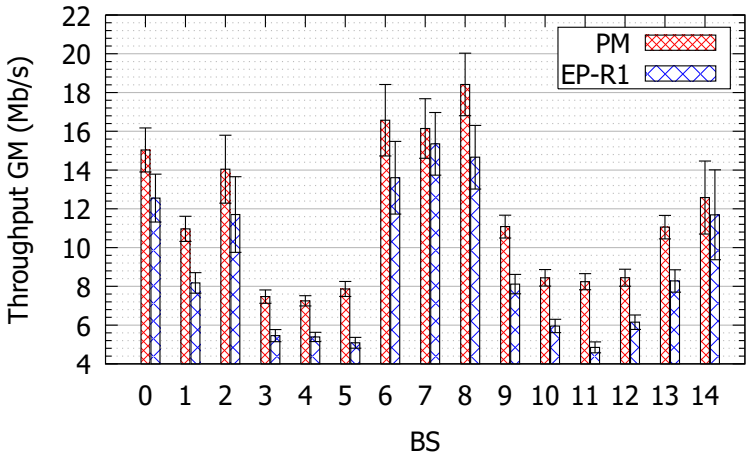


Figure 3.11: The geometric mean of the throughput of users associated to each BS, with a soft frequency reuse scheme with the obtained power map (PM), and a classical equal power with reuse 1 scheme (EP-R1).

While $b = 5$ sub-bands seemed reasonable for the previous topology of 5 BSs, we verify in the following if this is still the case for the topology with 15 BSs. For the topology in Fig. 3.9, we vary b, and compute for each value the GM throughput, $\Phi_{PM}(b)$, achieved with the resulting power map, for $|\Pi| = 100$ test snapshots, and $N_\pi = 150$ users in each snapshot. Results are shown in Fig. 3.12. The behavior is the same as previously seen in Fig. 3.6. The relative gain in the GM throughput, with respect to the EP-R1 scheme, increases with the number of sub-bands, and the increase in the gain becomes less important. The gain is around 40% with $b = 5$ sub-bands, and can go up to 46% with $b = 15$ sub-bands. Thus, having 15 BSs does not necessarily mean that there should be more sub-bands. The same value of b can be picked for the different topologies. For example, $b = 5$ still seems to be a good compromise between good performance and the size of problem $\mathcal{P}_1$.

3.6 Dynamic simulation

In the previous sections, we evaluated the power map performance on network snapshots, in which users are already in the network, and associated to the BSs. Solving these snapshot problems allowed us to verify the advantages of using an SFR-based power map, as well as the robustness of the latter. In order to verify the performance of the power map in more practical scenarios, we consider in this section a dynamic network setting. We will also use this setting to verify that the previous results hold for different user associations.

For the simulations, we use a home-built Python simulator based on “SimPy”, a process-based discrete-event simulation library [121].

3.6.1 Simulation setting

For a given network topology, we compute a power map with b sub-bands using Algorithm 1. Hence, the power of each BS on each sub-band is known. We then start the simulation. Users

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

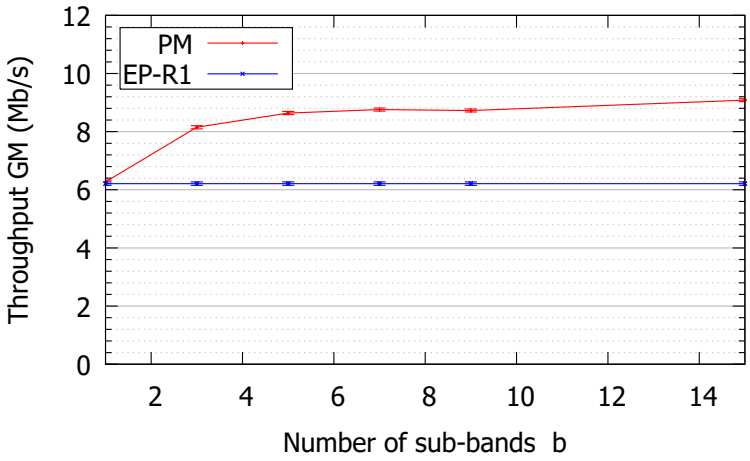


Figure 3.12: Throughput geometric mean function of the number of sub-bands b, for $|\Omega| = 10$ calibration realizations, and $N_\omega = 150$ users, tested on $|\Pi| = 150$ test realizations, with $N_\pi = 150$ users in each test realization.

appear in the network in random static positions, with their arrival following a Poisson process, with an average arrival rate λ (s^{-1}). Each user must first associate to a BS, and then download a file of fixed size. Once the file is downloaded, the user de-associate from the network. We compare two association policies. The first policy is "Best SINR" (given in Eq. 3.45), in which a user u associates to the BS j' from which it receives the highest SINR (Eq. 3.1) on one of the sub-bands. The second is denoted "Load Aware" (given in Eq. 3.46), in which a user u associates to the BS j^* in which it would see the the highest ratio of the maximum rate (Eq. 3.4) over the number of users associated to that BS, $|\mathcal{U}_j|$, on one of the sub-bands. Unlike "Best SINR", this user association policy not only accounts for the peak rate a user can get from a BS on each sub-band, but also the BS load, represented by the number of users associated to it.

$$j' = \arg \max_{\forall j \in \mathcal{J},\, \forall s \in S} \left(\gamma^s_{u,j}\right) \tag{3.45}$$

$$j^* = \arg \max_{\forall j \in \mathcal{J},\, \forall s \in S} \left(\frac{R^s_{u,j}}{|\mathcal{U}_j|}\right) \tag{3.46}$$

Once user u associates to BS j, the scheduling is done locally at BS j, for all the users associated to it, independently of the other BSs. Proposing online scheduling schemes is out of the scope of this chapter. Hence, without loss of generality, we consider that users of each BS are optimally scheduled, with the objective of maximizing the proportional fair objective function in a BS (see Section 3.4.2, Eq. 3.41-Eq. 3.44). When the scheduling is known, the throughput ϕ_u of each user can be computed (Eq. 3.5). The scheduling is done by the BS on a per-frame basis, meaning the values of $\alpha^s_{u,j}$ and, consequently, ϕ_u can be different in each frame, depending on the number of users actually associated to the BS. Based on the different per-frame throughputs of a user u, we compute the sojourn time of user u in the network, i.e., the time it takes u to download the file of fixed size. Finally, we compute the average user sojourn time in the network. The user sojourn time is in fact an indication of the user throughput, since users with higher throughputs have a lower sojourn time.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

3.6.2 Simulation results

We consider the topology with 5 BS depicted in Fig. 3.2. For comparison, we repeat the same simulation scenario for two initial power map configurations: (1) the SFR-based power map with $b = 5$ sub-bands (Fig. 3.3), (2) a classical EP-R1 scheme. Then, we compute the average user sojourn time with each of those power map configurations. Fig. 3.13 shows the average sojourn time τ in each of the studied cases, function of the average user arrival rate, with a confidence interval at 95%. The average user sojourn time is computed based on the batch means method, i.e., running a very long simulation, dividing it up into several batches of equal duration, then computing the mean over the batches in the steady state [122].

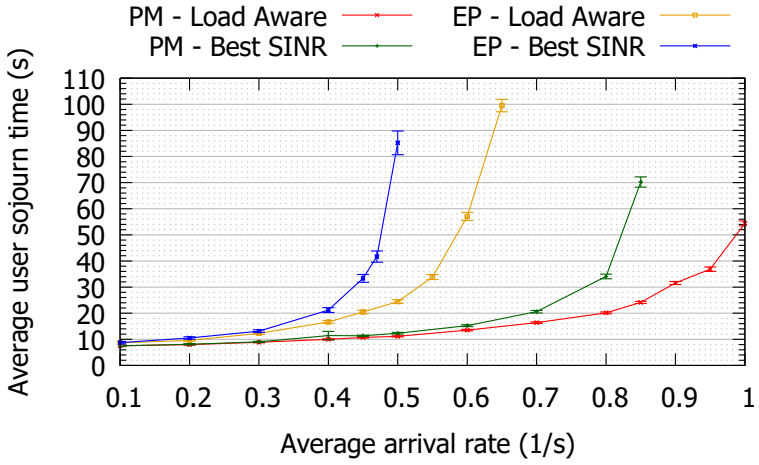


Figure 3.13: Average user sojourn time to download a file of 50 MB, as a function of the user arrival rate, with a classical equal power reuse 1 scheme and with the power map obtained for $|\Omega| = 10$ calibration realizations, $N_\omega = 100$, and $b = 5$ sub-bands, for different user association schemes (Best SINR and Load Aware).

We can make three conclusions when observing the results in Fig. 3.13. First, our power map leads to significant gains in terms of user sojourn time, which corroborates the previous results based on snapshots. For example, the maximum arrival rate to get an average sojourn of 10 s is about 0.2 users per second with the classical EP-R1 power map while it is about 0.45 users per second for our power map. Second, as expected, a load-aware association outperforms a best SINR association. Third, while the power map was obtained for a user association policy based on channel gains in the calibration phase, we show that this power map is performing well, even when different user association policies are adopted during network operation.

3.7 Global scheduling problem

In this section, we consider the optimal system performance with full BS coordination. We solve the system-wide user scheduling problem, in order to compare the achieved GM throughput with our pre-computed SFR-based power map with respect to the optimal case. The question we are looking to answer is: how far is our proposed scheme from optimality?

In fact, in an OFDMA frame consisting of a set of channels, denoted $\mathcal{M}$, and a set of time slots, denoted $\mathcal{T}$, scheduling is practically done per frame. The smallest scheduling unit that can be assigned to a user is a PRB, and it consists of one time slot $t \in \mathcal{T}$ and one channel $c \in \mathcal{M}$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

(Fig. 2.2, p. 11). Each BS serves its users by allocating the PRBs to the users on a per-frame basis [18] (see Section 2.1.2.3).

To get the optimal system performance, the BS transmission power on each PRB must be determined. That way, the interference is determined on a PRB basis, and the power is adjusted accordingly, also on a PRB-basis. This is only possible if the schedules of all the BSs are coordinated. Nevertheless, in practice, full BS coordination and power adjustment with PRB granularity is not feasible in real-time. One way to implement it is to have a central unit collect information from all the BSs, and all the channel states of the UEs in the network, determine the PRB allocation and the transmission power per PRB based on a defined objective function, then send the allocation information back to the BSs. This is not practical due to the long delays it causes, since the global scheduling problem is very hard to solve, as explained in the following, in addition to the generated communication overhead.

3.7.1 Problem formulation

We denote by $\mathcal{P}_{OPT}$ the global scheduling problem. It is a MINLP that includes a high number of binary variables and non-linear constraints. In $\mathcal{P}_{OPT}$, we aim to maximize a proportional fair objective function among all the users, within a frame of M channels and T time slots. We note that the following formulation considers the BSs that are using the same channels. Hence, for networks with a frequency reuse factor $r > 1$, the scheduling problem can be separated into r distinct problems, equivalent to $\mathcal{P}_{OPT}$.

Given the set of users associated to each BS j, $\mathcal{U}_j$, and the channel gains $G_{u,j}$, the problem outputs the binary variables $x_u^{c,t}$, that indicate if a user u is assigned the PRB at channel c and time slot t, and $P_{u,j}^{c,t}$, the BSs transmit powers on each of their PRBs. The objective function in Eq. 3.48 maximizes the product of the throughput of all the users in all the BSs (which is equivalent to maximizing the GM throughput). The throughput of a user, which is the sum of the rates at all PRBs assigned to that user, is computed in constraint 3.49. The SINR and the rate on each PRB are computed in constraint 3.50, constraint 3.51, and constraint 3.52, respectively. The constraint 3.53 ensures that the total power used by BS j cannot exceed P_j, the total power available to BS j. The constraint 3.54 ensures that a BS cannot allocate power to a PRB to serve a user if that PRB is not allocated to that user. The constraint 3.55 ensures that a PRB of a BS cannot be allocated to a user associated to another BS. The constraint 3.56 enforces that a PRB can be allocated to one and only one user.

With a large set of integer variables, and a non-linear objective function, $\mathcal{P}_{OPT}$ cannot be solved using a commercial solver, except when very small. Ozcan *et al.* [76] proposed a method to solve it. First, the integer variables $x_u^{c,t}$, used for PRB allocation, are eliminated by replacing the constraints from Eq. 3.54 to Eq. 3.56 by the following constraint:

$$P_{u,j}^{c,t} P_{v,j}^{c,t} \leq \epsilon \tag{3.47}$$

This forces the product of the powers allocated for two users on the same PRB to be close to zero, so that only one user is actually served on that PRB. Then, the problem is transformed to belong to the class of signomial programming problems, and more specifically to a complementary GP problem. Finally, the problem is solved in the same way explained in details in Section 3.3.2 to solve $\mathcal{P}_1$.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

$$\mathcal{P}_{\mathbf{OPT}} : \max_{P_{u,j}^{c,t}, x_{u,j}^{c,t}} \prod_{u \in \mathcal{U}_j} \prod_{j \in \mathcal{J}} \phi_u \tag{3.48}$$

$$\phi_u = \frac{1}{T} \sum_{c \in \mathcal{M}} \sum_{t \in \mathcal{T}} R_u^{c,t}, \ \forall j \in \mathcal{J}, u \in \mathcal{U}_j \tag{3.49}$$

$$\gamma_{u,j}^{c,t} = \frac{P_{u,j}^{c,t} \cdot G_{u,j}}{\mathcal{N}_0 + \sum_{h \in \mathcal{J}, h \neq j} \sum_{v \in \mathcal{U}_h} P_{v,h}^{c,t} \cdot G_{u,h}} \ , \forall j \in \mathcal{J} \ , \ \forall u \in \mathcal{U}_j(\omega), \ \forall c \in \mathcal{M}, \ \forall t \in \mathcal{T} \tag{3.50}$$

$$R_{u,j}^{c,t} = \nu \cdot \left(\gamma_{u,j}^{c,t}\right)^{\Delta} \ , \ \forall j \in \mathcal{J}, \ \forall u \in \mathcal{U}_j, \ \forall c \in \mathcal{M}, \ \forall t \in \mathcal{T} \tag{3.51}$$

$$R_{u,j}^{c,t} \leq R_{max} \ , \ \forall j \in \mathcal{J}, \ \forall u \in \mathcal{U}_j, \ \forall c \in \mathcal{M}, \ \forall t \in \mathcal{T} \tag{3.52}$$

$$\sum_{u \in \mathcal{U}_j} \sum_{c \in \mathcal{M}} P_{u,j}^{c,t} \leq P_{BS} \ , \ \forall j \in \mathcal{J}, \ \forall t \in \mathcal{T} \tag{3.53}$$

$$P_{u,j}^{c,t} \leq x_{u,j}^{c,t} P_{BS} \ , \ \forall j \in \mathcal{J}, \ \forall u \in \mathcal{U}_j(\omega), \ \forall c \in \mathcal{M}, \ \forall t \in \mathcal{T} \tag{3.54}$$

$$x_{u,j}^{c,t} = 0 \ , \ \forall j \in \mathcal{J}, \ \forall u \notin \mathcal{U}_j(\omega), \ \forall c \in \mathcal{M}, \ \forall t \in \mathcal{T} \tag{3.55}$$

$$\sum_{u \in \mathcal{U}_j} x_{u,j}^{c,t} \leq 1 \ , \ \forall j \in \mathcal{J}, \ \forall c \in \mathcal{M}, \ \forall t \in \mathcal{T} \tag{3.56}$$

3.7.2 Numerical results

We consider the topology with 5 BS depicted in Fig. 3.2. We keep the same numerical values for all the system parameters as presented in Section 3.4, except the total number of channels M, which we fix in this section to $M = 45$ channels, to reduce the problem size. We consider a set Π of 20 test snapshots, with each snapshot $\pi \in \Pi$ corresponding to a random deployment of UEs in the network, with $N_\Pi = 50$ UEs. For each of the snapshots in Π, we compute the GM throughput obtained by solving the global scheduling problem P_{OPT}, with a frequency reuse factor $r = 1$. We note that solving time for P_{OPT} is more than a day for only one snapshot. For comparison, we also compute, for the same snapshots, the GM throughput obtained with a pre-computed SFR-based power map with $b = 5$ sub-bands, and with an EP-R1 scheme.

Results are shown in Fig. 3.14. On average, the GM throughput obtained with our map is at 70% of the optimal case. However, the power map is computed offline prior to the network operation, the solving time is much faster, it does not require coordination between the BSs, and the scheduling is done by each BS independently of the others. Considering all those factors, the power map performance is relatively good in comparison with the optimal case.

3.8 Conclusion

We addressed in this chapter the BSs frequency and power allocation scheme in a self-deployable mobile network. In other words, we determined which channels are used by which BSs, and the transmit power of each BS on each of its channels. We proposed an offline algorithm capable of finding a robust SFR-based power map, computed prior to the network operation, with the sole knowledge of the network topology, and no coordination between the BSs. The algorithm consists

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

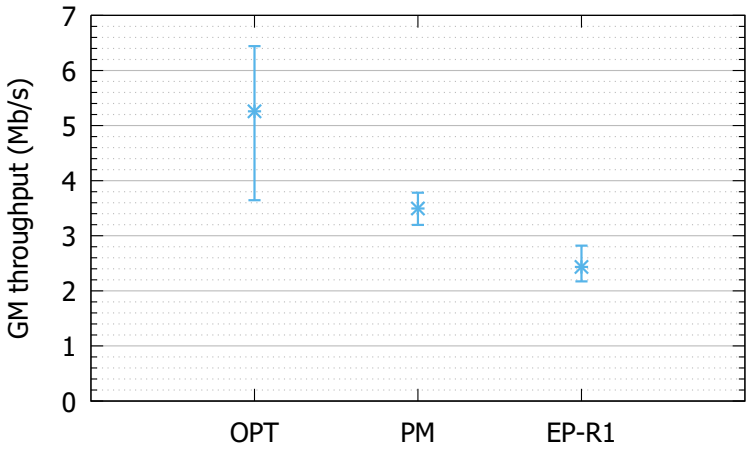


Figure 3.14: Comparison of the average, minimum, and maximum throughput geometric mean in the global scheduling optimal problem (OPT), with an SFR-based power map (PM), and a classical equal power with reuse 1 (EP-R1) scheme.

of solving a non-convex, non-linear optimization problem, using an iterative approach that transforms it into a series of geometric programs, easily transformed into convex problems and solved efficiently. The algorithm takes as input a set of realizations, denoted as calibration realizations, representing different network snapshots. Then, it computes the power map that maximizes the arithmetic mean, over all those calibration realizations, of the corresponding GM throughputs. Once computed, the power map is fixed and followed by the BSs in the network throughout their operation. We showed that this power map is robust, meaning it maintains a good performance, regardless of varying conditions, such as the number of users, and the association policy. We analyzed the different algorithm parameters, and studied their significance, as well as their impact on the performance of the algorithm and the resulting power map. Performance evaluation was conducted on networks of different sizes and topologies. Numerical results showed that a power map obtained with our algorithm outperforms a classical EP-R1 scheme, in both independent snapshots and dynamic simulations, achieving a gain of more than 40% in GM throughput. Moreover, with this power map, the system is at 70% of the system-wide optimal performance, where scheduling and power allocation are optimized for each BS on a PRB basis.

In this work, the power map is pre-computed and fixed throughout the network operation. A more dynamic approach could be adopted, in which changes in the power map are applied depending on the network state, based on performance thresholds for example. Reactivity and practicality of such schemes are key. Moreover, based on these results, further analysis of the obtained power maps structures could serve as basis for the development of simpler and faster heuristics, allowing the easier construction of robust power maps, on the fly, for given topologies. Furthermore, we note that, while we particularly focused on the frequency reuse scheme on the downlink, the uplink is not to be neglected, with the uplink interference having serious consequences on the uplink users' throughputs. Determining interference on the uplink is trickier because, unlike the downlink, the interferers in this case are users actually sharing the same channel at the same time. Hence, adequate uplink-aware reuse schemes should be further investigated. Finally, we note that the frequency and power allocation scheme presented in this chapter and the corresponding algorithm can be further generalized and applied to any cellular network with irregular topology to determine the BSs power map.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 4

Core Network Functions Centralized Placement

4.1 Introduction

Self-deployable mobile networks, like any other cellular network, comprise a RAN and a CN. While the RAN is responsible of radio-related functions, e.g., radio resource management, the CN handles several functionalities, such as authentication, paging, and routing, to name a few. However, the architecture of a self-deployable network is inherently different from that of a classical networks. In fact, in classical networks, the split between the RAN and the CN is both functional and physical. The CN, essential to enable the network to provide services to users, is a physical separate entity from the RAN, and comprises several logical and/or physical entities (e.g., MME, HSS, S-GW, P-GW, in the EPC architecture, as detailed in Chapter 1, Section 2.1.2, p. 8). Usually, each BS of the RAN has a dedicated backhaul link towards the CN, cautiously dimensioned to support the BS load, and to guarantee peak data rates and high-speed services. With this architecture, careful planning is needed before a network deployment, notably for the backhaul and CN dimensioning. In contrast, in self-deployable networks, the functional split between the RAN and the CN is not necessarily physical. Indeed, BSs do not necessarily have a dedicated backhaul connection to a traditional CN [12]. Instead, a BS is co-located with, or at least has access to, a local core network, referred to hereafter as Local CN, which is an entity analogous to the traditional CN, providing the same basic functionalities as the latter, in addition to housing the application servers [123]. Unlike a traditional CN, the Local CN can be co-located with the BS, using function virtualization technologies [13].

We remind that saying that we have one Local CN serving the network is an analogy to the single CN serving a classical operator network. Nevertheless, that does not mean that the Local CN cannot comprise multiple instances of the same function, similarly to how an EPC, for example, comprises a pool of MMEs and a pool of S-GWs, all belonging to the same EPC. Hence, one or several instances of the same function of a Local CN can exist in the same network. Accordingly, the Local CN can be either centralized (all functions co-located with the same BS) or distributed (functions and their different instances co-located with several BSs). In this chapter, we only address a centralized Local CN. A comparison with the distributed approach is kept for the next chapter.

Focusing on the Local CN configuration in a self-deployable network, we address a crucial problem in the deployment phase: where should the Local CN functionalities be placed in the network? The relevance of this problem emanates from the particular architecture of self-deployable

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

networks. With the BSs being interconnected, and the Local CN co-located with one of them, all traffic (e.g., data and signaling), usually exchanged between each BS and the standalone CN, is now routed locally on the links interconnecting the BSs. Those inter-BS links form henceforth the backhaul network of the system, with a potentially limited bandwidth (Fig. 4.1). Thus, in a self-deployable network with multiple interconnected BSs, one of the BSs must be co-located with a Local CN, in order to provide the necessary CN functionalities. The other BSs in the network must then be able to reach the designated BS co-located with the Local CN. Consequently, a placement problem arises, questioning where should the Local CN be placed in the network.

The driving idea is that the Local CN should be able to receive the traffic destined to it from all the BSs, via the inter-BS backhaul links. Reciprocally, the BSs must also be able to receive the traffic destined to them originating from the Local CN. However, the backhaul network, having limited bandwidth (i.e., limited link capacities), may represent a bottleneck by limiting the amount of traffic that can be exchanged between the BSs and the Local CN. Hence, the limited backhaul bandwidth plays an essential role in determining where the Local CN should be placed, in a way that ensures that the traffic can be exchanged without losses due to backhaul saturation. Moreover, the amount of traffic routed in the network depends on the number of users to be served, or more precisely, the number of user requests. Nevertheless, the number of users and their requests are dynamic. This means that the best placement of the Local CN might change depending on the user traffic corresponding to each BS. Since such information may not be available at the early deployment phase, we propose to place the Local CN at first in a general way, independently of user behavior. Once deployed and operational, a migration of the Local CN may or may not be recommended. This, however, is out of the scope of this chapter.

For now, we aim to maximize the amount of traffic each BS can send towards (resp. receive from) the Local CN, by treating all the BSs as equals from a traffic volume point of view. We suppose the inter-BS backhaul links have the same bandwidth in both directions. In this case, since we are treating BSs equally, the maximum traffic that can circulate in one direction is equivalent to the maximum traffic that can circulate in the other direction. Hence, for brevity, we focus in the remainder of the chapter on one traffic direction, with BSs sending traffic towards the Local CN. The problem in the other direction is equivalent. Suppose that each BS in the network must be able to send the same amount of traffic $\rho(d)$ towards the Local CN co-located with BS d. Our goal is to find the BS d for which $\rho(d)$ is maximized. In other words, we are aiming to maximize the possible amount of traffic that all BSs are capable of forwarding to the Local CN simultaneously, while respecting the limited backhaul bandwidth.

The contributions of this chapter can be summarized as follows. First, we propose a new centrality metric in a network, denoted as flow centrality. This metric measures the capacity of a node in the network (i.e., a BS in our case) in receiving the total amount of flows. The flow centrality of a node is represented by the maximum traffic that can be sent simultaneously by every other node in the network towards this node, while respecting link capacity constraints. This definition is based on an underlying linear optimization problem. We argue that a Local CN should be co-located with the BS in the network having the maximum flow centrality. This would allow it to receive the maximum traffic from each BS in the network, under certain capacity and load distribution constraints [124].

Second, we elaborate an analytical study allowing the computation of flow centrality in canonical non-random graphs, such as path graphs and balanced trees. Then, building on extensive simulations with different parameter settings, we further deduce general properties of the flow centrality metric, such as the dependence between the latter and the backhaul network properties, and the position of the node with the maximum flow centrality in a network.

Third, we benchmark the flow centrality metric by comparing it to different state of the art

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

centrality metrics. We highlight the loss in the total amount of traffic received by the Local CN when the latter is placed on a node not having the maximum flow centrality. Numerical results validate the advantages of placing the Local CN on the node with the maximum flow centrality, as the traffic loss occurring otherwise is significant, reaching 55% in some cases.

4.2 System model

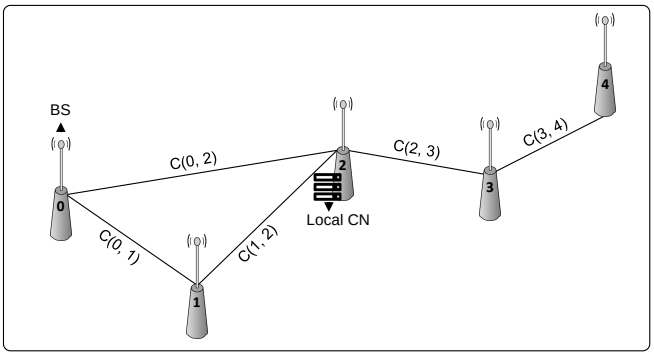


Figure 4.1: An example of a self-deployable network topology with a centralized Local CN co-located with one BS and inter-BS backhaul links of capacity $c(u, v)$.

We consider a self-deployable network consisting of BSs having no backhaul connectivity to a traditional CN. All these BSs must access a Local CN in order to be able to serve users. There is a centralized Local CN in the network, and it is co-located with one of the BSs. A Local CN provides the same functionalities as the traditional CN, supports bearer services, and houses the application servers. In case a BS is not co-located with, nor can reach a Local CN instance, it cannot serve users. We consider that the BSs are interconnected, meaning two BSs can be directly connected via a link. As the Local CN is co-located with one of the BSs, the inter-BS links are referred to as backhaul links, since they will be responsible of forwarding all data and signaling traffic between each BS and the Local CN, respectively. Exchanged traffic on the backhaul links between a BS and the Local CN is routed either directly, if the Local CN is at one hop from the BS, or through interconnected BSs, via the inter-BS backhaul links. Fig. 4.1 illustrates an example of a network topology with five interconnected BSs, and a Local CN co-located with one of them.

We keep the same properties of the backhaul network, as explained in Chapter 2 (Section 2.6, p. 16). To summarize, we consider that, regardless of the wireless technology used for inter-BS links, there is no contention among them for resource utilization. We assume that interfering wireless links are operating on distinct channels, allowing parallel transmissions on the different links, with no interference [49]. We suppose that the backhaul bandwidth is limited, meaning the amount of traffic that can be routed on these links from both data flows and signaling messages is limited.

4.2.1 Mathematical notation

Let $G(\mathcal{J}, \mathcal{E})$ be an undirected graph modeling the network, with $|\mathcal{J}| = n$ nodes. Each BS is a node of the graph, and the inter-BS links are the graph edges. The BS co-located with the Local CN serving the network, denoted by d, is considered as the destination node in the graph, while all the other BSs in the network are sources. Let $\mathcal{O}, \mathcal{D} \subseteq \mathcal{J}$ be the set of sources and destinations

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

of graph G, respectively. In this case, we have $\mathcal{O} = \mathcal{J} \setminus \{d\}$, and $\mathcal{D} = \{d\}$. To model the inter-BS links with limited bandwidth, we consider graph edges with limited capacities. The capacity of an edge represents the maximum amount of traffic that can pass through that edge (in units of traffic). We denote by $c(u, v)$ the capacity of the edge $(u, v) \in \mathcal{E}$. The flow through this edge is denoted by $f(u, v)$. All data and signaling traffic from a BS is forwarded to the Local CN. Hence, all the BSs in the network have an amount of traffic to route on the backhaul links connecting them to the Local CN. To model this traffic, we define the supply function z, such that $z(v, d)$ is the flow that a source node $v \in \mathcal{O}$ sends towards the destination d. We suppose that BSs do not have limited capacity constraints.

4.3 Local CN placement problem

As one of the BSs must be co-located with a Local CN serving the other BSs in the network, the centralized Local CN placement problem questions with which of the BSs must the Local CN be co-located. With the objective of maximizing the traffic received by the Local CN from all the BSs in the network, a logical reasoning is to place the Local CN on a "central" node. However, the literature is rich with different definitions of what is "centrality" in a network [125]. Indeed, there are several ways to measure node centrality, each one highlighting a different characteristic of the node, and its relative importance in the network [125].

In this section, we first present a non-exhaustive list of the main centrality metrics, highlighting the reason why they might not be adequate for a Local CN placement. Then, we illustrate a simple numerical example that justifies the need for a novel centrality metric.

4.3.1 Centrality in networks

1. **The degree centrality** of a node is equal to the degree of the node, i.e., the number of links the node has with other nodes [126]. While the degree centrality gives an idea on the node connectivity in the network, it does not take into consideration the limited link capacities. Hence, a node with the maximum degree centrality is not necessarily capable of receiving high amount of flows.

2. **The weighted degree** is defined as the sum of the weights (e.g., capacity) of the links connecting the node to its direct neighbors.

$$W(u) = \sum_{v \in \mathcal{J}} c(u, v) \tag{4.1}$$

   The node with the maximum weighted degree is potentially capable of receiving the maximum amount of traffic. However, this traffic is not necessarily achievable, since the amount of flows actually received depends on the whole network topology, and on the other potentially limited links routing these flows.

3. **The closeness centrality** of a node measures how *close* a node is to all the other nodes in the network [127]. Formally, the closeness centrality of a node is defined as the reciprocal of the sum of all the shortest distances from that node to all the other nodes in the network. If we denote by $\delta(u, v)$ the shortest path between nodes u and v, then the closeness centrality is defined as:

$$C(u) = \frac{1}{\sum_{v \in \mathcal{J}} \delta(u, v)} \tag{4.2}$$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

The node with the maximum closeness centrality is a geographically central node. This position gives the node a relative advantage for easily communicating with all the other nodes in the network, making it seem the most adequate for the Local CN placement. However, the limited link capacities leading to this central node could limit the amount of traffic it is capable of receiving.

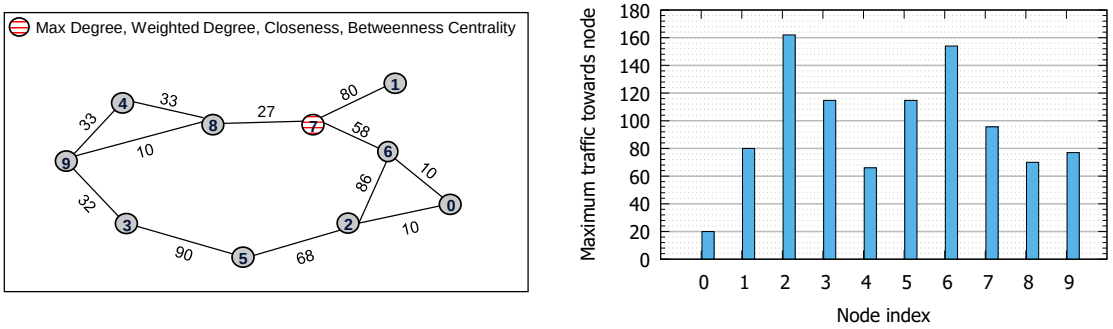
4. **The betweenness centrality** of a node quantifies the number of times a node belongs to the shortest path between all the pairs of nodes in the graph [86]. Let $\alpha(v, w)$ be the total number of shortest paths from node v to node w, and $\beta(u, v, w)$ the number of those paths that pass through node u, then the betweenness centrality of a node u is defined as:

$$B(u) = \sum_{v,w \in \mathcal{J}} \frac{\beta(u, v, w)}{\alpha(v, w)} \tag{4.3}$$

The node with the maximum betweenness centrality can be seen as an anchor node, well placed for forwarding communications between the nodes. However, once again, the limitation of this centrality, in our case, resides in the fact that link capacities around the node with the maximum betweenness centrality may limit its ability to forward traffic.

4.3.2 Numerical example

In the following, we consider an example of a self-deployable network, with BSs served by a Local CN, and verify whether placing the Local CN on a node maximizing one of the above-mentioned centrality metrics is the best option, or there is a way to do better.



(a) A network topology example, with the numbers on the links representing the link capacity $c(u, v)$.

(b) Traffic received by the node co-located with the Local CN from all the other nodes in the network

Figure 4.2: Limitations of state of the art centrality metrics.

We consider the topology in Fig. 4.2a, with 10 nodes and randomly distributed link capacities such that $c \in [0, 100]$ (in units of traffic). The node that maximizes the degree, the weighted degree, the closeness and the betweenness centrality is the node 7. To verify whether a placement of the Local CN on the node 7 is suitable, we compare the maximum traffic received from all the BSs in the network by each node, if the Local CN is placed on that node. Numerical results are shown in Fig. 4.2b. We notice that placing the Local CN on node 7 is not the best option, since there are multiple nodes that achieve better results. For instance, it is node 2 that maximizes the traffic that can be received by the Local CN. Placing the Local CN on node 7 instead of node 2 would cause a significant relative traffic loss equal to 41%.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

From this observation, we deduce that there may be a node in the network more suitable for the Local CN placement than a node maximizing one of the centrality metrics. Therefore, in Section 4.4, we formally characterize that node by proposing a novel centrality metric that we call flow centrality, such that the Local CN should be placed on the node maximizing this metric. Then, in Section 4.7, we benchmark the proposed metric against the state of the art centralities.

4.4 Flow centrality: a metric for local CN placement

As shown in the previous section, conventional centrality metrics are not always a suitable option for the Local CN placement. In this section, we first list the key placement criteria in a self-deployable mobile network. In order to meet those criteria, we propose a novel centrality metric in a network, called flow centrality, such that a Local CN is placed on the node having the maximum flow centrality. Then, we formulate in details the different computation methods for this metric.

4.4.1 Local CN placement criteria

The CN must be able to handle the entire user traffic, meaning it must be able to receive all the data and signaling traffic generated by the BSs. Traffic is forwarded by intermediate BSs towards the Local CN. The limited inter-BS links capacities set a threshold on the amount of traffic that can be transported. In this case, the load of the inter-BS links is upper bounded by their respective limited capacities, creating a bottleneck. Consequently, the placement of the Local CN must take into consideration the links between the BSs and their capacity, in order to ensure that all user traffic can circulate in the network without losses.

Recall that, in this work, we do not consider a placement based on the number of users in the network and their requests, due to the lack of such information at the early deployment phase. We propose to statically place the Local CN at first, in a general way, by treating all BSs as equals and maximizing the amount of traffic they can send, such that the achievable traffic is the same for each BS.

4.4.2 Flow centrality

We propose flow centrality as a novel centrality metric measuring the capacity of a node in receiving the total amount of flows in the graph by accounting for the limited backhaul bandwidth [124]. To compute the flow centrality of node d, it is sufficient to compute the maximum amount of traffic that node d can receive from all the other nodes in the graph if it were the unique destination node. The total amount of flows received at destination node d must be equal to the total flows sent by all the sources. As each node $v \in \mathcal{J}$ has a supply $z(v,d) = \rho(d)$ to send towards d, then the total flow value received at node d is:

$$|f|_d = \sum_{v \in \mathcal{O}} z(v,d) = (n-1) \cdot \rho(d) \tag{4.4}$$

We denote by $\bar{\rho}(d)$ the maximum achievable value of $\rho(d)$. When the supply $\rho(d)$ is maximized, the total flow received at d is maximum, such that:

$$|f|_{d_{max}} = (n-1) \cdot \bar{\rho}(d) \tag{4.5}$$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Eventually, the flow centrality of a node d is defined as $\bar{\rho}(d)$. The maximum flow centrality is then expressed as:

$$\rho_{max} = \max_{d \in \mathcal{J}} \left(\bar{\rho}(d) \right) \tag{4.6}$$

Practically, this new metric distinguishes the node that, when set as destination, is capable of receiving the largest amount of flows, equally from all the other nodes, in comparison to when other nodes are set as destination. Hence, in order to better serve the network, the Local CN must be co-located with the BS having the maximum flow centrality.

4.4.3 Computing flow centrality

To compute the flow centrality of a node d, we first suppose that node d is the unique destination of the graph, and all the other nodes are sources that can send the same amount of traffic $\rho(d)$ towards node d. As stated previously, the flow centrality of a node d is defined as $\bar{\rho}(d)$, the maximum achievable value of $\rho(d)$. In the following, we propose two approaches for the flow centrality computation, one based on solving a linear optimization problem, and the other based on a minimum cut problem.

4.4.3.1 Linear optimization problem

One way to compute the flow centrality of a node is through a linear optimization problem. The objective in Eq. 4.7 maximizes $\rho(d)$. The constraint 4.8 fixes the supply of all the sources in the graph to $\rho(d)$. The constraint 4.9 ensures that the flow on a edge in the graph does not surpass the edge capacity. Constraint 4.10 concerns flow conservation, such that the total flows entering a node must be equal to the flows exiting a node. The constraint 4.11 makes sure the total flow value received at the destination node is equal to the sum of all the supplies of the sources. The problem is formulated as follows:

$$\max_{\rho(d) \in [0,\infty)} \rho(d) \tag{4.7}$$

$$z(v,d) = \rho(d), \ \forall \ v \in \mathcal{O} \tag{4.8}$$

$$f(u,v) \leq c(u,v), \ \forall \ (u,v) \in \mathcal{E} \tag{4.9}$$

$$\sum_{u \in \mathcal{O}} f(u,v) = \sum_{w \in \mathcal{O}} f(v,w), \ \forall \ v \in \mathcal{O} \tag{4.10}$$

$$\sum_{v \in \mathcal{O}} f(v,d) = (n-1) \cdot \rho(d) \tag{4.11}$$

After computing the flow centrality of each of the nodes, the node d^* with the maximum flow centrality is the node for which $\bar{\rho}(d)$ is maximum, such that:

$$d^* = \arg\max_{d \in \mathcal{J}} \left(\bar{\rho}(d) \right) \tag{4.12}$$

4.4.3.2 Finding minimum cut value

Another approach to computing the flow centrality of a node is through the minimum cut problem, i.e., finding the cut of minimum value that separates the node from the others, as detailed in the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

following. A cut in the graph is a partition of the graph nodes into two non-empty disjoint subsets, joined by at least one edge.

We denote by $\mathcal{C} = (\mathcal{A}, \mathcal{B})$ a cut that partitions $\mathcal{J}$ into 2 subsets $\mathcal{A}, \mathcal{B} \subset \mathcal{J}$, such that $d \in B$. We define the size of a cut as $|\mathcal{C}| = |\mathcal{A}|$, i.e., the number of nodes in the subset not containing the destination. We define the capacity of a cut, $\gamma(\mathcal{C})$, as the sum of the capacities of the edges traversed by that cut, such that:

$$\gamma(\mathcal{C}) = \sum_{(u,v)\in E: u\in A, v\in B} c(u,v) \tag{4.13}$$

We define the traffic of a cut, $\upsilon(\mathcal{C})$, as the sum of the traffic sent by nodes $v \in \mathcal{A}$ towards the destination node $d \in \mathcal{B}$, such that:

$$\upsilon(\mathcal{C}) = \sum_{v\in\mathcal{A}} z(v,d) = \rho(d) \cdot |\mathcal{C}| \tag{4.14}$$

The capacity of a cut $\gamma(\mathcal{C})$ must be able to support the traffic of a cut $\upsilon(\mathcal{C})$, hence:

$$\gamma(\mathcal{C}) \geq \upsilon(\mathcal{C}) \tag{4.15}$$

Which leads to:

$$\rho(d) \leq \frac{\gamma(\mathcal{C})}{|\mathcal{C}|} \tag{4.16}$$

We refer to the ratio $\frac{\gamma(\mathcal{C})}{|\mathcal{C}|}$ as the value of a cut $\mathcal{C}$. Finding $\bar{\rho}(d)$ is equivalent to finding the cut $\mathcal{C} = (\mathcal{A}, \mathcal{B})$, with the minimum cut value. Therefore:

$$\bar{\rho}(d) = \min_{\mathcal{C}} \left(\frac{\gamma(\mathcal{C})}{|\mathcal{C}|} \right) \tag{4.17}$$

4.5 Computing flow centrality in canonical topologies

We continue in this section to focus on the flow centrality metric computation. As previously detailed in Section 4.4.3, computing the flow centrality of a node requires solving a linear optimization problem, or finding a minimum cut. In the following, we show how in some particular network topologies, such as path graphs and balanced trees, some properties emerge, allowing the computation of the flow centrality of a node via analytical expressions.

4.5.1 Path graph

A path graph is a trail of distinct nodes (Fig. 4.3). It is a sequence of n nodes, denoted by $v_0, v_1, ..., v_{n-1}$. Adjacent nodes, v_i and v_{i+1}, are linked with an edge of capacity $c_i = c(v_i, v_{i+1})$, $\forall\, i \in [0, n-2]$. We consider that link capacities follow a certain distribution, such that $c_i \in [c_{min}, c_{max}]$. In a path graph, a node v_i has i nodes to its left, and $n-i-1$ nodes to its right. The flow centrality of a node v_i on the path is equal to the maximum traffic $\bar{\rho}(v_i)$ that can be generated by all the other nodes on the path, and directed towards this node.

For a node $v_i \neq v_0$, each node v_k to the left of node v_i, such that $k \in [0, i-1]$, has a traffic $\rho(v_i)$ to send towards v_i. Consequently, each edge (v_k, v_{k+1}) of capacity c_k is responsible of forwarding the totality of the traffic generated by the $k+1$ nodes on its left towards node v_k. Thus, the total

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

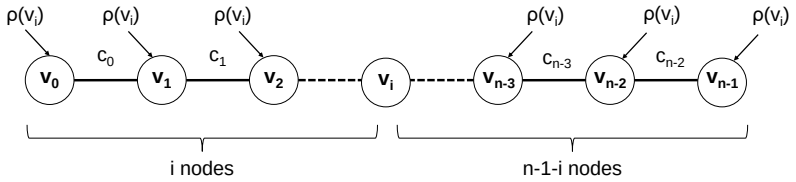


Figure 4.3: Path graph with n nodes.

traffic generated by the $k+1$ nodes arriving at each edge (v_k, v_{k+1}) is upper-bounded by the limited capacity c_k of that edge, such that:

$$\rho(v_i) \leq \frac{c_k}{k+1} \ , \ \forall \ k \in [0, i-1] \tag{4.18}$$

Similarly, for a node $v_i \neq v_{n-1}$, each node v_j to the right of node v_i, such that $j \in [i+1, n-1]$, has a traffic $\rho(v_i)$ to send towards v_i. Consequently, each edge (v_{j-1}, v_j) of capacity c_{j-1} is responsible of forwarding the totality of the traffic generated by the $n-j$ nodes on its right towards node v_j. Thus, the total traffic generated by the $n-j$ nodes arriving at each edge (v_{j-1}, v_j) is upper-bounded by the limited capacity c_{j-1} of that edge, such that:

$$\rho(v_i) \leq \frac{c_{j-1}}{n-j} \ , \ \forall \ j \in [i+1, n-1] \tag{4.19}$$

Since both expressions in Eq. 4.18 and Eq. 4.19 upper bounding $\rho(v_i)$ must be verified, the flow centrality of a node v_i, i.e., the maximum value of $\rho(v_i)$, is:

$$\bar{\rho}(v_i) = \begin{cases} \min_{j \in [1, n-1]} \left(\frac{c_{j-1}}{n-j} \right) & \text{if } i = 0 \\ \min\limits_{\substack{k \in [0, i-1] \\ j \in [i+1, n-1]}} \left(\frac{c_k}{k+1}, \frac{c_{j-1}}{n-j} \right) & \text{if } 0 < i < n-1 \\ \min_{k \in [0, n-2]} \left(\frac{c_k}{k+1} \right) & \text{if } i = n-1 \end{cases} \tag{4.20}$$

We consider now the particular case where links in the path graph have uniform capacities, i.e., $c(v_i, v_{i+1}) = c \ , \ \forall \ i \in [0, n-2]$. Based on Eq. 4.20, we get:

$$\begin{aligned} \bar{\rho}(v_i) &= \min_{\substack{k \in [0, i-1] \\ j \in [i+1, n-1]}} \left(\frac{c}{k+1}, \frac{c}{n-j} \right) \\ &= \min \left(c, \frac{c}{2}, ..., \frac{c}{i}, \frac{c}{n-i-1}, \frac{c}{n-i-2}, ..., \frac{c}{2}, c \right) \\ &= \min \left(\frac{c}{i}, \frac{c}{n-i-1} \right) \end{aligned} \tag{4.21}$$

From Eq. 4.21, we can deduce that, in a path graph with uniform capacities, the node(s) at the center of the path is (are) always the node(s) with maximum flow centrality. When the number of nodes is odd, i.e., $n = 2x+1$, the maximum value of $\bar{\rho}(v_i)$ is achieved for $i = x$. Thus, the node with the maximum flow centrality is node v_x. When the number of nodes is even, i.e., $n = 2x$, the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

maximum value of $\bar{\rho}(v_i)$ is achieved for $i = x - 1$ and $i = x$. Thus, two nodes have the maximum flow centrality: v_{x-1} and v_x, respectively. In both cases, the maximum flow centrality is:

$$\rho_{max} = \frac{c}{x} \tag{4.22}$$

4.5.2 Balanced tree

A balanced tree is a rooted tree having all its leaves at a distance h from the root. The left and right subtrees of any node have the same height. In a tree with a branching factor τ, all nodes, except the leaves, are branched into τ subtrees. All the nodes have a degree $\tau + 1$, except the root, which has a degree τ, and the leaves, which have a degree 1. Fig. 4.4 shows an example of a balanced tree of height $h = 4$ and a branching factor $\tau = 2$.

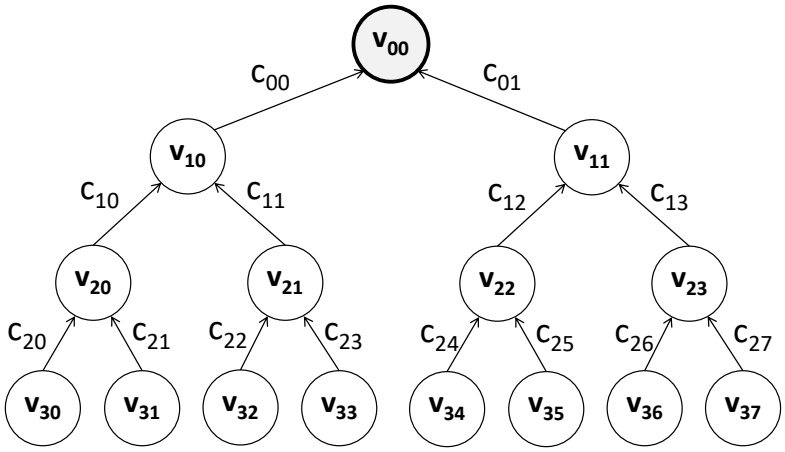


Figure 4.4: A balanced tree rooted at v_{00}, with height $h = 4$ and branching factor $\tau = 2$.

Let i be the depth of a node within the tree, such that $i \in [0, h-1]$. The node at depth $i = 0$ is the root, and nodes at depth $i = h - 1$ are the leaves. At each depth $i \in [0, h-1]$, there are τ^i nodes, denoted by v_{ij}, such that $j \in [0, \tau^i - 1]$. Each node v_{ij} is linked to its parent via a single link, of capacity $c_{i-1,j}$. The number of descendants of a node at depth i, denoted by $des(i)$, is:

$$des(i) = \begin{cases} 0 & \text{if } i = h-1 \\ \sum_{k=1}^{h-1-i} \tau^k & \text{if } 0 \leq i < h-1 \end{cases} \tag{4.23}$$

If we suppose that all nodes in a tree have traffic to forward towards one destination node, then each link exiting at a node must be able to forward the total traffic aggregated by this node, i.e., its traffic and the traffic of all its descendants. Let us consider the example shown in Fig. 4.5 of a node v_{ij} at a depth i, in a balanced tree of branching factor τ. The node v_{ij} is branched into τ symmetrical subtrees, and has $des(i)$ descendants in total (Eq. 4.23), divided equally among the subtrees. Thus, each subtree has $\frac{des(i)}{\tau}$ nodes.

Each link with a capacity $c_{i,k}$ exiting one of the subtrees and going towards v_{ij} is responsible of forwarding the total traffic of all the nodes in that subtree. If we denote by ρ the traffic of each

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

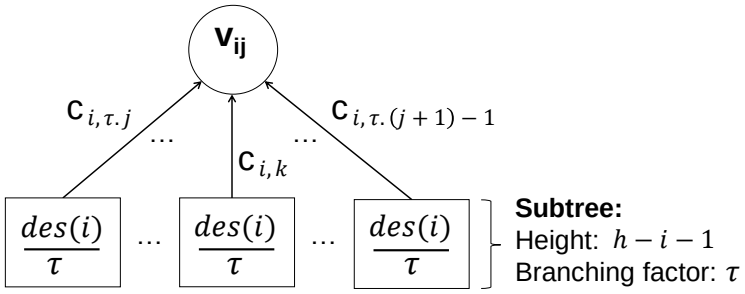


Figure 4.5: A node v_{ij} at depth i within a balanced tree, branched into τ symmetrical subtrees, each having a height $h-i-1$, and comprising $des(i)/\tau$ nodes.

node, then the following expression must be verified:

$$\rho \leq \tau \frac{c_{i,k}}{des(i)} \ , \ \forall \ k \in [\tau \cdot j, \tau(j+1)-1] \tag{4.24}$$

In the following, we formulate analytical expressions to compute the flow centrality of nodes in a balanced tree of height h and branching factor τ. The flow centrality value of a node depends on the depth i at which the node is within the tree. As an illustrative example, we consider the balanced tree with height $h=4$ and branching factor $\tau=2$, shown in Fig. 4.4.

A first step in flow centrality computation of a node v_{ij} is to consider that this node is the unique destination in the tree. In the following, we first explain the analytical expressions of the flow centrality of nodes at depths $i=0$ and $i=1$, before deducing the general expression for a node at depth i.

- **Node at depth** $i=0$

  When computing the flow centrality of the root v_{00} at depth $i=0$, all the nodes in the tree, except the leaves, are branched into τ symmetrical subtrees, and match the case shown in Fig. 4.5. Consequently, the expression in Eq. 4.24 applies to all the links in the tree, such that:

$$\rho \leq \tau \frac{c_{i,k}}{des(i)} \ , \ \forall \ i \in [0, h-2] \ , \ \forall \ k \in [0, \tau^{i+1}-1] \tag{4.25}$$

  From Eq. 4.25, we deduce that the flow centrality of node v_{00} is:

$$\bar{\rho}(v_{00}) = \min_{\substack{i \in [0,h-2] \\ k \in [0,\tau^{i+1}-1]}} \left(\tau \frac{c_{i,k}}{des(i)} \right) \tag{4.26}$$

- **Node at depth** $i=1$

  We compute in the following the flow centrality of nodes at depth $i=1$. As an example, we first compute the flow centrality of node v_{10}. We consider that v_{10} is the unique destination node in the tree. Fig. 4.6 shows the corresponding tree structure. In this case, node v_{10} is branched into $\tau+1$ subtrees:

  – τ balanced subtrees that encompass all the descendants of v_{10} (in the original tree), whose number is equal to $des(1)$. Each of the subtrees has $\frac{des(1)}{\tau}$ nodes.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

– One subtree, denoted $\mathcal{X}$ (see Fig. 4.6), that encompasses v_{00} and all its descendants (in the original tree), except the subtree started by v_{10}. Thus, the number of nodes in $\mathcal{X}$ is $n - \frac{des(0)}{\tau}$. $\mathcal{X}$ is linked to v_{10} via a link of capacity $c_{0,0}$. This link is responsible of forwarding the total traffic of all the nodes in $\mathcal{X}$. If we denote by ρ the traffic forwarded by each node, then the following expression must be verified:

$$\rho \leq \frac{c_{0,0}}{n - \frac{des(0)}{\tau}} \tag{4.27}$$

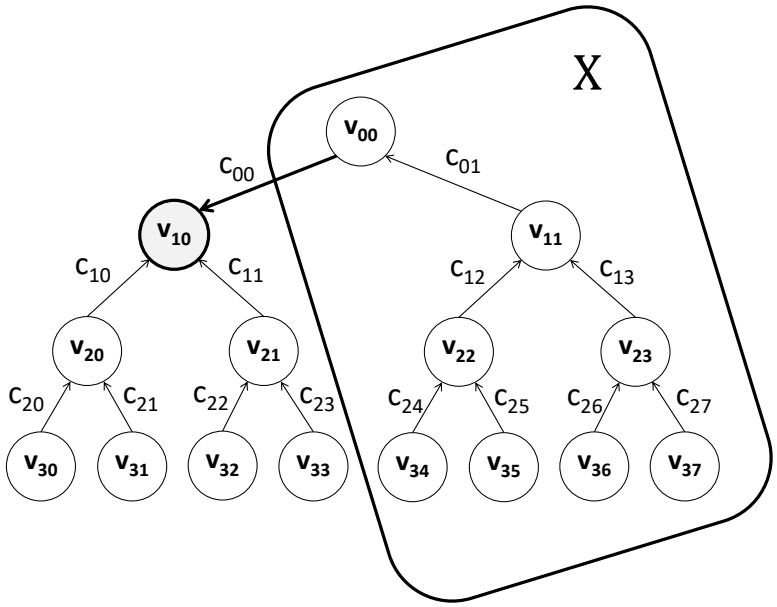


Figure 4.6: Tree structure with node v_{10} set as unique destination in the tree for flow centrality computation.

All the nodes in the tree in Fig. 4.6, except v_{00}, v_{10} and the leaves, are branched into τ symmetrical subtrees, and match the case shown in Fig. 4.5. Consequently, the expression in Eq. 4.24 applies to all of the links in the tree except $c_{0,0}$. Following the same reasoning as Eq. 4.24, we deduce:

$$\rho \leq \tau \frac{c_{i,k}}{des(i)} \ , \ \forall \ i \in [0, h-2] \ , \ \forall \ k \in [0, \tau^{i+1} - 1], \ (i,k) \neq (0,0) \tag{4.28}$$

From Eq. 4.27 and Eq. 4.28, we deduce that the flow centrality of node v_{10} is:

$$\bar{\rho}(v_{10}) = \min_{\substack{i \in [0,h-2] \\ k \in [0,\tau^{i+1}-1] \\ (i,k) \neq (0,0)}} \left(\frac{c_{0,0}}{n - \frac{des(0)}{\tau}}, \tau \frac{c_{i,k}}{des(i)} \right) \tag{4.29}$$

This could be generalized to all nodes v_{1j} at depth $i = 1$, with $j \in [0, \tau - 1]$. The flow centrality of node v_{1j} is:

$$\bar{\rho}(v_{1j}) = \min_{\substack{i \in [0,h-2] \\ k \in [0,\tau^{i+1}-1] \\ (i,k) \neq (0,j)}} \left(\frac{c_{0,j}}{n - \frac{des(0)}{\tau}}, \tau \frac{c_{i,k}}{des(i)} \right) \ , \ \forall \ j \in [0, \tau - 1] \tag{4.30}$$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

- **General case, at depth i**

  To generalize, we follow the same reasoning at each depth $i \in [1, h-1]$. We define $\lfloor x \rfloor$ as the floor function that gives the greatest integer less than or equal to x. To compute the flow centrality of a node v_{ij} at depth i, this node is considered as unique destination. All the nodes in the tree, except those lying on the path between v_{ij} and the original root v_{00} are branched into τ symmetrical subtrees, and match the case shown in Fig. 4.5. Consequently, the expression in Eq. 4.24 applies to all of the links in the tree except the following links that are on the path between v_{ij} and the original root v_{00}: $\left\{(0, \lfloor \frac{j}{\tau^{i-1}} \rfloor), (1, \lfloor \frac{j}{\tau^{i-2}} \rfloor), ..., (i-2, \lfloor \frac{j}{\tau} \rfloor), (i-1, j)\right\}$. Let $w(k) = \lfloor \frac{j}{\tau^{i-1-k}} \rfloor$. Each link $(k, w(k))$ is responsible of forwarding the total traffic of $n - \frac{des(k)}{\tau}$ nodes.

  Eventually, the flow centrality of a node v_{ij} is:

$$\bar{\rho}(v_{ij}) = \min_{\substack{p\in[0,h-2] \\ q\in[0,\tau^{p+1}-1] \\ k\in[0,i-1] \\ (p,q)\neq(k,w(k))}} \left(\frac{c_{k,w(k)}}{n - \frac{des(k)}{\tau}}, \tau \frac{c_{p,q}}{des(p)} \right) , \ \forall\, i \in [1, h-1] \ , \ \forall\, j \in [0, \tau^i - 1] \quad (4.31)$$

We consider now the particular case where all links in the tree have uniform capacities. In this case, the flow centrality values of nodes at the same depth are identical. We consider at first the node v_{00}. Since $des(0) > des(k),\ \forall\, k \in [1, h-2]$, we deduce from Eq. 4.26:

$$\bar{\rho}(v_{00}) = \tau \frac{c}{des(0)} = \tau \frac{c}{n-1}. \quad (4.32)$$

For nodes v_{ij},$\forall\, i \in [1, h-1]$, $\forall\, j \in [0, \tau^i - 1]$, the flow centrality is:

$$\bar{\rho}(v_{ij}) = \min_{\substack{p\in[0,h-2] \\ k\in[0,i-1]}} \left(\frac{c}{n - \frac{des(k)}{\tau}}, \tau \frac{c}{des(p)} \right) = \frac{c}{n - \frac{des(i-1)}{\tau}} \quad (4.33)$$

However, $\tau \frac{c}{n-1} > \frac{c}{n - \frac{des(i-1)}{\tau}}$, $\forall\, i \in [1, h-2]$. Consequently, the maximum flow centrality is:

$$\rho_{max} = \max_{i\in[0,h-1]} \left(\bar{\rho}(v_{ij})\right) = \bar{\rho}(v_{00}) = \tau \frac{c}{n-1} \quad (4.34)$$

Thus, the root v_{00} of a balanced tree with uniform capacities is the node with the maximum flow centrality.

4.6 Flow centrality properties

After detailing the computation methods of the flow centrality metric, we provide in this section some insights on its properties. By varying the inter-BS backhaul links characteristics, i.e, the link capacities distribution, the average link capacity, and the link capacities range, we deduce several properties on the dependence between the backhaul network, the flow centrality values, and the position of the node with the maximum flow centrality.

We model self-deployable networks using random geometric graphs, i.e., graphs whose nodes are randomly distributed, with an edge existing between two nodes if and only if the distance

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

between them is smaller than a certain radius, denoted η. The reason we use geometric graphs to model the backhaul network is to determine which BSs are directly connected via a backhaul link, in a general way, independently of the used backhaul technology. In a geometric graph, this is done by relying on a generic metric, which is the distance between the BSs.

Let us consider an example of a self-deployable network, with BSs served by a Local CN. We represent in Fig. 4.7 a network topology, based on a random geometric graph with 10 nodes, in a surface of one unit square, with a network radius $\eta = 0.2$, and randomly distributed link capacities such that $c(u, v) \in [0, 100]$ (in units of traffic). A null capacity $c(u, v) = 0$ means that the link is not usable.

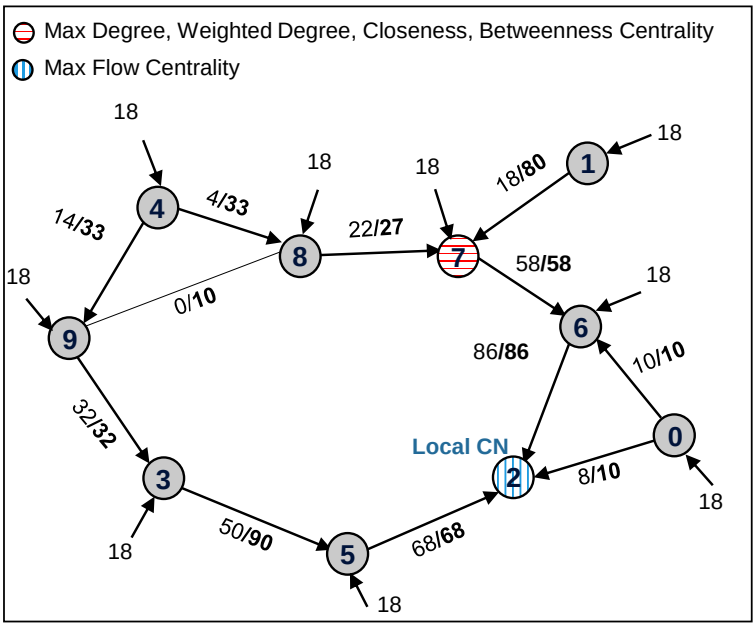


Figure 4.7: A random geometric graph topology, with random link capacities $c(u, v) \in [0, 100]$. All nodes transmit $\rho_{max} = 18$ units of traffic towards the Local CN co-located with node 2. The values on the links represent $f(u, v)/c(u, v)$.

We compute the flow centrality value of each node by solving the linear optimization problem in Section 4.4.3, using the commercial solver "CPLEX" [128]. Results show that node 2 has the maximum flow centrality, such that $\rho_{max} = \bar{\rho}(2) = 18$. This means that the Local CN in this network must be co-located with node 2.

For this network with a relatively small number of nodes, the computation time is in the order of milliseconds. This small number of nodes correspond to the small size of self-deployable networks, where only a few BSs are needed. For a more general evaluation of the flow centrality metric, larger scale networks are further studied in Section 4.8. We note that with the increase of the number of nodes, the computation time increases accordingly, but remains in the order of seconds even for networks with up to 100 nodes.

4.6.1 Position of the node with maximum flow centrality

Following an extensive study of different topologies, with different link capacities distribution, some guidelines emerge on the position of the node with the maximum flow centrality. Indeed, nodes at the network edges, usually linked to fewer nodes than the others, have lower flow cen-

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

trality values due to their "isolated" position. Such links are not central, nor have the maximum flow centrality. On the other hand, a node having a central position, i.e., maximizing the closeness centrality, or well served by links, i.e., maximizing the weighted degree, but that has one (or more) link(s) used only by a single node, cannot be the node with the maximum flow centrality. This can be seen in the example shown in Fig. 4.7, where we identify node 7 as the node having the maximum weighted degree centrality, the maximum betweenness centrality, the maximum closeness centrality, and the maximum degree centrality, simultaneously. However, having the link $(1, 7)$ of relatively high capacity only used by node 1 prevents this node from having the maximum flow centrality. In this example, it is node 2 that has the maximum flow centrality and not node 7.

4.6.2 Link capacities distribution

In Fig. 4.7, we show on each link (u, v), the routed traffic $f(u, v)$ in comparison with the maximum link capacity $c(u, v)$, when the Local CN is placed at node 2, and each node sends a traffic equal to ρ_{max} towards the Local CN. We can notice that some links are saturated, while others are under-used.

The saturated links are mostly the ones that play a major role in upper bounding the value of ρ_{max}. In Fig. 4.7, these are the links $(9, 3)$, $(5, 2)$, $(7, 6)$ and $(6, 2)$. For this particular topology, and this particular capacity distribution, those are the links that control the value of ρ_{max}. We observe that by increasing (decreasing) the capacity of those links, the value of ρ_{max} increases (decreases) respectively.

Other links, such as $(0, 6)$ and $(0, 2)$, which are only responsible of delivering the flow of one node, have enough capacity in this case to not upper bound the value of ρ_{max}. We should note that, since node 0 is directly linked to node 2 co-located with the Local CN, it is practically capable of forwarding more traffic than ρ_{max}, without affecting the other nodes traffic, as long as $c(0, 2)$ allows it.

On the other hand, link $(8, 9)$, for example, is not used at all. Removing this link would not affect the position of the node with the maximum flow centrality, nor its value. The capacity of link $(1, 7)$, which is only used to route traffic from node 1, is under-used. However, increasing the traffic of node 1 renders the total flow unfeasible, due to other saturated links leading to destination node. Moreover, due to the position of node 1 in the network, increasing the capacity of link $(1, 7)$ would not have any effect on the value of the maximum flow centrality.

Therefore, even for the same network topology, the nodes flow centrality values and the position of the node with the maximum flow centrality are impacted by the link capacities distribution.

4.6.3 Link capacities average

From the previous results, we know that the value of the flow centrality of a node d, $\bar{\rho}(d)$, is significantly affected by the links capacities. In fact, as shown in Section 4.4.2 and illustrated in the previous example, the value of $\bar{\rho}(d)$ of a node d is dictated by the capacity of the links leading to that node, as well as by the capacity of all the links in the graph. In this section, we further highlight this proportional relation.

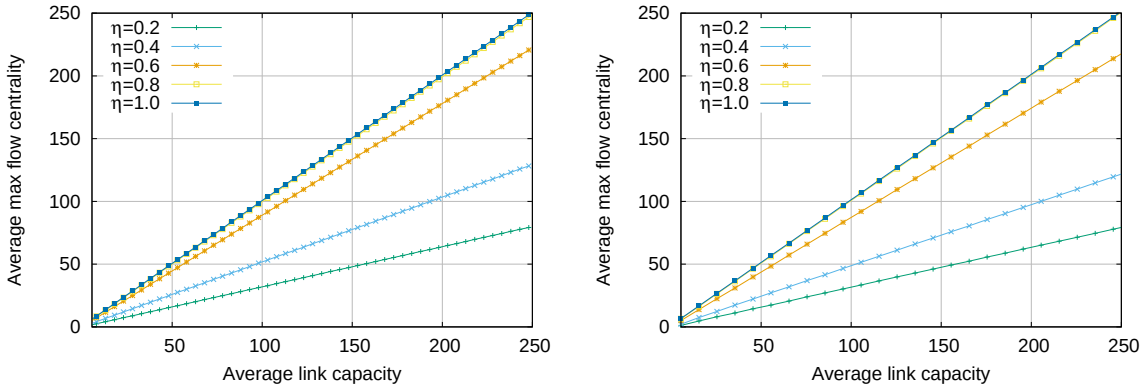
We consider a sample of random geometric graphs of radius η, with uniform link capacities distribution such that $c(u, v) \in [c_{min}, c_{max}]$. We denote by c_{avg} the average link capacity, such that $c_{avg} = \frac{c_{min}+c_{max}}{2}$, and by Δc the capacity range, such that $\Delta c = c_{max} - c_{min}$. To highlight the existing relation between the maximum flow centrality ρ_{max} and the link capacities, we show in Fig. 4.8 the average value of ρ_{max} function of the average link capacity c_{avg}. We vary the values of c_{avg}, but keep a constant interval Δc. We show this variation for different values of the graph

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

radius η. We remind that the more η increases, the higher the number of links is in the network.

Results show that the average value of ρ_{max} increases almost linearly with c_{avg}. For $\eta = 1$, where almost all pairs of nodes are connected, the value of ρ_{max} is approximately equal to c_{avg}. This suggests that the value of the maximum flow centrality is upper bounded by the average link capacity. Moreover, as the radius η decreases, the number of edges decreases, and the value of ρ_{max} decreases accordingly.

Indeed, for the same value of c_{avg}, the less there are links in the network, the less traffic can circulate. Furthermore, the linear relation between c_{avg} and ρ_{max} comes in accordance with the analytical expressions obtained in Section 4.5 that show a linear relation between the maximum flow centrality and the link capacities.



(a) $\Delta c = 5$ (b) $\Delta c = 10$

Figure 4.8: Variation of the average maximum flow centrality ρ_{max} function of the average link capacity c_{avg}, in random geometric graphs, for different values of the radius τ, with a constant capacity range Δc.

4.6.4 Link capacities range

We observe now the variation of ρ_{max} function of c_{avg}, for different values of capacity intervals Δc. Fig. 4.8a and Fig. 4.8b correspond to two capacity intervals, $\Delta c = 5$ and $\Delta c = 10$, respectively. We notice that the value of ρ_{max} and its variation depending on c_{avg} are very similar in both cases. Thus, while the value of ρ_{max} depends on the average link capacity, it is not significantly affected by the capacity range.

However, we show in the following that the capacity range does have an impact on the position of the node with the maximum flow centrality. Indeed, changing the capacity range also changes the link capacities distribution. As discussed in Section 4.6.2, link capacities distribution do impact the node with the maximum flow centrality. For example, we consider the network topology in Fig. 4.9. While the nodes distribution is identical to the one in Fig. 4.7, the link capacities, and more specifically the capacity range Δc, are different. In Fig. 4.9, $\Delta c = 0$ such that all links (u, v) have the same capacity $c(u, v) = 50$ (as opposed to $\Delta c = 100$ in Fig. 4.7). In this network, nodes 2, 6, 8, and 9 all have the maximum flow centrality (as opposed to only node 2 in Fig. 4.7). We note that the flow distribution on the different links, represented by the $f(u, v)/c(u, v)$ notations in Fig. 4.9, correspond to the case where node 2 is set as destination node and co-located with the Local CN.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

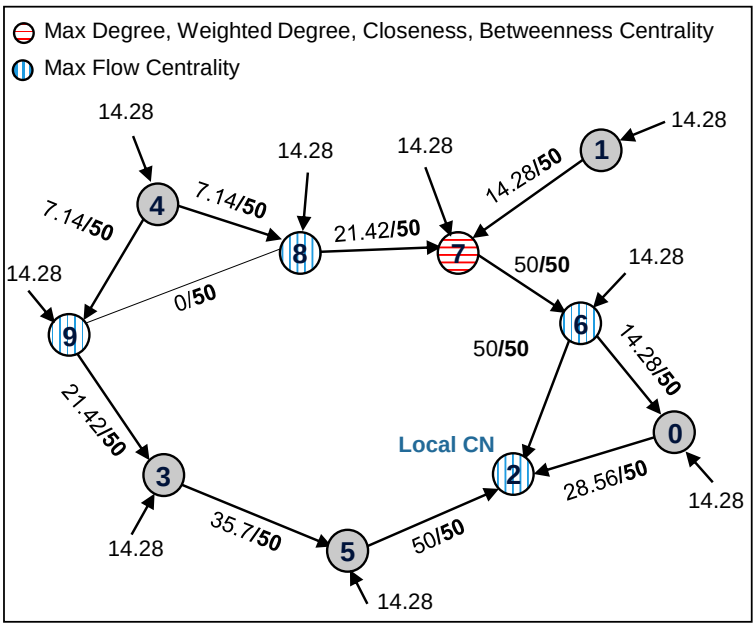


Figure 4.9: A random geometric graph topology, with link capacities $c(u,v) = 50$. All nodes transmit $\rho_{max} = 14.28$ units of traffic towards the Local CN co-located with node 2. The values on the links represent $f(u,v)/c(u,v)$.

4.7 Benchmarking flow centrality

In this section, we benchmark the proposed flow centrality metric against the state of the art centralities previously presented in Section 4.3. Moreover, we also compare our metric to another metric from literature, proposed in [90] for the placement of a gateway in a wireless mesh network, referred to hereafter as the weighted path (see Chapter 2, Section 2.9, p. 22). In the weighted path metric computation, a link weight represents the likelihood of transmission failure on that link. We represent it in our case by the link capacity. For each node, the sum of the link weights of the shortest paths from all the other nodes toward this node is computed. Then, the destination node that minimizes this sum is selected for the placement.

Since we focus on the Local CN placement problem, we are mostly interested in checking if the node having the maximum flow centrality is the same as the node maximizing one of the other centralities. We denote by **matching percentage** the percentage of cases where the node with maximum flow centrality is identical to the node maximizing another centrality.

Moreover, we verify whether one of the tested metrics could be directly used for the Local CN placement instead of the flow centrality. Hence, we compute the total traffic received by the Local CN if the latter was placed on a node maximizing one of these metrics, but not the flow centrality. If ρ_{max} is the maximum flow centrality in the graph, and $\bar{\rho}(u)$ is the flow centrality of a node u, then placing the Local CN on node u, instead of the node with the maximum flow centrality, causes a **relative traffic loss** $\epsilon_\rho(u)$, such that:

$$\epsilon_\rho(u) = \frac{\rho_{max} - \bar{\rho}(u)}{\rho_{max}} \tag{4.35}$$

For the evaluation, we consider random geometric graphs with 10 nodes, on a total surface of one unit square, and a radius $\eta = 0.2$. We fix the average link capacity $c_{avg} = 50$, and vary the capacity range Δc, such that $\Delta c = \{0, 10, 40, 100\}$.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

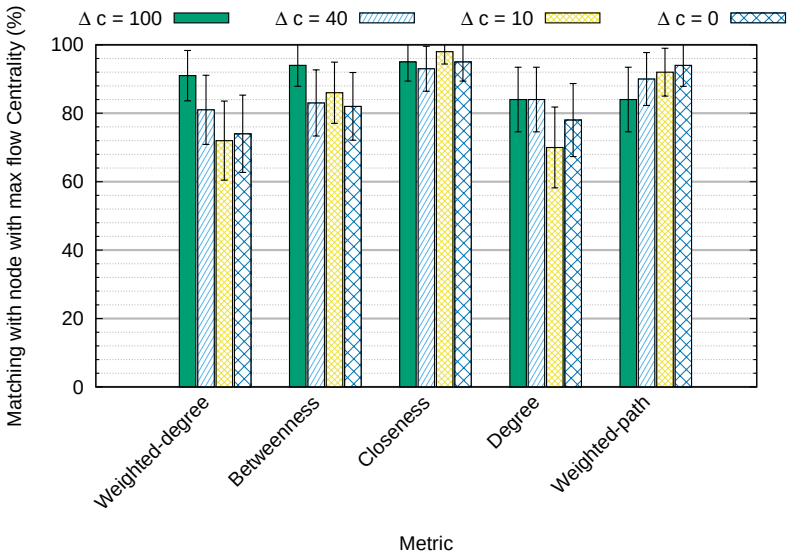


Figure 4.10: The percentage of scenarios where the node with maximum flow centrality is identical to the nodes maximizing other centralities, in random geometric graphs, for different capacity ranges, with constant average capacity $c_{avg} = 50$.

We present in Fig. 4.10 the percentage of scenarios where the node with maximum flow centrality is identical to the nodes maximizing the other centrality metrics. The results show that, for all values of Δc, the closeness centrality is the most similar to the flow centrality in terms of matching percentage. In contrast, the degree centrality is the farthest. For example, for $\Delta c = 0$, the node with maximum flow centrality matches the node with the maximum closeness centrality in 95% of the cases, and the node with the maximum degree 78% of the time. We also note that the capacity range Δc does not have a significant impact on the matching percentage. Thus, the closeness centrality, with its geographically central position, has advantages over the other metrics and is seemingly more suitable for the Local CN placement and the closest to the flow centrality.

However, even when the matching percentage is high, what validates if a metric is as suitable as the flow centrality for the Local CN placement is the incurred traffic loss when the Local CN is placed on the node maximizing that metric instead of the node maximizing the flow centrality, when the two are different. In Fig. 4.11, we show this relative traffic loss ϵ_ρ (Eq. 4.35). Even though nodes with maximum closeness centrality had the highest matching percentage with the nodes with maximum flow centrality, results show that the average traffic loss incurred when these nodes are different is relatively high. Fig. 4.11 indicates that placing the Local CN on the node with the maximum closeness centrality instead of the node with the maximum flow centrality would cause a relative traffic loss of 46.5%, which is the highest loss in comparison with the other centrality measures. On the other hand, the relative losses incurred by the other centrality measures are lower, but still important, around 30% on average.

Therefore, opting for the node with the maximum closeness (or any other) centrality as a placement for the Local CN, instead of the node with the maximum flow centrality, is not recommended. In the cases where the two nodes are different, the incurred loss is significant.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

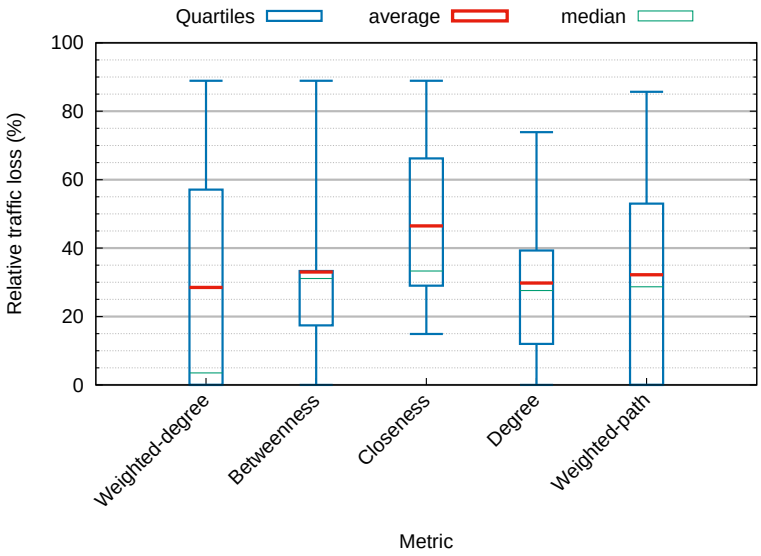


Figure 4.11: Relative traffic loss when Local CN is placed on the node maximizing centrality metrics other than the flow centrality, in random geometric graphs, for $\Delta c = 100$.

4.8 Large scale networks

Until now, we limited our study to smaller networks corresponding to our specific use case of self-deployable networks, where only few BSs are needed. In order to thoroughly study the flow centrality metric and generalize the obtained results, we vary in the following the number of nodes in the network. We observe in Fig. 4.12 the matching percentage between the node with the maximum flow centrality and the nodes maximizing other centrality metrics, function of the number of nodes n.

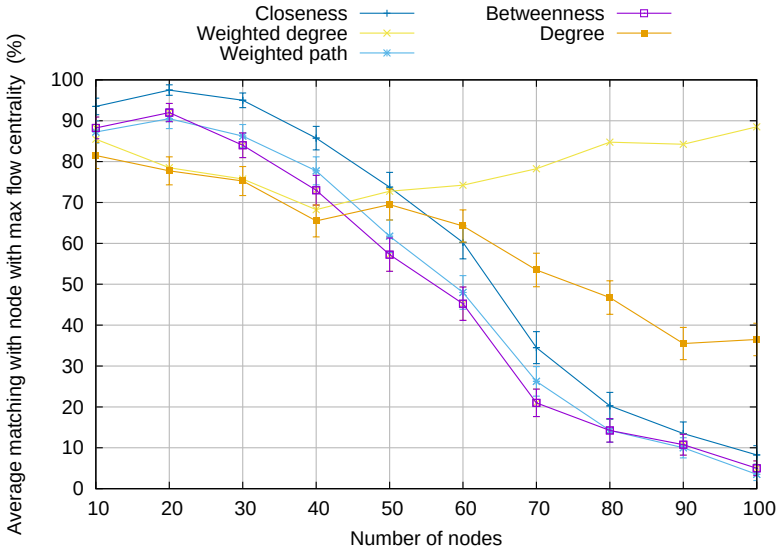


Figure 4.12: The percentage of scenarios where the node with maximum flow centrality is identical to the nodes maximizing other centralities, function of the number of nodes.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

We notice in Fig. 4.12 that for all metrics, except the weighted degree centrality, the matching percentage decreases with the increase of the number of nodes. For larger values of n, this matching percentage is negligible. Conversely, the matching percentage is conserved for the weighted degree centrality, even with the increase of n. The fact that flow centrality accounts for the link capacities make it inherently different than most of the other metrics. The more the network size increases, the number of links interconnecting the nodes also increases, and the divergence between the node maximizing the flow centrality and the nodes maximizing the other centralities is emphasized. In contrast, the weighted degree centrality conserves a relatively high matching percentage, since it also takes into account the capacity of the links surrounding the node. However, similarly to the smaller scale networks, despite having high matching percentages, the relative traffic loss when the two nodes are different is high.

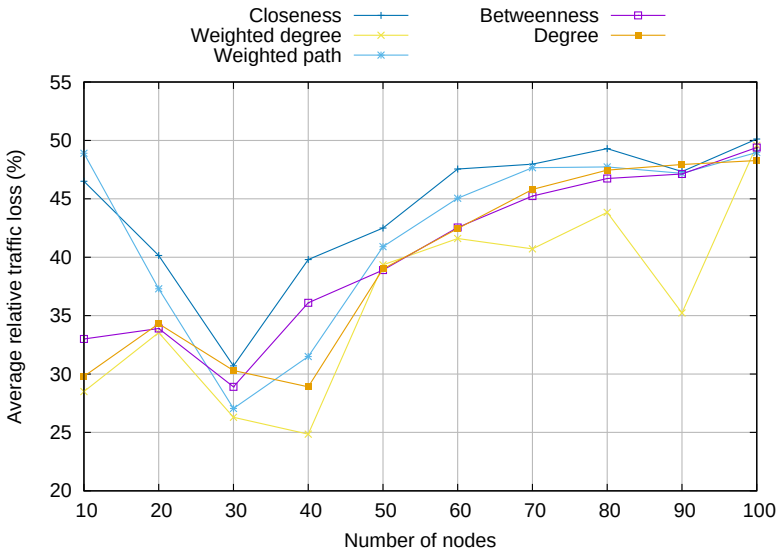


Figure 4.13: The average relative traffic loss when the Local CN is placed on the node maximizing centrality metrics other than the flow centrality, function of the number of nodes.

Indeed, we show in Fig. 4.13 the average traffic loss occurring when the Local CN is placed on the node maximizing one of the other metrics instead of the flow centrality. As n increases, the average traffic loss increases accordingly.

These observations generalize the previous conclusions, and further distinguish the node with the maximum flow centrality by its ability to receive the maximum flow in the network and its suitability for the Local CN placement, as opposed to the other metrics.

4.9 Conclusion

In a self-deployable network, consisting of BSs having no backhaul connection to a traditional core network, network services are provided by a Local CN co-located with the BSs. The Local CN functions can be centralized or distributed within the network. In this chapter, we focused on a centralized placement, such that the Local CN is co-located with only one BS. We tackled the placement problem to determine with which of the BSs it should be co-located. Following an overview of state of the art centrality metrics, we showed that placing the Local CN on a node maximizing one of these centralities does not necessarily allow the Local CN to receive the maximum amount of traffic from the BS. Therefore, we proposed flow centrality, a novel centrality

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

metric that measures a node capacity of receiving the total flow in the network. We showed that co-locating the Local CN with the BS having the maximum flow centrality maximizes the total amount of traffic that the Local CN is capable of receiving from all the BSs, while respecting the limited link capacities. We presented a detailed analytical study on different flow centrality computation methods, in different graph topologies. While the initial computation method is based on solving a linear optimization problem, analytical expressions in canonical topologies, such as paths and balanced treed, are deduced. We further highlighted the existing dependencies between the flow centrality metric and the backhaul network capacity. The latter notably affects the position of the node with the maximum flow centrality. Finally, we compared the flow centrality to the other state of the art centralities, to further establish the distinction between them. We highlighted the loss incurred when the Local CN is placed on a node maximizing one of the centrality metrics and not the flow centrality. We showed that, in some cases, the relative traffic loss can be as high as 55%.

In this work, we treated the BSs in the network equally by maximizing their capabilities in the same way. This is suitable for a network where the Local CN is statically placed at the initial deployment phase, when no sufficient information on the BSs demands due to the number of users and their requests is available, or when the Local CN position is fixed for the entire network operation. However, in practical deployments, during network operation, BSs generally have different demands, which directly affects the Local CN optimal position. Hence, it would be interesting to study non-uniform BS demands and possible migrations of the Local CN during network operation. Another important aspect in the communication between the BSs and the Local CN is the end-to-end delay. Delay constraints must be further included in the Local CN placement criteria. On the other hand, the possibility to co-locate the CN with the BSs opens up new perspectives on having the CN functions distributed in the network, rather than centralized on a single BS, as presented in this chapter. We evaluate the distribution of different CN functionalities and their optimal placement within the network in the next chapter.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 5

Core Network Functions Distribution

5.1 Introduction

The particular architecture of self-deployable mobile networks, with a Local CN co-located with the BSs, and an interconnection among BSs forming the backhaul network, opens up a multitude of possibilities regarding the Local CN functions placement, and raises questions on the reliability of a limited-bandwidth backhaul, and how to avoid its saturation. In the previous chapter, we showed how those two problems are correlated, with the limited backhaul bandwidth serving as a key criterion for placing the local CN. We focused on a centralized placement by considering that all the Local CN functions are co-located with only one BS. Furthermore, we ignored user requests, and determined the placement under some assumptions of demand similarity among the BSs. In this chapter, we further study the advantages of a distributed Local CN, with functions placed on multiple BSs in the network. We tackle the problem form a different point of view, by explicitly taking into consideration the number of users in the network, their distribution, and their requests.

From the moment we mention users in the network, addressing the user association problem becomes inevitable. User association consists of selecting the most appropriate BS to serve a user in the network [94]. Conventionally, user association is based solely on RAN metrics, and more specifically on the DL signal strength, with the assumption of an over-provisioned backhaul [99]. However, in self-deployable networks, with a potentially limited backhaul, the latter may represent a bottleneck. Indeed, multiple BSs share the same backhaul links to forward the traffic of their associated users. Ignoring the backhaul when associating users can eventually lead to early backhaul saturation. Hence, we address this problem by investigating the advantages (or lack thereof) of a backhaul-aware user association, that considers the backhaul bandwidth in addition to the RAN when associating users. The main idea is to have users associated in a way that minimizes the bandwidth consumption caused by their traffic on the backhaul.

In addition to associating to a BS, each active user in the network must also attach to the CN, more specifically to an MME and a S-GW. One of the objectives of the attachment is to assign the flows of each user to a specific S-GW [16]. In a classical cellular network, each user is attached to one default S-GW responsible of locally routing all of his flows. The selection of the S-GW to which a user attaches determines the data traffic routing path on the backhaul. Likewise, the placement of the MME handling all signaling procedures, such as gateway selection and data session management, affects the signaling traffic routing path. Thus, user attachment could affect the backhaul load and the backhaul bandwidth consumption. In order to assess the impact of the attachment procedure, we analyze the performance of different attachment optimization schemes. Moreover, we evaluate the impact of the signaling traffic caused by the signaling procedures.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

The Local CN functions placement, user association, and user attachment are somehow correlated problems. In fact, data and signaling traffic, originating from or destined to a user, are routed on the inter-BS backhaul links, between the BS the user is *associated* to and the *Local CN functions* the user is *attached* to. Reducing the amount of traffic routed on the backhaul requires a careful investigation of each of those three problems.

The endgame of this chapter is threefold: (i) validate assumptions regarding the network architecture, particularly the S-GW placement, (ii) understand the worthiness of optimizing user association and/or attachment, and (iii) ascertain the relevance of accounting for signaling traffic. Based on the obtained results, this work served as a stepping stone for the development of a more elaborate network-aware user association policy, presented in details in the next chapter.

The contributions of this chapter can be summarized as follows. First, we determine the optimal placement of the Local CN functions. We compare two strategies: a centralized Local CN, with one S-GW and one MME co-located with one BS, and a distributed Local CN, with S-GWs co-located with multiple BSs. The placement objective is to minimize backhaul bandwidth consumption caused by user data and signaling traffic exchanged among the BSs and the different Local CN entities. In case the Local CN is distributed, we also determine the optimal number of needed S-GWs. Results indicate that having S-GWs distributed in the network, such that each BS is co-located with one S-GW, outperforms a centralized Local CN from a backhaul bandwidth consumption point of view.

Second, we study optimized backhaul-aware user association. The corresponding optimization problem balances different constraints: the limited resource availability on the RAN, the throughput requested by flows between two users, and the data and signaling traffic routed on the backhaul. The problem is solved with the objective of minimizing the backhaul bandwidth consumption. Results show that optimizing association can reduce bandwidth consumption on the backhaul by 30%, in comparison with a traditional RAN-based association scheme [129].

Finally, we evaluate different attachment schemes, and formulate the corresponding optimization problems, with the same objective as described above. We show that an optimized attachment further reduces backhaul bandwidth consumption when the signaling traffic is significant. However, in comparison to optimizing user association, the contribution of optimizing the attachment to the overall gain is marginal [129].

5.2 System model

We consider a self-deployable mobile network, where $\mathcal{J}$ is the set of BSs, $\mathcal{L}$ is the set of directional links interconnecting the BSs, and $\mathcal{U}$ is the set of fixed UEs. When two BSs are linked, there are 2 directional links between them, one in each direction.

5.2.1 Core network and backhaul

The Local CN functions are co-located with the BSs. We focus on two main entities of the Local CN using the EPC architecture of an LTE system: the MME and the S-GW.

- The MME handles network management such as paging, authentication, user mobility and gateway selection. We consider that network management is ensured by a single MME entity, co-located with one of the BSs. Indeed, in a classical network, it is common to have a pool of distributed MMEs across different sites, serving a number of BSs. This enables geographical redundancy, increases overall capacity, achieves load balancing among MMEs, and eliminates single point of failure between an BS and MME [130]. The required signaling

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

exchanges among the MMEs of the same pool do not impact the overall network operation, as the interconnection among the MME is dedicated, planned, and usually over-provisioned. However, this is not the case in a self-deployable network. Due to the fact that the CN is local, the backhaul is impacted by the signaling exchanges among multiple MMEs. With MMEs co-located with different BSs, the signaling traffic between them is routed on the inter-BS backhaul links, further increasing the backhaul load. On the other hand, the small size of self-deployable networks, and their deployment in a constrained geographical area, do not necessitate several MMEs. For all of these reasons, we limit the number of MME functions within the Local CN to one.

- The S-GW handles local data routing. We consider that each BS of the network has the possibility to be co-located with a S-GW. One of our goals is to determine the number of needed S-GWs (when they are distributed), and their optimal placement (when centralized and when distributed).

All signaling traffic passes through the MME, and all data traffic passes through the S-GW. Both entities also exchange signaling traffic, as detailed afterwards in Section 5.2.2.2.

The same assumptions regarding the backhaul network, i.e., the inter-BS links, as explained in Chapter 2 (Section 2.6, p. 16), are made. To summarize, the backhaul and the RAN resources are independent. Regardless of the used wireless technology, there is no contention among backhaul links for resource utilization.

As previously explained, one of the main motivations behind the studied topics in this chapter is the limited backhaul bandwidth in a self-deployable network. We are looking to optimize the Local CN placement, the user association, and the user attachment in a way that minimizes the backhaul bandwidth consumption. We use this total bandwidth consumption on the backhaul as an evaluation metric to compare the performance of the different strategies and select the one that consumes less. In this context, there is no need to quantify the limited backhaul bandwidth. In other words, we do not set limited link capacities for the inter-BS backhaul links, since we are interested in the overall consumption of a studied strategy rather than its feasibility or not for a particular scenario with preset values for the limited backhaul links. Indeed, adding limited capacities would just add additional constraints in the problem formulation, making a problem feasible or infeasible depending if the backhaul consumption surpasses the backhaul limited bandwidth.

5.2.2 Traffic model

5.2.2.1 Data traffic

We adopt a data traffic model consisting of bidirectional flows between two parties (e.g., two UEs, or a UE and an application server), that is two directional flows, one in each direction. Flows can be intra-network (between two parties belonging to the same network) or inter-network (one of the parties belong to another network).

For brevity, the given examples, notations, and the subsequent numerical applications are limited to intra-network flows between two UEs. However, the problem formulation is general enough to include both intra and inter-network flows, as well as flows between a UE and an application server. The reason we retain intra-network flows between two UEs is to represent scenarios with the most loaded backhaul and RAN. In fact, in intra-network flows, both UEs are associated to BSs in the network and both consume RAN resources. Their data and signaling traffic, exchanged between the BSs and the Local CN functions, are only routed locally on the inter-BS backhaul links. Hence, this type of flows is supposed to be more costly, both on the RAN and on the backhaul, than an inter-network flow with only one UE in the network and part of the traffic being external.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Let $\mathcal{F}$ be the set of directional flows, and $f = \{u, v\} \in \mathcal{F}$ a directional flow from UE u to UE v. For each $f = \{u, v\} \in \mathcal{F}$, there exists $f' = \{v, u\} \in \mathcal{F}$ in the opposite direction. A flow exists between two UEs, in both directions, with a probability p. One UE can have several simultaneous flows with different UEs. We denote by d_f the requested data rate of a flow f, in bits/second.

5.2.2.2 Signaling traffic

Besides data traffic, signaling traffic takes up an important part in mobile networks. It is due to a multitude of signaling procedures handling the management of different aspects of the network [131]. While some signaling procedures have a minimum imposed frequency, such as tracking area update, others are timely, such as attach, detach, paging, handover, and session management (bearer setup and release) [131]. As detailed in Chapter 2 (Section 2.1.2, p. 8) the main signaling traffic is exchanged:

- between the MME and the BS to which the UE is associated, on the signaling interface S1-MME in LTE;
- between the MME and the S-GW to which the UE is attached, on the signaling interface S11 in LTE.

The frequency of signaling procedures is highly dependent on the users' activity in the network, e.g., their requests and their mobility. Hence, quantifying the amount of signaling traffic in a network is not a trivial task, since it depends on the particular scenario in question. This quantification requires details on the number of exchanged signaling messages, their frequency and their respective sizes, not only between the RAN and the CN, but also between the CN entities themselves.

Since we do not have worthy signaling traffic quantification, notably for inside the CN (e.g., between the MME and the S-GW), and to avoid limiting the study to a particular use case, we adopt, in this chapter, a simplified signaling model. We consider that signaling is closely related to data traffic. We suppose that each data flow f is accompanied by the two aforementioned types of signaling traffic, between the MME and the BS, and between the MME and the S-GW. Both of these communications consume backhaul bandwidth on the inter-BS links, in both directions. We suppose that the amount of signaling traffic is proportional to the corresponding data traffic. We denote by Si_f the bit rate of a signaling traffic i, accompanying flow f. Si_f is a linear function of the flow data rate d_f, and σ a percentage, such that:

$$Si_f = \sigma \cdot d_f \tag{5.1}$$

For the evaluation afterwards, we vary σ to observe the impact of the increase of the signaling traffic. For the signaling traffic between the BS and the UE, we suppose it is taken into consideration implicitly within the RAN resources.

5.2.2.3 Routing

Depending on the network topology, two BSs may or may not be directly connected via a backhaul link. Traffic (both data and signaling) is exchanged between a BS and a Local CN entity. This traffic is routed on the inter-BS links either directly, if the two end-entities are at one hop from each other, or through the interconnected BSs, otherwise. In the latter case, a routing policy is needed to determine on which links traffic is routed. Any routing policy could be adopted. We define $Z^l_{j,j'}$ as a boolean, such that $Z^l_{j,j'} = 1$ if link $l \in \mathcal{L}$ belongs to the routing path between BSs

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

j and j', and $Z^l_{j,j'} = 0$, otherwise. To simplify routing, we assume that, for all bidirectional flows, the route is the same in both directions. In other words, $Z^l_{j,j'} = Z^{l'}_{j',j}$, where l' is the link in the opposite direction of l.

5.2.3 Radio access network

We consider an OFDMA-based system with a total of M orthogonal channels allocated to the set of BSs and a time frame made of T time-slots. Distinct orthogonal channels are used for the DL and the UL. The channels are equally divided among the BSs, following a given frequency reuse scheme with reuse factor r[1]. The number of channels reserved for each BS on the DL and on the UL are denoted $\mathcal{K}^{DL}_j$ and $\mathcal{K}^{UL}_j$, respectively. We assume that all BSs are identical in terms of maximum transmit power, P_{BS}, and antenna gain G^a. BSs granted the same channels interfere with each other. We denote by $\mathcal{I}^{DL}_j$ and $\mathcal{I}^{UL}_j$ the set of BSs interfering with BS j on the DL and the UL, respectively.

To simplify notation, we use hereafter the notation DL/UL as a superscript on the variables within an equation, when the latter is written in the same way for both the DL and the UL counterparts. We define $R^{DL}_{u,j}$ and $R^{UL}_{u,j}$ as the per channel rates seen by UE u from BS j, on the DL and the UL, respectively. Let $R^{DL/UL}_{u,j} = \Psi(\gamma^{DL/UL}_{u,j})$, with $\Psi(\cdot)$ a discrete function mapping the per channel SINR $\gamma^{DL/UL}_{u,j}$ to the per channel data rate, determined by the MCS.

The per channel SINR on the DL, $\gamma^{DL}_{u,j}$, between UE u and BS j is defined as:

$$\gamma^{DL}_{u,j} = \frac{\frac{P_{BS}}{\mathcal{K}^{DL}_j} \cdot G_{u,j}}{\mathcal{N}_0 + \sum_{h \in \mathcal{I}^{DL}_j} \frac{P_{BS}}{\mathcal{K}^{DL}_h} \cdot G_{u,h}} \ , \tag{5.2}$$

where $\mathcal{N}_0$ is the per channel additive white Gaussian noise power, $G_{u,j}$ is the channel gain between UE u and BS j that accounts for the path loss $\Gamma_{u,j}$, shadow fading, antenna gain G^a, and equipment losses E, all expressed in dB. The path loss $\Gamma_{u,j}$ between u and j is written as $\Gamma_{u,j} = a + b \cdot \log(D_{u,j})$ [116], where a and b are standard coefficients that depend on the path loss model, and $D_{u,j}$ is the distance between UE u and BS j. We model the shadow fading through a normal distribution $\mathcal{N}(0, sd)$.

On the DL, computing $R^{DL}_{u,j}$ is straightforward under the assumption that all the channel gains are available. The SINR from each BS is first computed (Eq. 5.2), then $R^{DL}_{u,j}$ is deduced based on the MCS step function.

On the UL, computing $R^{UL}_{u,j}$ in network snapshots, when user association is not known, is trickier. $R^{UL}_{u,j}$ is function of the SINR $\gamma^{UL}_{u,j}$. Computing $\gamma^{UL}_{u,j}$ requires an estimation of the UL interference, which in turn requires the number of users associated to interfering BSs to be known (which is however unknown when association is not given in offline network snapshots). Moreover, in the UL SINR computation, the number of channels actually allocated to the user must be known. However, this depends on the still unknown UL rate (which we are trying to compute). These intricate computations prevent having the UL rates estimations as an input to the problem, similarly to the DL rates. Computing them within the problem leads to the inclusion of additional non-linear constraints, further complicating the problem formulation. Nonetheless, our goal in this chapter is only evaluative. In other words, we are looking to study the usefulness of a backhaul-aware association scheme, not to propose a novel association policy that requires an evaluation based on a precise physical layer model. We argue that a granular UL model, in this case, would

[1] An evaluation with the frequency and power allocation scheme based on the SFR power map proposed in Chapter 3 is kept for the next chapter.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

only complicate the computations without having any effect on the awaited outcome. Therefore, we leave the presentation of a detailed UL model for the next chapter, where an association policy is proposed and evaluated accordingly. We settle, in this chapter, for a simpler assumption, such that $R_{u,j}^{UL}$ and $R_{u,j}^{DL}$ are equal.

Typically, for each of its flow, the UE would receive a number of physical resource blocks, via scheduling, depending on its requirement and its SINR. To maintain the tractability of our framework, we simplify the model by assuming that for each flow the UE is granted a fraction of the channels available on the BS, for the UL and the DL, depending on the rate it gets from that BS, and on the throughput it asks for. To retain the model as an upper bound for the total bandwidth consumption on the backhaul, we impose that each UE is granted for each flow the exact throughput it requests in each direction (on the DL and on the UL). That is, for a flow f of throughput d_f, a user u associated to BS j is granted $\frac{d_f}{R_{u,j}^{DL/UL}}$ channels out of the $\mathcal{K}_j^{DL/UL}$ channels of that BS. We do not go into further details of user scheduling. Resource allocation is abstracted by an allocation of the available resources among the users (each user gets a percentage of the resources). The constraint on the RAN is such that the sum of channel fractions allocated to all UEs associated to a BS cannot exceed the maximum number of channels available at that BS.

5.3 Problem overview

Recall that our goals in this chapter are threefold: study the Local CN functions placement, assess the impact of an optimal user association, and analyze different attachment schemes, with the objective of minimizing backhaul bandwidth consumption. For each of the studied problems, a number of strategies unfold. The combinations of the different strategies lead to the formulation of different optimization problems. The comparison of the different problems allows us to evaluate each of the studied aspects and the correlation between them. Those problems are summarized in Table 5.1. Each problem is denoted $\mathcal{P}_{\mathbf{a/b/c}}$, with the indexes a, b, and c respectively representing the strategies adopted for: (a) determining the number of S-GWs and their placement, (b) user association, and (c) user attachment.

(a) For the Local CN functions placement, we determine the number of needed S-GWs and their distribution. We compare three strategies for the S-GWs placement:

- a = $\mathbf{1}$: there is one and only one S-GW co-located with one BS. The raised question is whether this S-GW should be co-located with the MME (i.e., a centralized Local CN), or if the optimal placements of those entities are different;
- a = $\mathcal{J}$: there are $|\mathcal{J}|$ S-GWs, as many as the BSs in the network, such that each BS is co-located with an S-GW;
- a = $\mathbf{o}$: there are multiple S-GWs distributed in the network. The number of the S-GWs is optimized to determine whether all the BSs of the network should be co-located with an S-GW, or only a subset.

(b) For the association, we also compare two strategies:

- b = $\mathbf{g}$: the user association is given based on a common association policy;
- b = $\mathbf{o}$: the user association is backhaul-aware, with users optimally associated in order to minimize the bandwidth consumption on the backhaul.

(c) For the attachment, three strategies unfold:

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

- c = **g**: there is no attachment decision to make. In case there is one S-GW in the network, the attachment is trivial. In case there are $|\mathcal{J}|$ S-GWs co-located with all the BSs (i.e., a = $\mathcal{J}$), the attachment follows the association, such that a UE is attached to the S-GW co-located with the BS it is associated to;
- c = **o**: the attachment is optimized per user. A UE is attached to one S-GW for all of its flows. The S-GW can be co-located with any BS, even one that is different from the BS it is associated to;
- c = **of**: the attachment is optimized on a per-flow basis. A UE is attached to one S-GW per flow. A UE can attach to different S-GWs for its different flows. UEs at both ends of a flow are attached to the same S-GW.

Problem $\mathcal{P}_{a/b/c}$	Number of S-GWs	Association	Attachment
$\mathcal{P}_{1/g/g}$	1	Given	Given
$\mathcal{P}_{\mathcal{J}/g/g}$	$\|\mathcal{J}\|$	Given	Given
$\mathcal{P}_{o/g/o}$	Optimized	Given	Optimized per user
$\mathcal{P}_{\mathcal{J}/o/g}$	$\|\mathcal{J}\|$	Optimized	Given
$\mathcal{P}_{o/o/o}$	Optimized	Optimized	Optimized per user
$\mathcal{P}_{o/g/of}$	Optimized	Given	Optimized per flow
$\mathcal{P}_{o/o/of}$	Optimized	Optimized	Optimized per flow

Table 5.1: Optimization problems nomenclature & summary.

Typically, each BS of the network must be served by an MME co-located with one of the BSs. We define vector W, such that $W_j = 1$ if the MME is co-located with BS j. Vector W is an output in all the problems. Each UE must associate to one and only one BS, and attach to at least one S-GW. We define association vector X, with $X_{u,j}$ a boolean, such that $X_{u,j} = 1$ if UE u is associated to BS j, and $X_{u,j} = 0$, otherwise. In each problem, vector X is either an output or an input, depending on whether we are looking to optimize association, or it is just given.

All the problems presented in Table 5.1 have the same objective function (Eq. 5.3), and several shared constraints (Eq. 5.4 - Eq. 5.9). The inputs and the outputs differ from one problem to another, as well as the remaining constraints, as detailed in the following.

5.4 Number and placement of S-GWs

To highlight the problem relevance, Fig. 5.1 illustrates in a simple example how the placement of the Local CN functions can impact the traffic on the backhaul, and the consequent bandwidth consumption,

The four cases in Fig. 5.1 correspond to a flow between two UEs associated to BSs 2 and 3, respectively, accompanied by the two types of signaling traffic: between the BSs and the MME, and between the MME and the S-GW(s). When the MME and the S-GW are co-located, signaling traffic between them is not routed on the backhaul (cases (a) and (d)), further economizing bandwidth. When the S-GWs are co-located with the BSs to which the UEs are associated (case (b) and (c)), or belong to the shortest path between the BSs (case (d)), the flow takes a shortest path between the two BSs. Hence, it consumes less backhaul bandwidth than the one having to go through a S-GW that does not belong to the shortest path between the two BSs (case (a)). The same can be said about the MME and the resulting signaling traffic path: the shortest the path

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

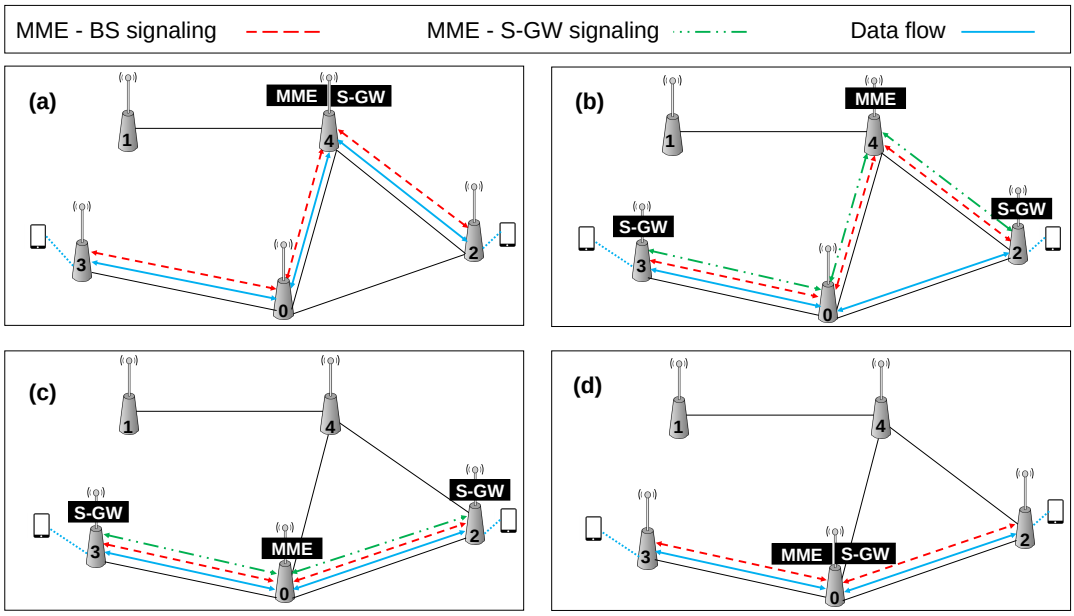


Figure 5.1: An example on the different data and signaling traffic paths for a flow between two UEs, in scenarios with different Local CN functions placement.

between the BSs and the MME, the less is the MME-BS signaling backhaul consumption (cases (c) and (d)). While this example shows the trade-offs for a particular flow, the placement must actually take into account the overall traffic in the network, i.e., all the flows between the different BSs and the consumption incurred by the corresponding signaling and data flows. The optimal placement minimizing the backhaul bandwidth consumption is found by solving the optimization problems formulated in the following.

In this section, we address the number of needed S-GWs and their placement in the network. At this step, we ignore the user association, i.e., we do not try to optimize it, by supposing that association is given. We compare three S-GW placement schemes, by solving three mixed integer linear programming (MILP) problems: ($\mathcal{P}_{\mathbf{1/g/g}}$) where only one S-GW is optimally placed in the network, ($\mathcal{P}_{\mathcal{J}\mathbf{/g/g}}$) where all BSs are co-located with S-GWs, such that each UE attaches to the S-GW co-located with the BS it is associated to, and ($\mathcal{P}_{\mathbf{o/g/o}}$) where all BSs are co-located with S-GWs, but the attachment is optimized such that each UE can attach to any S-GW (not necessarily the one on the BS it is associated to).

5.4.1 One default S-GW in the network

In this scenario, only one S-GW exists, and we want to optimally place it in the network in a way that minimizes backhaul bandwidth consumption. Hence, we formulate the corresponding problem, denoted $\mathcal{P}_{1/g/g}$. We define vector G, such that $G_j = 1$ if the S-GW is co-located with BS j. Similarly to vector W that outputs the MME placement, vector G is an output of the problem. No attachment decision is needed since all UEs attach to the only available S-GW in the network. User association is given, such that vector X is an input, with $X_{u,j}$ known $\forall u \in \mathcal{U},\ j \in \mathcal{J}$.

To better illustrate the problem in this scenario, Fig. 5.2 shows the data and signaling traffic paths for a flow between UEs u and v, respectively associated to BSs j_1 and j_2, with the MME

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

at j_0 and the S-GW at j_5. We denote by d_f the bit rate of the data flow between UEs u and v. $S1_f$ and $S2_f$ are used to denote the bit rate of the signaling traffic between the MME, from one side, and BSs j_1 and j_2, respectively, from the other side. We denote by $S5_f$ the bit rate of the signaling traffic between the MME and the S-GW. Recall from Eq. 5.1, that the signaling bit rates Si_f, accompanying flow f, are proportional to the flow data rate d_f.

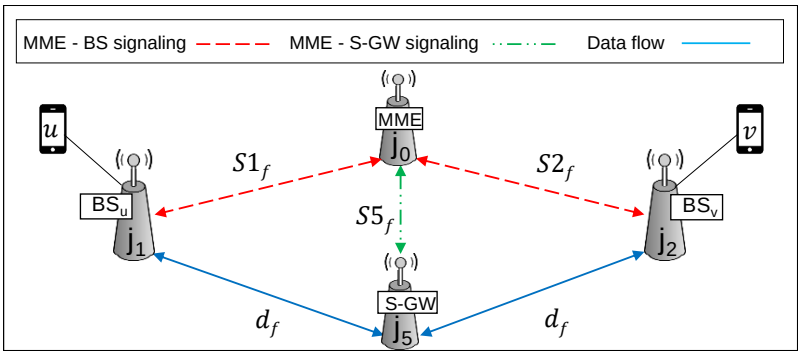


Figure 5.2: Data and signaling traffic paths between BSs, and their corresponding bit rates d_f and Si_f, for a flow f, when there is one S-GW in the network ($\mathcal{P}_{1/g/g}$), or when attachment is optimized per flow ($\mathcal{P}_{o/g/of}$, and $\mathcal{P}_{o/o/of}$)

The bandwidth consumed on a backhaul link l by all the flows is denoted C_l and computed in constraint 5.13. It is the sum of the bandwidth consumed by all data and signaling traffic on that link. Our objective is to minimize the total bandwidth consumed on the backhaul, on all links by all flows, as formulated in the objective function in Eq. 5.3.

Constraints 5.4 and 5.5 state that there is one and only one MME in the network, and that a UE is associated to one and only one BS, respectively. Constraints 5.6 and 5.7 state that the total flows received from UEs associated to a BS on the DL should not exceed the DL BS capacity, and the total flows sent by UEs associated to a BS on the UL should not exceed the UL BS capacity, respectively. If, for a given set of UEs $\mathcal{U}$, and flows $\mathcal{F}$, there are no sufficient resources on the RAN to support all the flows (i.e., constraint 5.6 or constraint 5.7 not satisfied), then the problem is unfeasible. Constraint 5.10 states that there is one and only one S-GW in the network.

The data path of a flow f goes from j_1 to j_2, passing through the only S-GW at j_5 (Fig. 5.2). In this case, a rate d_f is consumed by f on each link l on the routing path between j_1 and j_5, and between j_2 and j_5, that is on each link l with $Z^l_{j_1,j_5} = 1$, and each link l with $Z^l_{j_5,j_2} = 1$. We compute in constraint 5.11 the value of C^d_l, for each backhaul link l, which is the total bandwidth consumed by the data traffic of all the flows on a link l.

On the other hand, signaling bit rates $S1_f$, $S2_f$, and $S5_f$ are consumed on each link l that belongs to the routing path between MME j_0 and BS j_1, MME j_0 and BS j_2, and MME j_0 and S-GW j_5, respectively. In constraints 5.8, 5.9, and 5.12, we compute C^{S1}_l, C^{S2}_l, and C^{S5}_l, respectively representing the total bandwidth consumed by the signaling traffic, of bit rates $S1_f$, $S2_f$, and $S5_f$, accompanying all flows f on a single link l.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

$$\min \sum_{l \in \mathcal{L}} C_l \tag{5.3}$$

$$\sum_{j \in \mathcal{J}} W_j = 1 \tag{5.4}$$

$$\sum_{j \in \mathcal{J}} X_{u,j} = 1 \ , \ \ \forall u \in \mathcal{U} \tag{5.5}$$

$$\sum_{u \in \mathcal{U}} \frac{X_{u,j}}{\mathcal{K}_j^{DL} \cdot R_{u,j}^{DL}} \sum_{f \in \mathcal{F}/u \in f} d_f \leq 1 \ , \ \ \forall j \in \mathcal{J} \tag{5.6}$$

$$\sum_{u \in \mathcal{U}} \frac{X_{u,j}}{\mathcal{K}_j^{UL} \cdot R_{u,j}^{UL}} \sum_{f \in \mathcal{F}/u \in f} d_f \leq 1 \ , \ \ \forall j \in \mathcal{J} \tag{5.7}$$

$$C_l^{S1} = \sum_{f \in \mathcal{F}} \sum_{j_1 \in \mathcal{J}} X_{u,j_1} \sum_{j_0 \in \mathcal{J}} W_{j_0} \Big(Z_{j_1,j_0}^l + Z_{j_0,j_1}^l \Big) S1_f \ , \ \ \forall l \in \mathcal{L} \tag{5.8}$$

$$C_l^{S2} = \sum_{f \in \mathcal{F}} \sum_{j_2 \in \mathcal{J}} X_{v,j_2} \sum_{j_0 \in \mathcal{J}} W_{j_0} \Big(Z_{j_2,j_0}^l + Z_{j_0,j_2}^l \Big) S2_f \ , \ \ \forall l \in \mathcal{L} \tag{5.9}$$

$$\sum_{j \in \mathcal{J}} G_j = 1 \tag{5.10}$$

$$C_l^d = \sum_{j \in \mathcal{J}} G_j \sum_{f \in \mathcal{F}} \Big(\sum_{j_1 \in \mathcal{J}} X_{u,j_1} Z_{j_1,j}^l + \sum_{j_2 \in \mathcal{J}} X_{v,j_2} \cdot Z_{j,j_2}^l \Big) d_f \ , \ \forall l \in \mathcal{L} \tag{5.11}$$

$$C_l^{S5} = \sum_{f \in \mathcal{F}} \Big(\sum_{j \in \mathcal{J}} G_j \sum_{j_0 \in \mathcal{J}} W_{j_0} \cdot Z_{j_0,j}^l \cdot S5_f \Big) \ , \ \ \forall l \in \mathcal{L} \tag{5.12}$$

$$C_l = C_l^d + C_l^{S1} + C_l^{S2} + C_l^{S5} \tag{5.13}$$

5.4.2 All BSs co-located with S-GWs

In this scenario, each BS in the network is co-located with a S-GW. No attachment decision is needed since user attachment follows user association, such that a UE attaches to the BS to which it is associated. User association is given. The corresponding problem, $\mathcal{P}_{\mathcal{J}/g/g}$, only outputs the MME optimal placement.

Fig. 5.3 shows the data and signaling traffic paths for a flow between UEs u and v, respectively associated to BSs j_1 and j_2, co-located with their corresponding S-GWs. If u (resp. v) is associated to BS j_1 (resp. j_2), then u (resp. v) is attached to S-GW j_1 (resp. j_2). We denote by $S3_f$ and $S4_f$ the bit rates of the signaling traffic between the MME, from one side, and S-GWs j_1 and j_2, respectively, from the other side.

In $\mathcal{P}_{\mathcal{J}/g/g}$, the objective function is the same as the one in Eq. 5.3, minimizing the total bandwidth consumed on the backhaul. The constraints from 5.4 to 5.9 remain unchanged.

In this scenario, the data traffic cost, and the signaling between the MME and the S-GW to which a user is attached, are, however, formulated differently. The data path of a flow f goes directly from j_1 to j_2 (Fig. 5.3). In constraint 5.14, we compute C_l^d, the total bandwidth consumed by the data traffic of all the flows f on each link l. The total bandwidth consumed on each link l, by

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

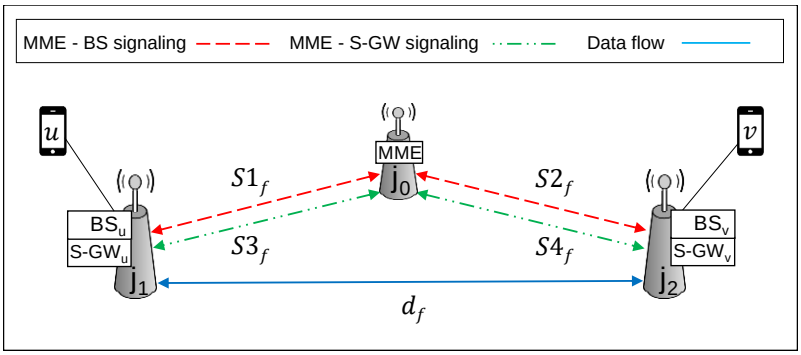


Figure 5.3: Data and signaling traffic paths between BSs, and their corresponding bit rates d_f and Si_f, for a flow f, when there are $|\mathcal{J}|$ S-GWs ($\mathcal{P}_{\mathcal{J}/g/g}$ and $\mathcal{P}_{\mathcal{J}/o/g}$).

the signaling between MME j_0 and S-GW j_1, and between MME j_0 and S-GW j_2, with respective data rates $S3_f$ and $S4_f$ for each flow f, are computed in constraint 5.15 and constraint 5.16, and denoted C_l^{S3} and C_l^{S4}, respectively. Hence, the constraints 5.10 to 5.13 are replaced by:

$$C_l^d = \sum_{f\in\mathcal{F}} \Big(\sum_{j_1\in v} X_{u,j_1} \sum_{j_2\in\mathcal{J}} X_{v,j_2} \cdot Z^l_{j_1,j_2} \cdot d_f \Big) \ , \ \forall l \in \mathcal{L} \tag{5.14}$$

$$C_l^{S3} = \sum_{f\in\mathcal{F}} \Big(\sum_{j_1\in\mathcal{J}} X_{u,j_1} \sum_{j_0\in\mathcal{J}} W_{j_0} \cdot Z^l_{j_0,j_1} \cdot S3_f \Big) \ , \ \ \forall l \in \mathcal{L} \tag{5.15}$$

$$C_l^{S4} = \sum_{f\in\mathcal{F}} \Big(\sum_{j_2\in\mathcal{J}} X_{v,j_2} \sum_{j_0\in\mathcal{J}} W_{j_0} \cdot Z^l_{j_0,j_2} \cdot S4_f \Big) \ , \forall l \in \mathcal{L} \tag{5.16}$$

$$C_l = C_l^d + C_l^{S1} + C_l^{S2} + C_l^{S3} + C_l^{S4} \tag{5.17}$$

5.4.3 Optimized S-GW placement

In this scenario, each BSs can be co-located with a S-GW. However, unlike the previous scenario, each UE can attach to any S-GW, not necessarily the one it is associated to. In this case, user attachment is optimized with the objective of minimizing backhaul bandwidth consumption. We formulate the corresponding problem $\mathcal{P}_{o/g/o}$. In addition to the optimal MME placement, the problem outputs which S-GWs have UEs attached to them: this could be all S-GWs, a subset or only one. Recall that user association is given.

Fig. 5.4 shows the data and signaling paths for a flow between u and v, respectively associated to j_1 and j_2, and attached to S-GWs j_3 and j_4, that may or may not be the same as j_1 and j_2.

In $\mathcal{P}_{o/g/o}$, the objective function that minimizes the total bandwidth consumed on the backhaul is the same as in Eq. 5.3. Likewise, the constraints from 5.4 to 5.9 remain unchanged.

Since we optimize user attachment, we define the attachment vector Y, with $Y_{u,j}$ a boolean such that $Y_{u,j} = 1$ if UE u is attached to S-GW j. We add constraint 5.18, stating that a UE is attached to one and only one S-GW. Since the BS to which a UE is associated and the S-GW to which it is attached are not necessarily co-located, the data path from BS j_1 to BS j_2 passes through S-GW j_3 then S-GW j_4 (Fig. 5.4). We compute the bandwidth consumption C_l^d caused by the data traffic of all flows on a link l in constraint 5.19.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

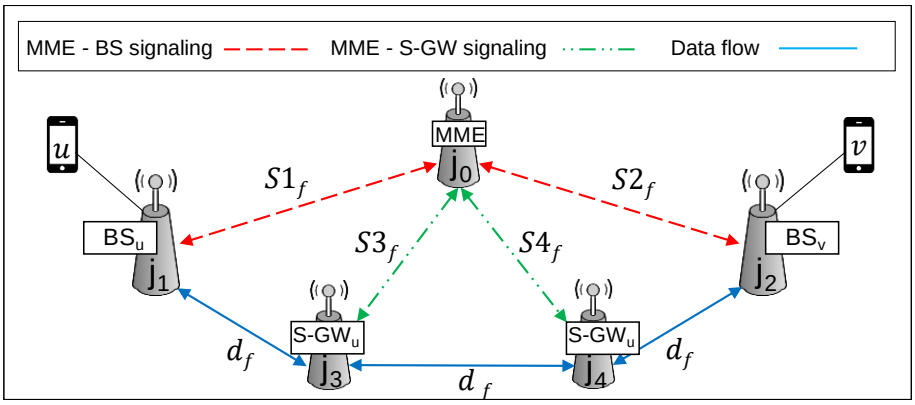


Figure 5.4: Data and signaling traffic paths between BSs, and their corresponding bit rates d_f and Si_f, for a flow f, when the number of S-GWs is optimized ($\mathcal{P}_{o/g/o}$ and $\mathcal{P}_{o/o/o}$).

In constraints 5.20 and 5.21, we compute C_l^{S3} and C_l^{S4}, which are the costs incurred by all flows on each link l from the signaling between MME j_0 and S-GW j_3, to which one UE is attached, and between MME j_0 and S-GW j_4, to which the other UE is attached, respectively.

The constraints 5.10 to 5.13 are replaced by the following constraints:

$$\sum_{j\in\mathcal{J}} Y_{u,j} = 1\ ,\quad \forall u \in \mathcal{U} \tag{5.18}$$

$$C_l^d = \sum_{f\in\mathcal{F}} d_f \Bigg(\sum_{j_1\in\mathcal{J}} X_{u,j_1} \sum_{j_3\in\mathcal{J}} Y_{u,j_3}\cdot Z^l_{j_1,j_3} + \sum_{j_4\in\mathcal{J}} Y_{v,j_4}\Big(\sum_{j_3\in\mathcal{J}} Y_{u,j_3}\cdot Z^l_{j_3,j_4} + \sum_{j_2\in\mathcal{J}} X_{v,j_2}\cdot Z^l_{j_4,j_2}\Big)\Bigg)\ ,\quad \forall l\in\mathcal{L} \tag{5.19}$$

$$C_l^{S3} = \sum_{f\in\mathcal{F}} \Big(\sum_{j_3\in\mathcal{J}} Y_{u,j_3} \sum_{j_0\in\mathcal{J}} W_{j_0}\cdot Z^l_{j_0,j_3}\cdot S3_f\Big)\ ,\quad \forall l\in\mathcal{L} \tag{5.20}$$

$$C_l^{S4} = \sum_{f\in\mathcal{F}} \Big(\sum_{j_4\in\mathcal{J}} Y_{v,j_4} \sum_{j_0\in\mathcal{J}} W_{j_0}\cdot Z^l_{j_0,j_4}\cdot S4_f\Big)\ ,\forall l\in\mathcal{L} \tag{5.21}$$

$$C_l = C_l^d + C_l^{S1} + C_l^{S2} + C_l^{S3} + C_l^{S4} \tag{5.22}$$

5.4.4 Numerical results

We set $M = 120$ channels, equally divided between DL and UL, and among the BSs with a reuse factor $r = 3$. We consider a distance-based path loss model for an urban setting, such that $\Gamma_{u,j} = 128.1 + 37.6\cdot\log(D_{u,j}/1000)$ dB [116]. We set $P_{BS} = 46$ dBm, $G^a = 10$ dB, $E = 20$ dB, $N_0 = -121$ dBm, and we model shadow fading using a normal distribution $\mathcal{N}(0, 8)$. Table 5.2 gives the numerical values of the mapping between the SINR and the rates for the discrete rate function $\Psi(.)$ [132].

When association is given, association vector X is determined using a traditional DL-oriented RAN-based association policy, in which a UE associates to the BS from which it gets the maximum

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

SINR threshold (dB)	-6.5	-4	-2.6	-1	1	3	6.6	10	11.4	11.8	13	13.8	15.6	16.8	17.6
Data rate (Kb/s)	25.2	38.6	63.8	100.8	147.8	198.2	248.6	320.9	404.9	458.6	557.8	655.2	759.4	860.2	932.4

Table 5.2: SINR thresholds and the corresponding data rates based on the MCS.

SINR. For the routing on inter-BS links, we adopt a shortest path routing policy, where the shortest path in terms of number of hops is selected between two BSs.

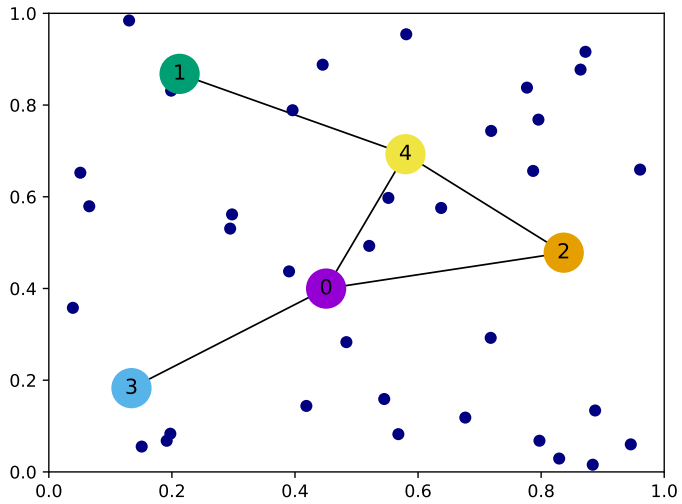


Figure 5.5: A network topology.

We consider as an example the network topology in Fig. 5.5. We model the network as a random geometric graph of radius $\eta = 0.4$, with 5 BSs deployed in an area of 1 unit square. Tests are conducted on 400 network snapshots. Each snapshot consists of a different combination of a user distribution and a flow distribution (out of 20 user distributions and 20 flow distributions). In each snapshot, there are 35 UEs, such that each UE has on average 3 simultaneous intra-network bidirectional flows with other users. We consider that all bidirectional flows are symmetric, such that for a directional flow f and its counterpart in the other direction f', we have $d_f = d_{f'} = 1$ Mb/s. The reason we consider symmetric traffic is to study scenarios where the RAN and backhaul are most loaded. The MME-BS and MME-S-GW signaling bit rates, namely $S1_f$, $S2_f$, $S3_f$, $S4_f$, and $S5_f$, are all considered equal. As explained in Section 5.2, we avoid limiting our study to a particular use case by using pre-defined values of signaling bit rates. To that end, for all Si_f, we vary the percentage σ in Eq. 5.1 that represents the percentage of signaling traffic with respect to the data traffic

Throughout the chapter, all optimization problems are solved using the commercial solver "CPLEX" [128], on a server with Intel Xeon CPU E5-649 @ 2.53GHz. Solving the problems, despite their complexity, is possible because of their small sizes, a direct consequence of the limited size of self-deployable networks.

For the network topology in Fig. 5.5, the solutions of the three problems returned either BS 0 or BS 4 as the optimal placement of the MME (depending on user/flow distribution of the tested snapshot), regardless of the signaling traffic. Recall that, in Chapter 4, we studied the optimal cen-

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

tralized placement of Local CN functions. We proposed placing them on the node that maximizes the flow centrality metric. By computing the flow centrality of the network nodes in Fig. 5.5, we do find that BS 0 and BS 4 both have the maximum flow centrality, which corroborates the optimal MME placement returned by the optimization problems. We note that, for the shown user distribution in Fig. 5.5, BS 4 is the optimal placement of the MME.

In order to compare the three S-GW placement schemes, Fig. 5.6 shows the total backhaul bandwidth consumption, obtained by solving each of the problems $\mathcal{P}_{1/g/g}$, $\mathcal{P}_{\mathcal{J}/g/g}$, and $\mathcal{P}_{o/g/o}$, function of the signaling traffic represented by σ, to further assess the impact of signaling traffic.

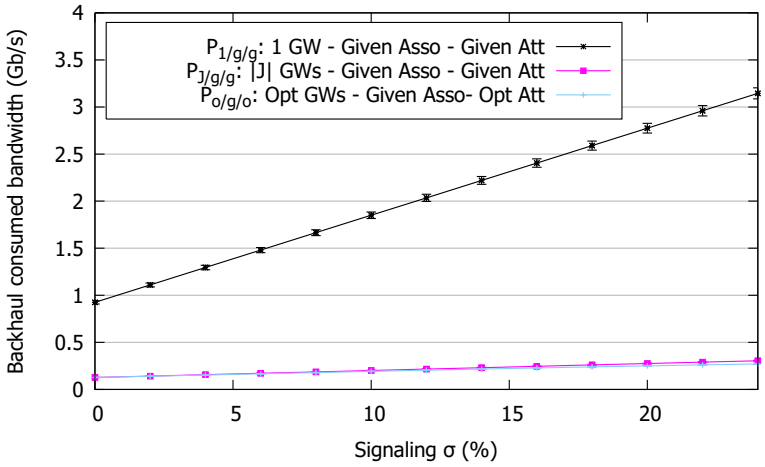


Figure 5.6: The total backhaul bandwidth consumption in $\mathcal{P}_{1/g/g}$, $\mathcal{P}_{\mathcal{J}/g/g}$, and $\mathcal{P}_{o/g/o}$, function of σ.

We observe in Fig. 5.6 that much more bandwidth is consumed on the backhaul when only one S-GW is placed in the network ($\mathcal{P}_{1/g/g}$), compared to when all BSs are co-located with S-GWs ($\mathcal{P}_{\mathcal{J}/g/g}$). Indeed, data traffic between two BSs is always routed through the S-GW. With only one S-GW in the network, the latter is not necessarily placed on the shortest backhaul path between the two BSs. This further increases backhaul consumption.

This observation is true for all values of signaling traffic represented by σ. Indeed, even when signaling traffic is not accounted for (i.e., $\sigma = 0$), co-locating S-GWs with all the BSs would cause a significant economy in the backhaul consumption, reducing the latter by around 86% with respect to having only one S-GW. With the increase of σ, this reduction percentage is not significantly impacted (remains around 90%). This means that the seen backhaul consumption reduction is mostly due to the decrease in data traffic on the backhaul, and not the signaling traffic.

These observations confirm that having distributed S-GWs in the network is evidently better than having one centralized S-GW. However, should all BSs be co-located with S-GWs or only a subset? To answer this question, we investigate the user attachment distribution when the S-GWs placement and user attachment are optimized ($\mathcal{P}_{o/g/o}$). In Fig. 5.7, we show the user attachment distribution, i.e., the percentage of users attached to each of the BSs, for $\mathcal{P}_{1/g/g}$, $\mathcal{P}_{\mathcal{J}/g/g}$, and $\mathcal{P}_{o/g/o}$. The results in Fig. 5.7 correspond to one of the tested network snapshots, with the user distribution shown in Fig. 5.5, and signaling traffic at $\sigma = 6\%$. For $\mathcal{P}_{1/g/g}$, the only S-GW in the network is co-located with BS 4. Recall that this is also the optimal placement of the MME in this network. For $\mathcal{P}_{o/g/o}$, all BSs in the network have users attached to them. This means that, with optimized attachment, all BSs are co-located with S-GWs. If we compare the attachment distribution in this case, with the one obtained for $\mathcal{P}_{\mathcal{J}/g/g}$, where attachment only

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

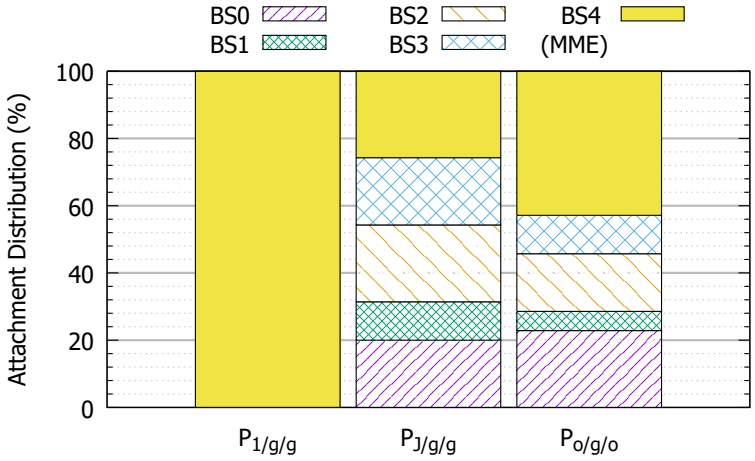


Figure 5.7: User attachment distribution in $\mathcal{P}_{1/g/g}$, $\mathcal{P}_{\mathcal{J}/g/g}$, and $\mathcal{P}_{o/g/o}$, for the studied topology and user distribution, with signaling traffic at $\sigma = 6\%$.

follows association, we notice that the difference mainly resides in BS 4. Indeed, the BS co-located with the MME, has more users attached to it. However, we notice in Fig. 5.6 that the backhaul consumption in $\mathcal{P}_{\mathcal{J}/g/g}$ and $\mathcal{P}_{o/g/o}$ are practically identical, notably for smaller values of σ. These results suggest that, for most users, optimizing attachment is not so different from attaching to the S-GW co-located with the BS to which they are already associated. This is mainly due to two reasons. First, when two UEs are associated to the same BS, if they also attach to the same S-GW co-located with it, the data they exchange is not routed on the backhaul. Second, when two UEs are associated to two different BSs, as data traffic goes from one BS to the other via the corresponding S-GWs, the data path is the shortest, i.e., causes a minimum backhaul bandwidth consumption, when the two S-GWs belong to the shortest path between the two BSs. One case is to have them directly co-located with the BSs on both ends.

5.5 Optimizing user association

Thus far, we considered that user association is given, based on a traditional RAN-based association policy. However, in self-deployable networks, backhaul may represent a bottleneck, since the links between BSs, forming this backhaul, may have a limited bandwidth. Hence, we investigate in the following the advantages of having a backhaul-aware user association policy that jointly considers the RAN and the backhaul. Our goal is to determine to which BS each UE associates, in such a way that the bandwidth consumption on the backhaul is minimized, and quantify the consumption reduction.

Based on the previous results, we suppose that all BSs are co-located with S-GWs. We set as reference the scheme in $\mathcal{P}_{\mathcal{J}/g/g}$, where both association and attachment are given, such that each UE associates to the BS from which it gets the maximum SINR, and attachment follows association. We compare the output of three particular schemes:

- $\mathcal{P}_{\mathbf{o/g/o}}$, where user association is given, but the attachment is optimized;
- $\mathcal{P}_{\mathcal{J}/\mathbf{o}/\mathbf{g}}$, where user association is optimized, but the attachment is not since it follows the association;

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

- $\mathcal{P}_{\mathbf{o/o/o}}$, where both user association and user attachment are optimized.

The only difference between $\mathcal{P}_{o/o/o}$ and $\mathcal{P}_{o/g/o}$ (presented earlier), is that in $\mathcal{P}_{o/g/o}$ user association is given (vector X is an input), whereas in $\mathcal{P}_{o/o/o}$ user association is optimized (vector X is an output). Likewise for problems $\mathcal{P}_{\mathcal{J}/o/g}$ and $\mathcal{P}_{\mathcal{J}/g/g}$. Therefore, $\mathcal{P}_{o/o/o}$ and $\mathcal{P}_{\mathcal{J}/o/g}$ are MIQP problems.

Since our goal is to investigate backhaul bandwidth consumption, we use as evaluation metric the relative backhaul bandwidth consumption reduction with respect to the benchmark $\mathcal{P}_{\mathcal{J}/g/g}$, denoted δ, and expressed as a percentage. If $C_{\mathcal{P}_{\mathcal{J}/g/g}}$ is the total backhaul bandwidth consumption in $\mathcal{P}_{\mathcal{J}/g/g}$, and $C_{\mathcal{P}_{a/b/c}}$ is the total backhaul bandwidth consumption in $\mathcal{P}_{a/b/c}$, then:

$$\delta = 100 \cdot \frac{C_{\mathcal{P}_{\mathcal{J}/g/g}} - C_{\mathcal{P}_{a/b/c}}}{C_{\mathcal{P}_{\mathcal{J}/g/g}}} \tag{5.23}$$

We consider the same example as in Sec. 5.4.4, with the same input data and network snapshots for the same network topology in Fig. 5.5. Fig. 5.8 shows the relative backhaul bandwidth consumption reduction δ function of the signaling traffic.

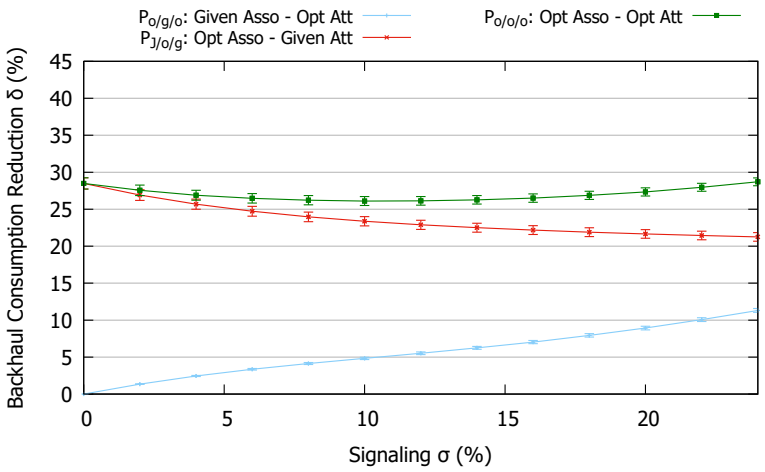


Figure 5.8: The relative backhaul bandwidth consumption reduction δ function of σ, in $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, and $\mathcal{P}_{o/o/o}$.

For $\mathcal{P}_{\mathcal{J}/o/g}$, we notice that, for $\sigma = 0$, the backhaul bandwidth consumption is reduced by 30%. In other words, a gain of 30% in backhaul capacity is achieved when UEs are optimally associated, which is a relatively important gain.

When attachment is also optimized, the results are exactly the same in $\mathcal{P}_{o/o/o}$ as in $\mathcal{P}_{\mathcal{J}/o/g}$, for $\sigma = 0$. This is expected because when signaling traffic is not accounted for, the only traffic on the backhaul comes from data following the path as shown in Fig. 5.4: BS j_1, S-GW j_3, S-GW j_4, BS j_2, and vice-versa. In this case, the optimal solution is for each UE to attach to the BS it is associated to, similarly to what happens in $\mathcal{P}_{\mathcal{J}/o/g}$, meaning no need for j_1 and j_3 (resp. j_2 and j_4) to be distinct. However, when signaling is accounted for, the previous statement does not hold anymore, since signaling between S-GW and MME is also routed on the backhaul. We notice that, as the signaling traffic increases, the backhaul consumption reduction with $\mathcal{P}_{o/o/o}$ becomes higher than $\mathcal{P}_{\mathcal{J}/o/g}$. This suggests that, the more signaling traffic is significant, the more optimizing attachment can help reduce the overall traffic on the backhaul. Indeed, it reduces both signaling and data traffic, leading to improved gains in backhaul capacity. Nonetheless, the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

gain achieved by optimizing attachment is still marginal in comparison to the one achieved by optimizing association. This can be seen in Fig. 5.8 by focusing on $\mathcal{P}_{o/g/o}$, where user association is not optimized but attachment is. The consumption reduction in this case is much lower than the cases where association is optimized. Moreover, we can notice how the consumption reduction with $\mathcal{P}_{o/g/o}$ does not become significant until higher values of σ.

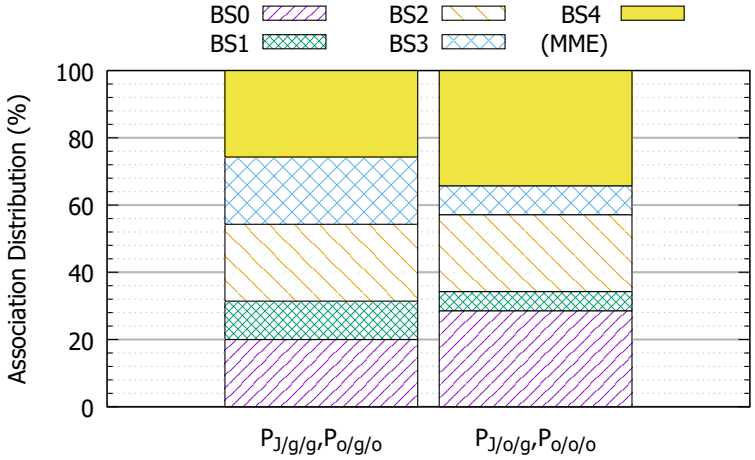


Figure 5.9: User association distribution in $\mathcal{P}_{\mathcal{J}/g/g}$, $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, and $\mathcal{P}_{o/o/o}$, for the studied topology and user distribution, with signaling traffic at $\sigma = 6\%$.

Fig. 5.9 shows the percentage of users associated to each of the BSs, for the user distribution in Fig. 5.5, and for $\sigma = 6\%$. In comparison with the given best SINR association in $\mathcal{P}_{\mathcal{J}/g/g}$ and $\mathcal{P}_{o/g/o}$, we notice that more UEs are associated to BS 4 and BS 0, when user association is optimized. One of the reasons is that BS 4 is the BS co-located with the MME, as returned by the optimal solution. Indeed, backhaul bandwidth consumption is reduced when UEs associate to BS 4 because the signaling traffic between the BS and the MME is not routed on the backhaul. Eventually, it seems that it is more beneficial to associate and/or attach to the BS co-located with the MME, if this is an option, as long as that BS can handle all the UEs from a RAN point of view. While this could raise questions on the achieved load balancing, we remind that RAN constraints are also accounted for in our problems, meaning no BS can take more UEs than allowed by its capacity.

5.6 Attachment per flow

Typically, each UE is attached to a default S-GW, and all the flows of that UE are forwarded to that S-GW to handle their routing locally. In this section, we rethink user attachment, and propose to optimize it on a per flow basis. We propose an attachment policy that handles attachment flow by flow. This means that the notion of a default S-GW for each UE no longer stands, since a UE can attach to different S-GWs for its different flows. Both UEs at both ends of a flow are attached to the same S-GW. Adding more flexibility to the attachment by allowing UEs to attach to multiple S-GWs at a time, in addition to each flow being handled by only one S-GW, are expected to reduce the backhaul consumption caused by signaling.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

5.6.1 Problem formulation

In such a scenario, the data and signaling paths for a flow between u and v, respectively associated to j_1 and j_2, are the same as the ones shown in Fig. 5.2. In this case, the single S-GW shown in Fig. 5.2 represents the S-GW to which a flow is attached (as opposed to representing the only S-GW in the network like in the context of $\mathcal{P}_{1/g/g}$). The difference between $\mathcal{P}_{o/g/of}$ and $\mathcal{P}_{o/o/of}$ is that user association is given in the first (X is an input), and optimized in the second (X is an output). Both problems have the same objective function and constraints explained in the following, and both are MIQP problems.

In both problems, the objective function in Eq. 5.3 remains the same, minimizing the total consumed bandwidth on the backhaul. Likewise, the constraints from 5.4 to 5.9 remain unchanged. What distinguishes those problems is that each flow, and not each UE, must be attached to one S-GW. We define flow attachment vector A, with $A_{f,j}$ a boolean such that $A_{f,j} = 1$ if flow f is attached to S-GW j. Constraint 5.24 states that a flow is attached to one and only one S-GW.

The data path of the flow f, as seen in Fig. 5.2, follows BS j_1, S-GW j_5, BS j_2. The corresponding cost of data traffic on each link, C_l^d, is computed in constraint 5.25. With one S-GW per flow f, there is signaling traffic between the MME and this S-GW, with a bit rate $S5_f$. In constraint 5.26, we compute C_l^{S5}, the total bandwidth consumed by the signaling traffic of all the flows $f \in \mathcal{F}$ on a given link l belonging to the signaling path between MME j_0 and S-GW j_5.

The constraints 5.10 to 5.12 are replaced by the following constraints, while constraint 5.13 remains the same:

$$\sum_{j \in \mathcal{J}} A_{f,j} = 1 \ , \ \ \forall f \in \mathcal{F} \tag{5.24}$$

$$C_l^d = \sum_{f \in \mathcal{F}} 2d_f \Big(\sum_{j_5 \in \mathcal{J}} A_{f,j_5} (\sum_{j_1 \in \mathcal{J}} X_{u,j_1} \cdot Z^l_{j_1,j_5} + \sum_{j_2 \in \mathcal{J}} X_{v,j_2} \cdot Z^l_{j_5,j_2}) \Big) \ , \ \ \forall l \in \mathcal{L} \tag{5.25}$$

$$C_l^{S5} = \sum_{f \in \mathcal{F}} \Big(\sum_{j_5 \in \mathcal{J}} A_{f,j_5} \sum_{j_0 \in \mathcal{J}} W_{j_0} \cdot Z^l_{j_0,j_5} \cdot S5_f \Big) \ , \ \ \forall l \in \mathcal{L} \tag{5.26}$$

5.6.2 Numerical results

Similarly to the previous section, we set as reference $\mathcal{P}_{\mathcal{J}/g/g}$, where both association and attachment are given. We consider the same example as in Section 5.4.4, with the same input data, network snapshots, and network topology in Fig. 5.5. Fig. 5.10 shows the relative backhaul bandwidth consumption reduction δ, with respect to $\mathcal{P}_{\mathcal{J}/g/g}$, function of signaling traffic.

For $\sigma = 0$, with an optimized association in $\mathcal{P}_{o/o/of}$, the backhaul consumption reduction is at 30%, the same when attachment is not optimized. For $\mathcal{P}_{o/g/of}$, with no association optimization, there is no consumption reduction. This proves, once more, that when signaling is not accounted for, attachment optimization is not necessary. On the other hand, we notice that the backhaul consumption reduction increases with the increase of signaling traffic. An attachment per flow outperforms both given and optimized attachment per user. For signaling traffic at $\sigma = 24\%$, the consumption reduction goes up to 40%, a 10% increase with respect to an attachment per user.

In order to decouple the gain brought by optimizing the association and the gain by optimizing attachment per flow, we compare δ for $\mathcal{P}_{o/g/of}$ and $\mathcal{P}_{o/o/of}$. Without an optimized association, the attachment per flow still seems to reduce backhaul bandwidth consumption, even when signaling

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

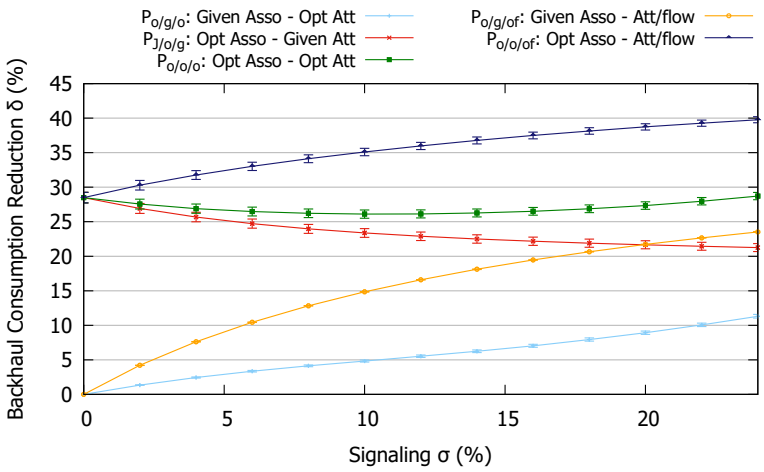


Figure 5.10: The relative backhaul bandwidth consumption reduction δ function of σ, in $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, $\mathcal{P}_{o/o/o}$, $\mathcal{P}_{o/g/of}$, and $\mathcal{P}_{o/o/of}$.

traffic is low, but more notably when the latter is high. However, this reduction remains less than the reduction achieved when association is optimized (whether with or without optimizing attachment, per user or per flow). Hence, while user association plays a major role in economizing bandwidth, the contribution of attachment optimization is secondary and signaling-dependent.

Eventually, the attachment per flow scheme, while seemingly profitable, is faced with two main obstacles rendering it less applicable: the computation complexity of the problem, on the one hand, and the implementation complexity it may require in practice, on the other hand. Indeed, selecting a different S-GW for each flow of a UE, instead of one default S-GW for all its flows as it is the case today, is a laborious task: it increases control traffic, as well as the delay.

5.7 Impact of network topology

Up to this point, all of the above results were obtained for the network topology in Fig. 5.5, but for different UE and traffic distributions. In order to verify that all of the conclusions can be generalized, regardless of the topology in question, we repeat the tests described above on various random network topologies with 5 BSs. The difference between one topology and another lies in the nodes positions and the interconnection among them. We consider a set of 21 distinct topologies, representing all the possible connected graphs topologies with 5 unlabeled nodes [133]. For each topology, tests are conducted on 100 network snapshots, each consisting of a user distribution/flow distribution combination, in the same conditions described in Section 5.4.4. Similarly to the previous tests, we set $\mathcal{P}_{\mathcal{J}/g/g}$ as reference, and compute the relative backhaul bandwidth consumption reduction δ, with respect to $\mathcal{P}_{\mathcal{J}/g/g}$. Fig. 5.11 shows the obtained average over the network topologies and snapshots, function of σ, with confidence level at 95%.

We observe in Fig. 5.11 that, for all the problems, the corresponding curves have the same shape as previously observed for one particular topology. The confidence intervals are also tight. Moreover, the numerical values of δ are at the same order of magnitude. Hence, these results corroborate the previous conclusions, and allow their generalization to different network topologies.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

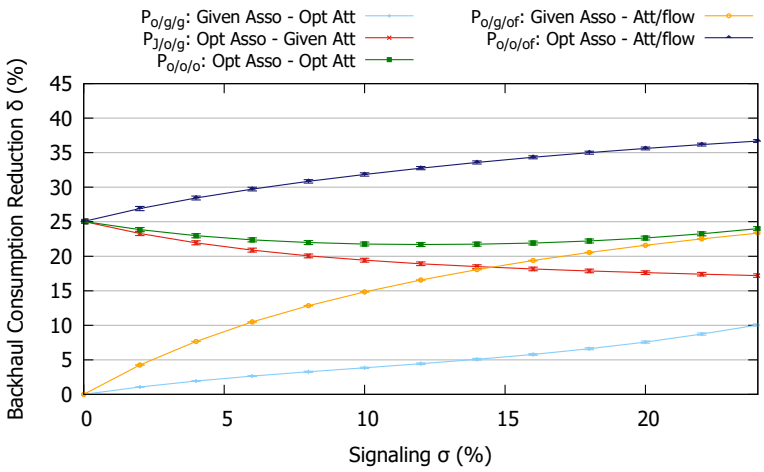


Figure 5.11: The average relative backhaul bandwidth consumption reduction δ function of σ, in $\mathcal{P}_{o/g/o}$, $\mathcal{P}_{\mathcal{J}/o/g}$, $\mathcal{P}_{o/o/o}$, $\mathcal{P}_{o/g/of}$, and $\mathcal{P}_{o/o/of}$, for different network topologies.

5.8 Conclusion

We addressed in this chapter three correlated problems in self-deployable networks: the Local CN functions placement, user association and user attachment, with the common objective of minimizing backhaul bandwidth consumption. For the placement problem, we compared two Local CN placement schemes: centralized and distributed. Results showed that having S-GWs distributed in the system, such that each BS is co-located with one S-GW is less costly from a backhaul point of view. Then, we assessed the backhaul consumption reduction when applying a backhaul-aware user association in comparison with a RAN-centric association. The backhaul-aware association achieved promising results, with a significant backhaul consumption reduction of 30%, in a signaling-free scenario. Finally, we compared two optimization attachment schemes, one per user, and one per flow, and a simplified attachment where each user attaches to the BS to which it is associated. When signaling is not accounted for, attachment schemes are practically worthless. The optimized attachment further reduced the backhaul bandwidth consumption, especially when signaling traffic was significant. The gain achieved by optimizing attachment, however, remained marginal in comparison to the gain achieved by the association scheme.

This work served as basis to validate a number of assumptions regarding the Local CN architecture, the unnecessity of sophisticated attachment schemes when signaling traffic is not accounted for, and the need for advanced association policies adapted to self-deployable networks. Therefore, building on the obtained results, we developed an elaborate network-aware user association policy, that we present in details in the next chapter.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 6

Network-Aware User Association Policy

6.1 Introduction

Once a self-deployable network is operational, users connect to the network as they arrive to benefit from the provided services. The first step is identifying the BS to which the user must associate. When a user is covered (i.e., can receive a sufficiently strong signal) from only one BS, there is no choice but to associate to that BS. However, when multiple BSs are available, a decision must be made [94]. This is the user association problem, that we tackle in this chapter.

User association is a prevalent problem in cellular networks. The importance of this problem emanates from two prominent challenges facing cellular networks in general: delivering a consistent QoE for users, on one hand, and coping with the scarce available spectrum, on the other hand. Indeed, when a user associates to a BS, a number of resources (e.g., PRBs) from that BS are allocated to that user, depending on his requirements, the strength of the received signal, and the amount of available resources. All these criteria must be balanced to guarantee that the association is done efficiently, in the sense that user's requirements are met while resources are efficiently used. Other noteworthy criteria could also affect user association, such as load balancing, delay, and energy efficiency [93]. Self-deployable networks further extend the association criteria list, whether because of the specific self-deployable network architecture (e.g., inter-BS backhaul with potentially limited bandwidth), or the data usage trends (e.g., minimum guaranteed throughput).

Numerous association policies have been proposed and adapted to different architectures [94]. In classical cellular networks, the most common association policy is based on the DL signal strength. The default association rule is to associate a user to the BS that provides the maximum DL SINR [99]. Such a basic association policy has been criticized for its poor load balancing [93] and for solely accounting for the DL, while ignoring both the UL [97], and the backhaul [106].

In self-deployable networks, in particular, we argue that user association must be completely rethought. Key parameters must be included when associating users, other than the DL signal strength, such as the backhaul capacity, the UL available resources, and users' requests. In fact, due to the self-deployable network architecture, and the potentially limited backhaul, the latter may occasionally create a bottleneck. We demonstrated, in the previous chapter, the significant impact of a backhaul-aware user association. We build on the obtained results to further include the backhaul state in the association process, such that the remaining backhaul resources are accounted for. On the other hand, UL traffic is becoming more and more important with UL-oriented asymmetric traffic applications, such as live video broadcasting, and symmetric traffic applications, such as video calls [109]. This increase in the UL traffic requires careful consideration of the limited UL resources, which, otherwise, can have a serious impact on the network performance. Hence, only considering DL resources when a user associates is not enough anymore, and UL should not be

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

overlooked. Furthermore, with users having higher QoS expectations, the user requirements in terms of bit rate must also be accounted for when associating users. In general, services fall back on the default bearer when a dedicated bearer with the required bit rate cannot be set up. With such a best effort approach, users are not always granted the bit rate they request for a particular service, causing a bad QoE. On the other hand, in some use cases of self-deployable networks, such as disaster relief networks, the traffic has stringent requirements in terms of bit rate constraints. For example, video streaming during a rescue operation cannot freeze or lag because of network problems. In that case, it is recommended to provide the users with the throughput they ask for or, by default, with the minimum acceptable throughput as requested by the service in question. Thus, we adopt flow blocking probability as the metric of interest when associating users, defined as the probability that a flow is rejected if it cannot be granted the throughput it asks for.

The contributions of this chapter can be summarized as follows. First, we propose a network-aware user association policy (NAS) that takes into account all the parameters discussed above, and balances the different requirements and constraints. The end result is an algorithm executed at the arrival of each flow request, that first enforces a flow admission procedure. By exploiting information on the requested flow data rate and the channel gains of the users involved in that flow, it is determined whether the available resources, on the access links and on the backhaul, can guarantee the required flow throughput. If the flow request is accepted, a decision on user association, and eventually on the routing path on the backhaul, is made. The association objective is to maximize a combination of the remaining available resources on the DL, the UL and the backhaul. The overall goal is to keep the flow blocking probability as low as possible.

Second, we show that applying the proposed network-aware association policy reduces the flow blocking probability remarkably, up to nine times, in comparison to a traditional association scheme, under different parameter settings. Furthermore, we show that our proposed policy adapts to different traffic types, network topologies, and network constraints, by mitigating the bottlenecks causing degraded performance, on the RAN and/or the backhaul.

Finally, we investigate several variants of the proposed association policy. We show that allowing re-association of already associated users with ongoing flows slightly improves a scheme where re-association is not allowed. Furthermore, we show that our proposition fits well with emerging network features, such as DL and UL split association [111], where a UE can be associated simultaneously to two distinct BSs, one for the DL, and one for the UL.

6.2 System model

We consider a self-deployable mobile network, where $\mathcal{J}$ is the set of BSs, $\mathcal{L}$ is the set of directional links interconnecting the BSs, and $\mathcal{U}$ is the set of fixed UEs. When two BSs are linked, there are 2 directional links between them, one in each direction.

6.2.1 Core network and backhaul

In the previous chapter, we showed that S-GWs should be distributed in the network, and that a good and simple attachment scheme consists of each UE attaching to the S-GW co-located with the BS it is associated to. Hence, we adopt this attachment scheme, and suppose that each BS in the network is co-located with a S-GW. One MME, co-located with one of the BSs, is serving the network. As previously shown in Chapter 5, signaling traffic mostly impacts user attachment and not user association. Since we only focus on user association, with a given attachment policy, we ignore signaling traffic and, consequently, the MME placement plays no part in our problem.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

For traffic destined outside the network, i.e., to an external packet data network, such as Internet or an IP multimedia subsystems (IMS), the inter-BS backhaul links carry data from one BS to another towards the P-GW. The P-GW is co-located with a designated BS that has a dedicated backhaul link to the solicited packet data network. This BS serves as point of entry and exit, to and from the network. We suppose that the P-GW and the MME are co-located with the same BS.

The same assumptions regarding the backhaul network formed by inter-BS links, as explained in Chapter 2 (Section 2.6, p. 16), are made. The backhaul and the RAN resources are independent, and there is no contention among backhaul links for resource utilization. We further consider, in this chapter, that links interconnecting the BSs have limited bandwidth. The maximum amount of traffic that can be carried by the link is also referred to as link capacity, in bits/second. We denote by C_l the capacity of a backhaul link $l \in \mathcal{L}$.

6.2.2 Traffic model

We model traffic as two co-exiting categories of flows: intra-network and inter-network. Intra-network flows are flows between two UEs that belong to the same network (Fig. 6.1, flow (a)). In this case, both UEs need to be associated to BSs within the network, and the association of both UEs is within the scope of our interest. Since core network functions are co-located with the BSs, data traffic of intra-network flows is only routed locally, i.e., on the inter-BS backhaul links.

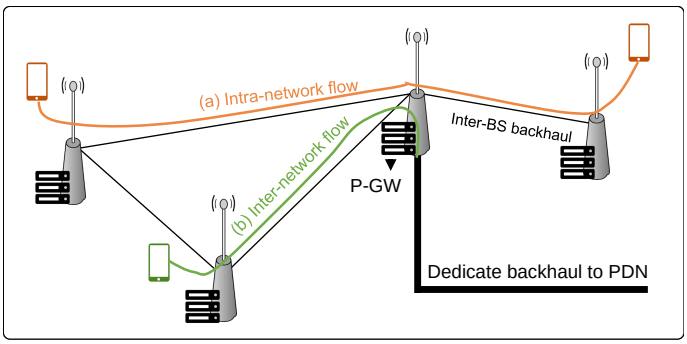


Figure 6.1: Intra-network and inter-network flows.

Inter-network flows are flows between a UE belonging to the network, on one side, and either a UE belonging to another network or an external server, on the other side. In other words, the flow is destined to (or originated from) an external packet data network (Fig. 6.1, flow (b)). In our architecture, only one designated BS, co-located with a P-GW, has a dedicated backhaul link towards the external packet data network. Hence, all inter-network flows are routed on the backhaul links towards (resp. from) this designated BS. Everything happening beyond this BS (i.e., outside the network) is not of interest to us, since our goal is to determine the association of the UE belonging to our network. In this case, an inter-network flow behaves exactly the same as an intra-network flow having one of the UEs already associated to the designated BS co-located with the P-GW (but without consuming any RAN resources on that BS). Therefore, for all inter-network flows, we model the party belonging to the external network as a permanently-associated UE to the designated BS with no RAN consumption. For brevity, when we talk hereafter about a flow between two UEs u and v, inter-network flows, even between a UE and a external server, are also included.

We model all flow requests as bidirectional, meaning they are composed of two directional flows, and the throughput requirements can be different on each direction. One UE can have

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

several simultaneous flows with different UEs within or outside the network, or with data servers. A UE with no ongoing flows is not associated to any BS. To transmit or/and receive data, a UE should be associated to one and only one BS. If a UE is already associated (because of an ongoing flow), it is not allowed to re-associate upon the arrival of a new flow. Moreover, we consider a joint DL/UL association, in which a UE is associated to the same BS for both DL and UL traffic. Both assumptions, on re-association and joint DL/UL association, will be revisited and investigated in Section 6.8.1 and Section 6.8.2, respectively.

Flow requests arrive between two random UEs according to a Poisson process, with an average flow arrival rate λ_f, and flow duration follows an exponential distribution, with an average μ_f. We denote by β the percentage of intra-network flow requests out of all flow requests. We denote by d_{uv} and d_{vu} the requested data rates, in each direction, respectively, for a flow between u and v.

When two UEs are associated to the same BS, the data they exchange does not need to be routed on the backhaul. However, when two UEs are associated to two different BSs, data traffic is routed on the backhaul links. Since two BSs are not necessarily directly connected by a link, a routing path between them must be determined. A routing path, denoted $P(i,j)$, consists of a succession of directional backhaul links on which traffic is routed from BS i to BS j. Routing paths in both directions, $P(i,j)$ and $P(j,i)$, are not necessarily identical.

6.2.3 Radio access network

We adopt an OFDMA-based model, similar to the one described in Chapter 5 for the DL (Section 5.2.3, p. 76). Furthermore, with respect to the previous chapter, we elaborate an UL model. Recall that M is the total number of orthogonal channels, divided among the BSs, following a given frequency reuse scheme, with reuse factor r. The number of channels reserved for each BS are denoted $\mathcal{K}_j^{DL}$ and $\mathcal{K}_j^{UL}$, on the DL and on the UL, respectively. We denote by $\mathcal{I}_j^{DL}$ and $\mathcal{I}_j^{UL}$ the set of BSs sharing the same channels as BS j, on the DL and UL, respectively. We suppose that all BSs are identical in terms of maximum transmit power, denoted P_{BS}, and antenna gain, denoted G^a. We denote by P_{UE} the maximum UE transmit power.

If UE u is the only UE associated to BS j, then u would receive its maximum per channel rates both on the DL and the UL. These values are denoted as $R_{u,j}^{DL}$ and $R_{u,j}^{UL}$, respectively. They are determined by function $\Psi(\cdot)$, that maps the per channel SINR $\gamma_{u,j}^{DL/UL}$ to the per channel data rate, such that $R_{u,j}^{DL/UL} = \Psi(\gamma_{u,j}^{DL/UL})$. Function $\Psi(.)$ is a discrete step function, such that:

$$R_{u,j}^{DL/UL} = \mathfrak{r}_\theta \text{ ,if } \gamma_{u,j}^{DL/UL} \in [\theta, \theta+1), \tag{6.1}$$

with the corresponding values of θ and $\mathfrak{r}_\theta$ determined based on the modulation and coding scheme.

On the DL, the per channel SINR, $\gamma_{u,j}^{DL}$, between UE u and BS j is defined as:

$$\gamma_{u,j}^{DL} = \frac{\frac{P_{BS}}{\mathcal{K}_j^{DL}} \cdot G_{u,j}}{\mathcal{N}_0 + \sum_{h \in \mathcal{I}_j^{DL}} \frac{P_{BS}}{\mathcal{K}_h^{DL}} \cdot G_{u,h}} \ , \tag{6.2}$$

where $\mathcal{N}_0$ is the per channel additive white Gaussian noise power, $G_{u,j}$ is the channel gain between UE u and BS j that accounts for the path loss $\gamma_{u,j}$, shadow fading, antenna gain G^a, and equipment losses E, all expressed in dB. The path loss $\Gamma_{u,j}$ between u and j is written as $\Gamma_{u,j} = a + b \cdot \log(D_{u,j})$ [116], where a and b are standard coefficients that depend on the path loss model, and $D_{u,j}$ is the distance between UE u and BS j. We model the shadow fading through a normal distribution.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

The per channel SINR on the UL between UE u and BS j is written as:

$$\gamma_{u,j}^{UL} = \frac{P_{UE} \cdot G_{u,j}}{m_{u,j}^{UL} \cdot (\mathcal{N}_0 + I_j^{UL})} , \tag{6.3}$$

where $m_{u,j}^{UL}$ is the number of UL channels that must be allocated to UE u on BS j to get the throughput it asks for, and I_j^{UL} is a conservative estimate of the per-channel UL interference.

Interference on the UL and DL are inherently different. The interference on the DL is mainly caused by the BSs sharing the same DL channels, whereas the interference on the UL is mainly caused by UEs transmitting on the same UL channels, at the same time (i.e., on the same PRB). Depending on the scheduling within a BS, not all UEs are using all the channels of that BS at a given time. Hence, not all UEs of a BS are interfering with the UEs of a neighboring BS sharing the same UL channels. Determining which of the UEs are the actual interferers is not trivial, and requires the knowledge of the instantaneous UL scheduling scheme (on a PRB level). Therefore, computing UL interference is generally based on conservative estimators.

The average per-channel UL interference estimation, denoted I_j^{UL}, is caused to UEs associated to BS j by the UEs associated to the interfering cells $h \in \mathcal{I}_j^{UL}$. We assume that the interference is the same on all channels. Since we do not know which UE in a BS h is the one actually interfering, we have to compute an "average" interference (we refer to the supposed interferer as an average user of the BS). Hence, we compute $\bar{G}_{j,h}$, the channel gain between BS j and an average user in BS h. $\bar{G}_{j,h}$ is computed as the average channel gain between a (large) number of possible user locations in BS h ($\mathfrak{n}_u$), and BS j, such that:

$$\bar{G}_{j,h} = \frac{\sum_{u=1}^{\mathfrak{n}_u} G_{u,j}}{\mathfrak{n}_u} \tag{6.4}$$

We assume that all UEs in BS h get equal number of channels. Each UE allocates its transmission power P_{UE} on its channels, such that the power is equally divided among those channels. Hence, if N_h UEs are actually associated to BS h, the average per channel power budget is written as $\frac{P_{UE} \cdot N_h}{\mathcal{K}_j^{UL}}$.

The average interference on a UE associated to BS j, caused by an average UE associated to BS $h \in \mathcal{I}_j^{UL}$, is equal to the product of the average channel gain and the average power budget. Therefore, the UL interference on a UE of BS j, denoted I_j^{UL}, is written as:

$$I_j^{UL} = \sum_{h \in \mathcal{I}_j} \frac{P_{UE} \cdot N_h \cdot \bar{G}_{j,h}}{\mathcal{K}_j^{UL}} \tag{6.5}$$

6.2.4 Resource allocation

Typically, for each of its flows, a UE would receive a number of PRBs, via scheduling. PRB allocation and scheduling is, however, out of the scope of this chapter. To maintain the tractability of the framework, we model resource allocation as an allocation of fraction of the available channels to each flow, on the UL and the DL. The number of channels allocated to a UE from a BS, for each of its flows, depends on its requirement and on the data rate it observes from the BS.

To simplify notation, we use hereafter the notation DL/UL as a superscript on the variables within an equation, when the latter is written in the same way for both the DL and the UL counterparts. We denote by $m_{u,j}^{DL}$ the number of channels needed by UE u from BS j on the DL, to

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

obtain a requested throughput d. Likewise, $m_{u,j}^{UL}$ is defined on the UL. We compute this as:

$$m_{u,j}^{DL/UL} = \frac{d}{R_{u,j}^{DL/UL}} \tag{6.6}$$

Computing $m_{u,j}^{DL}$ is straightforward under the assumption that all the channel gains are available. The SINR from each BS is computed (Eq. 6.2), then $R_{u,j}^{DL}$ is deduced (Eq. 6.1).

On the UL, computing $m_{u,j}^{UL}$ is trickier, since it depends on the UL per channel rate $R_{u,j}^{UL}$. To compute $R_{u,j}^{UL}$, the SINR on the UL must first be known, or at least estimated. However, $SINR_{u,j}^{UL}$ depends on $m_{u,j}^{UL}$ (Eq. 6.3). Hence, the computations of $\gamma_{u,j}^{UL}$ and $m_{u,j}^{UL}$ cannot be done separately, as in the DL case. We develop, in the following, a method to compute $m_{u,j}^{UL}$. Using step function $\Psi(.)$, to get the rate $\mathfrak{r}_{\theta_i}$, the SINR value must be in the range $[\theta_i, \theta_{i+1})$. Therefore, to get a throughput d with a rate $\mathfrak{r}_{\theta_i}$, we need to find the SINR level θ_i such that:

$$\frac{X}{\theta_{i+1}} < \frac{d}{\mathfrak{r}_{\theta_i}} \leq \frac{X}{\theta_i}, \tag{6.7}$$

where $X = P_{UE} \cdot G_{u,j}/(\mathcal{N}_0 + I_j^{UL})$, and the above relations are found using Eq. 6.3 and Eq. 6.6. Eq. 6.7 can be satisfied by zero, one or several values of $\mathfrak{r}_{\theta_i}$. If no rate satisfies Eq. 6.7, then u cannot get the requested throughput on the UL from BS j. If there is a unique solution, then $R_{u,j}^{UL} = \mathfrak{r}_{\theta_i}$. If there are several solutions $\mathfrak{r}_{\theta_i}$, the maximum rate is selected, since this leads to the minimum number of used channels. Once $R_{u,j}^{UL}$ is found, $m_{u,j}^{UL}$ can be computed from Eq. 6.6.

6.3 Association policy overview

The association policy presented in the following comprises two phases: a flow admission control and an association decision. The whole procedure is triggered by each arrival of a flow request.

Since we are assuming that re-association is not allowed (a UE remains associated to the same BS even when a new flow arrives), at each arrival of a flow request, the network has to check whether the UEs are already associated or not. Note that this assumption is revisited in Section 6.8.1. In the flow admission control phase, for the already associated UE(s), the network needs to check if there are enough resources on their BS(s) to accommodate the new flow. For the UE(s) that are not yet associated, the network needs to check if there are available BS(s) with enough resources, and available backhaul bandwidth (when needed), to accept the arriving flow. Recall that, for inter-network flows, all these tests only concern the UE that belongs to the network. For the sake of brevity, we represent in the following the general case where the UEs involved in a flow request are not already associated to a BS. The cases where one or both UEs are already associated can be simply deduced. This applies to all formulations in Section 6.4 and Section 6.5.

For each bi-directional flow request between UEs u and v, that are not already associated, of data rate requirements $d_{u,v}$ and $d_{v,u}$, the eventual goal is to find a pair of BSs (j_u, j_v), such that the following constraints are fulfilled simultaneously: *i)* UE u associates to BS j_u and it can get a throughput $d_{v,u}$ on the DL and $d_{u,v}$ on the UL; *ii)* UE v associates to BS j_v and it can get a throughput $d_{u,v}$ on the DL and $d_{v,u}$ on the UL; *iii)* there is a routing path $P(j_u, j_v)$ on the backhaul between j_u and j_v that has enough capacity to carry the flow of data rate $d_{u,v}$; and *iv)* there is a routing path $P(j_v, j_u)$ on the backhaul between j_v and j_u that has enough capacity to carry the flow of data rate $d_{v,u}$. When u and/or v are already associated, j_u and/or j_v are given.

If no pair (j_u, j_v) is found, then the flow is rejected. If only one pair (j_u, j_v) fulfills the constraints above, then UEs u and v are directly associated to BSs j_u and j_v, respectively. If more

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

than one pair fulfill the constraints, then the best pair is selected with the aim of maximizing the remaining network resources, by accounting for both the radio access and the backhaul networks.

Several input data are needed to make the admission and/or the association decision: the UEs at both ends of the flow, their required throughput, their channel gains with the BSs in the network, and the network state represented by the available resources on the DL, the UL, and the backhaul. One way to implement such a strategy is to have a centralized control entity that takes these parameters as input, from the UEs and the BSs, does the necessary computations, and outputs the association decision. The practical implementation, however, is out of the scope of this chapter.

6.4 Flow admission control

The flow admission control phase consists of verifying whether there are enough resources to accept an arriving flow, before trying to find a suitable association.

6.4.1 Association feasibility on the RAN

On the RAN, each UE is granted a fraction of the UL and DL channels of the BS it is associated to, depending on its requested throughput. We denote by M_j^{DL} (resp. M_j^{UL}) the remaining DL (resp. UL) channels at BS j. Therefore, UE u can consider BS j as a candidate for association on the DL if and only if $m_{u,j}^{DL} \leq M_j^{DL}$. Similarly, UE u can consider BS j as a candidate for UL association if and only if $m_{u,j}^{UL} \leq M_j^{UL}$. Since we are considering a joint DL/UL association, a BS j is a candidate for the association of UE u if and only if it can be considered as a candidate on both DL and UL.

The first step in the flow admission control is to check the candidate BSs for association, and eliminate the BS pairs that are not *access-feasible*. A BS pair (i, j) is said to be access-feasible if and only if BS i is a candidate for u, and BS j is a candidate for v, both on the DL and the UL. Let $\mathbb{F}_u$ be the set of candidate BSs for a UE u. Consequently, a BS pair (i, j) is access-feasible if and only if $(i, j) \in \mathbb{F}_u \times \mathbb{F}_v$. To determine $\mathbb{F}_u$ and $\mathbb{F}_v$, at the arrival of a flow request between u and v, the number of channels needed by each of the two UEs from each BS $j \in \mathcal{J}$ on the DL and the UL are computed, i.e., $m_{u,j}^{DL}$, $m_{v,j}^{DL}$, and $m_{u,j}^{UL}$, $m_{v,j}^{UL}$. Then, they are compared to the number of remaining channels on this BS j, on the DL and on the UL, i.e., M_j^{DL} and M_j^{UL}, respectively.

6.4.2 Association feasibility on the backhaul

On the backhaul, a directed flow consumes, on each link of the selected routing path, an amount of bandwidth equal to its data rate. We denote by A_l the remaining capacity of a link $l \in \mathcal{L}$. As flows start and end, the remaining capacities of the links vary. A flow between two UEs, associated to BSs i and j, respectively, follows one path $P(i, j)$. A path $P(i, j)$ is said to be feasible, i.e., a candidate to route the flow from i to j, if the remaining capacity on all of its links can accommodate the flow data rate d, i.e., $d \leq A_l \ , \ \forall \ l \in P(i, j)$.

The second step in the flow admission control is to check the feasible routing paths between the BSs of each access-feasible pair, and eliminate the BS pairs that are not *backhaul-feasible*, to reduce the search space. A BS pair (i, j) is said to be backhaul-feasible if and only if there exists at least one feasible routing path in each direction, i.e., a feasible routing path $P(i, j)$ and a feasible routing path $P(j, i)$. Let $\mathcal{P}_{i,j}$ be the set of feasible paths from i to j. Hence, a BS pair (i, j) is backhaul-feasible if and only if $\mathcal{P}_{i,j} \neq \emptyset$ and $\mathcal{P}_{j,i} \neq \emptyset$. We denote by $z(i, j)$ a binary variable that indicates if a BS pair (i, j) is backhaul-feasible or not. If $z(i, j) = 0$, the BS pair (i, j) is not

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

backhaul-feasible and cannot be considered for association. We define $z(i,i) = 1,\ \forall\ i \in \mathcal{J}$, since a pair (i,i) is always backhaul-feasible with no routing path actually needed.

6.4.3 Association feasibility

In order for it to be considered as a candidate for the association of UEs u and v, a BS pair must be both backhaul-feasible and access-feasible. Therefore, we define $\mathbb{F}$ as the set of all feasible BS pairs (i,j), such that:

$$\mathbb{F} = \{(i,j) \in \mathbb{F}_u \times \mathbb{F}_v : z_{ij} = 1\} \tag{6.8}$$

If $\mathbb{F} = \emptyset$, the flow is blocked. With this information, we can compute the flow blocking probability and use it as a performance evaluation metric. If $\|\mathbb{F}\| = 1$, the flow is accepted, and the UEs are associated accordingly. In this case, only a routing path between the BSs must be selected if there are multiple feasible paths in $\mathcal{P}_{i,j}$. Any routing metric can be applied in this case. We adopt the following: for each routing path, we find the link with the least remaining capacity, then, among those links, find the one with the maximum remaining capacity, and select the corresponding routing path. In other words, we select the path with the least most loaded link (more details on the routing path selection are presented in the next section). When multiple options are possible for the association, i.e., $\|\mathbb{F}\| > 1$, the flow is accepted but an association decision must be made, as well as a routing decision.

6.5 User association decision

Following the flow admission control, if a flow request between UEs u and v is accepted, a set of feasible BS pairs $\mathbb{F}$ for user association is returned (Eq. 6.8). For the case of $\|\mathbb{F}\| > 1$, an optimal BS pair, and the corresponding routing paths, must be selected from $\mathbb{F}$. Our goal is to have this selection maximize the remaining resources in the network, by jointly considering the remaining resources on the backhaul, the DL, and the UL. Hence, we compute for each of these three elements a metric representing the remaining resources following the association. All the metrics are normalized into values between 0 and 1.

RAN - We define $L^{DL}_{x,y} \in [0,1], \forall\ (x,y) \in \mathbb{F}$, as the normalized remaining DL channels, on the BS in pair (x,y) with the least remaining DL channels, if the pair was selected for the association of two UEs u and v. For each feasible BS pair (x,y), $L^{DL}_{x,y}$ is written as:

$$L^{DL}_{x,y} = \min\left(\frac{M^{DL}_x - m^{DL}_{u,x}}{\mathcal{K}^{DL}_x}, \frac{M^{DL}_y - m^{DL}_{v,y}}{\mathcal{K}^{DL}_y}\right) \tag{6.9}$$

Similarly, $L^{UL}_{x,y} \in [0,1], \forall\ (x,y) \in \mathbb{F}$ is defined for the UL. For each feasible BS pair (x,y), $L^{UL}_{x,y}$ is written as:

$$L^{UL}_{x,y} = \min\left(\frac{M^{UL}_x - m^{UL}_{u,x}}{\mathcal{K}^{UL}_x}, \frac{M^{UL}_y - m^{UL}_{v,y}}{\mathcal{K}^{UL}_y}\right) \tag{6.10}$$

The higher $L^{DL}_{x,y}$ (resp. $L^{UL}_{x,y}$), the better the corresponding pair (x,y) is from a DL (resp. UL) point of view. For all feasible BS pairs (i,j) in $\mathbb{F}$, we must compute $L^{DL}_{i,j}$ and $L^{UL}_{i,j}$, using Eq. 6.9 and Eq. 6.10, respectively.

Backhaul - We define $L^{BH}_{P(x,y)} \in [0,1],\ \forall\ P(x,y) \in \mathcal{P}_{x,y},\ \forall\ (x,y) \in \mathbb{F}$, as the normalized remaining capacity of the link with the least remaining capacity, among the links of a feasible routing path from a BS x to BS y, if it was selected as a routing path for a flow between to UEs of demand

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

d. Recall that the maximum capacity of a link l is denoted C_l, and the remaining capacity of the link is denoted A_l. Then, for each path $P(x,y)$, $L^{BH}_{P(x,y)}$ is written as:

$$L^{BH}_{P(x,y)} = \begin{cases} \min\limits_{l \in P(x,y)} \left(\frac{A_l - d}{C_l} \right), \text{ if } x \neq y \\ 1, \text{ if } x = y \end{cases} \tag{6.11}$$

The higher $L^{BH}_{P(x,y)}$, the better the path $P(x,y)$ is. For all feasible BS pairs (i,j) in $\mathbb{F}$, we must compute $L^{BH}_{P(i,j)}$, $\forall\ P(i,j) \in \mathcal{P}_{i,j}$, and $L^{BH}_{P(j,i)}$, $\forall\ P(j,i) \in \mathcal{P}_{j,i}$, using Eq. 6.11.

$L^{BH}_{P(i,j)}$, $L^{BH}_{P(j,i)}$, $L^{DL}_{i,j}$ and $L^{UL}_{i,j}$ are computed for all feasible pairs (i,j) in $\mathbb{F}$ and their corresponding feasible paths in $\mathcal{P}_{i,j}$ and $\mathcal{P}_{j,i}$. To establish some kind of fairness among the metrics, i.e., to equally consider the DL, the UL, and the backhaul, we choose to maximize their product. This is to avoid biased results when one metric is very high while the other is negligible, for example. It is enough for one metric to be negligible to lower the chances of a pair to be selected. Therefore, the selected pairs for association (j_u, j_v), and for routing $(P(j_u,j_v), P(j_v,j_u))$ are the pairs that maximize the product of the four metrics, such that:

$$\arg \max_{\substack{P(i,j) \in \mathcal{P}_{i,j} \\ P(j,i) \in \mathcal{P}_{j,i} \\ (i,j) \in \mathbb{F}}} L^{BH}_{P(i,j)} \cdot L^{BH}_{P(j,i)} \cdot L^{DL}_{i,j} \cdot L^{UL}_{i,j} \tag{6.12}$$

To summarize, the association algorithm unfolds as follows. At each flow request between UEs u and v, flow admission control is performed to determine whether the flow is to be accepted or rejected. If accepted, a list of feasible BS pairs, candidates for user association, is determined, together with their corresponding feasible routing paths. Then, one BS pair (j_u, j_v) and one routing paths pair $(P(j_u,j_v), P(j_v,j_u))$ are chosen according to Eq. 6.12, such that UEs u and v are associated to BS j_u and j_v, respectively, and the flows between j_u and j_v are routed on paths $P(j_u,j_v)$ and $P(j_v,j_u)$, in each direction respectively.

Once associated, BSs j_u and j_v must update the number of remaining channels on their DL and UL, i.e., $M^{DL}_{j_u}$, $M^{UL}_{j_u}$, $M^{DL}_{j_v}$, and $M^{UL}_{j_v}$, by subtracting the allocated channels for the flow. The remaining capacities of the backhaul links on paths $P(j_u,j_v)$ and $P(j_v,j_u)$ must also be updated. When the flow ends, the remaining channels on the the DL and the UL, as well as the remaining capacities on the backhaul links, must also be updated, by re-adding the resources that were allocated to the ending flow.

6.6 Problem formalization

The computation complexity to select one pair of BSs increases with the number of feasible BS pairs in $\mathbb{F}$. Typically, only few BS pairs are feasible in self-deployable networks, and only few routing paths between them are available, considering the network limited size. This simplifies the computation, and allows a relatively quick selection. However, in case of dense, or even ultra-dense networks, for example, where the number of BSs is high, and may even surpass the number of active users [101], brute force computations are not as practical. For such cases, the problem can be formulated as an MINLP, that is solved at each flow request to return the users' association and the routing path. In our use case of self-deployable networks, the network size allowed the simpler use of the above-mentioned brute force computations.

Given $\mathcal{J}$, $\mathbb{F}$, $\mathcal{P}_{i,j}$, $\mathcal{P}_{j,i}$, $\mathcal{K}^{DL}_j$, $\mathcal{K}^{UL}_j$, M^{DL}_j, M^{UL}_j, $m^{DL}_{u,j}$, $m^{DL}_{u,j}$, $m^{DL}_{v,j}$, and $m^{UL}_{v,j}$, we define the optimization problem $\mathcal{P}1$. For a flow between UEs u and v, the objective function in Eq. 6.13 returns a BS pair (i,j), such that UEs u and v are associated to i and j, respectively, and two routing paths p and p', one in each direction. We define x_j, y_j, a_p and b_p as binary variables, such

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

that $x_j = 1$ if UE u associates to BS j, $y_j = 1$ if UE v associates to BS j, $a_p = 1$ if routing path p is selected from u to v, and $b_p = 1$ if routing path p is selected from v to u. Constraints 6.14 and 6.15 state that a UE is associated to one and only one BS, while constraint 6.16 indicates that UEs u and v cannot associate to BSs i and j, if the pair (i, j) is not in the feasible pairs set $\mathbb{F}$. Constraints 6.17 and 6.18 state that only one routing path is selected for each direction. In constraint 6.19 and constraint 6.21, we compute the normalized remaining capacities on each link of all the feasible paths in case they were selected for routing, denoted Λ_p^l and ρ_p^l for each direction, respectively. Constraint 6.20 and constraint 6.22 return the metrics L_{BH} and L'_{BH}, as the minimum of Λ_p^l and ρ_p^l, respectively. In constraint 6.23 and constraint 6.26, we compute ζ_j^{DL} and ζ_j^{UL} as the normalized number of remaining DL and UL channels, respectively, on each BS in case the UE(s) associates to it. Constraints 6.24 and 6.25 (resp. constraints 6.27 and 6.28) return the metric L^{DL} (resp. L^{UL}) as the minimum of ζ_j^{DL} (resp. ζ_j^{UL}). This formulation represents the case where both UEs of a flow request are not already associated. The case where u is already associated to BS j is deduced by considering $x_j = 1$ as input parameter.

$$\mathcal{P}\mathbf{1}: \max_{\substack{(x_i),(y_i),(a_p),(b_{p'}) \\ L_{DL},L_{UL} \\ L_{BH},L'_{BH}}} \sum_{\substack{(i,j)\in\mathbb{F} \\ p\in\mathcal{P}_{i,j} \\ p'\in\mathcal{P}_{j,i}}} x_i y_j a_p b_{p'} L_{DL} L_{UL} L_{BH} L'_{BH} \tag{6.13}$$

$$\sum_{j\in\mathcal{J}} x_j = 1 \tag{6.14}$$

$$\sum_{j\in\mathcal{J}} y_j = 1 \tag{6.15}$$

$$x_i y_j = 0\ ,\ \forall (i,j) \notin \mathbb{F} \tag{6.16}$$

$$\sum_{p\in\mathcal{P}_{i,j}} a_p = 1 \tag{6.17}$$

$$\sum_{p\in\mathcal{P}_{j,i}} b_p = 1 \tag{6.18}$$

$$\Lambda_p^l = \frac{A^l - a_p d}{C^l}\ ,\ \ \forall l \in p,\ \forall p \in \mathcal{P}_{i,j},\ \forall (i,j) \in \mathbb{F} \tag{6.19}$$

$$\Lambda_p^l \geq a_p L_{BH}\ ,\ \ \forall l \in p,\ \forall p \in \mathcal{P}_{i,j},\ \forall (i,j) \in \mathbb{F} \tag{6.20}$$

$$\rho_p^l = \frac{A^l - b_p d}{C^l}\ ,\ \ \forall l \in p,\ \forall p \in \mathcal{P}_{j,i},\ \forall (i,j) \in \mathbb{F} \tag{6.21}$$

$$\rho_p^l \geq a_p L'_{BH}\ ,\ \ \forall l \in p,\ \forall p \in \mathcal{P}_{j,i},\ \forall (i,j) \in \mathbb{F} \tag{6.22}$$

$$\zeta_j^{DL} = \frac{M_j^{DL} - x_j m_{u,j}^{DL} - y_j m_{v,j}^{DL}}{\mathcal{K}^{DL}}\ ,\ \ \forall j \in \mathcal{J} \tag{6.23}$$

$$\zeta_j^{DL} \geq x_j L_{DL}\ ,\ \ \forall j \in \mathcal{J} \tag{6.24}$$

$$\zeta_j^{DL} \geq y_j L_{DL}\ ,\ \ \forall j \in \mathcal{J} \tag{6.25}$$

$$\zeta_j^{UL} = \frac{M_j^{UL} - x_j m_{u,j}^{UL} - y_j m_{v,j}^{UL}}{\mathcal{K}^{UL}}\ ,\ \ \forall j \in \mathcal{J} \tag{6.26}$$

$$\zeta_j^{UL} \geq x_j L_{UL}\ ,\ \forall j \in \mathcal{J} \tag{6.27}$$

$$\zeta_j^{UL} \geq y_j L_{UL}\ ,\ \forall j \in \mathcal{J} \tag{6.28}$$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

6.7 Numerical evaluation

In the following, we compare a traditional best SINR association policy (denoted as SINR in the figures), in which the UE associates to the BS from which it gets the maximum SINR on the DL, with our proposed network-aware policy (denoted NAS). The evaluation is based on network simulations, where flow requests arrive in the network, between two random users, according to a Poisson process. A flow admission control followed by an association decision, when needed, are executed for each flow request. The comparison between the two association policies relies on the flow blocking probability, i.e., the ratio of blocked flows over the total number of flow requests. We evaluate the performance of both policies with respect to various parameters, such as the intra-network flows percentage β, the average flow arrival rate λ_f, the backhaul links capacity C_l, and the network topology.

6.7.1 Simulation settings

We set $M = 120$ orthogonal channels, equally divided between the DL and the UL. Channels and power are equally divided among the BSs with an EP-R3 scheme. For the channel reuse scheme, we suppose that the set of interfering BSs on the DL and the UL, $\mathcal{I}_j^{DL}$ and $\mathcal{I}_j^{UL}$, are identical. To determine $\mathcal{I}_j^{DL}$ (resp. $\mathcal{I}_j^{UL}$), we rely on a Voronoi diagram that partitions the area into distinct regions, referred to as Voronoi cells. Each region corresponds to one BS, and consists of all the points closer to that BS than to any other. The channel reuse scheme is such that there are 3 sets of channels, and BSs with adjacent Voronoi cells do not use the same set. We consider a distance-based path loss model for an urban setting, with $\Gamma_{u,j} = 128.1 + 37.6 \cdot \log(D_{u,j}/1000)$ dB [116]. We set $P_{BS} = 46$ dBm, $G^a = 10$ dB, $E = 20$ dB, $N_0 = -121$ dBm, and we model shadow fading using a normal distribution with zero mean and standard deviation 8 dB. The numerical values of the discrete rate function $\Psi(.)$ are shown in Table 5.2 (Chapter 5, p. 84) [132]. We suppose there are 100 UEs, randomly distributed in the area. Flows arrive between random UEs following a Poisson process with an average flow arrival rate λ_f, and a duration following an exponential distribution of average $\mu_f = 180$ s. We fix μ_f, in the following, and vary λ_f. We suppose that all the flows request a data rate $d = 1$ Mb/s, in both directions. We revisit this assumption on traffic symmetry afterwards, and evaluate performance with asymmetric traffic.

We first consider the network topology shown in Fig. 6.2, with 6 BSs in a square of area 1 unit square. We model the network as a connected random geometric graph with radius $\eta = 0.4$. The Voronoi cells are shown in Fig. 6.2, and the interfering BSs are the ones highlighted with the same color. We note that several topologies are studied afterwards to generalize the results. We suppose that all backhaul links in $\mathcal{L}$ have the same capacity C_l.

Henceforth, all the results are averaged over 30 simulations, with each simulation corresponding to a different user distribution, with 3000 flow requests. The confidence level is set at 95%. All the simulations are done using a home-built Python simulator based on SimPy, the process-based discrete-event simulation library [121], on a server with Intel Xeon CPU E5-2697 v2 @ 2.7 GHz.

6.7.2 Simulation results

6.7.2.1 Impact of the traffic model

First, we study the effect of having inter-network and/or intra-network flows on the association policy performance, by varying the parameter β that represents the percentage of intra-network flows out of the total number of flow requests. We test different values of the inter-BS backhaul links capacities C_l, representing limited and well-provisioned backhaul capacities. We set the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

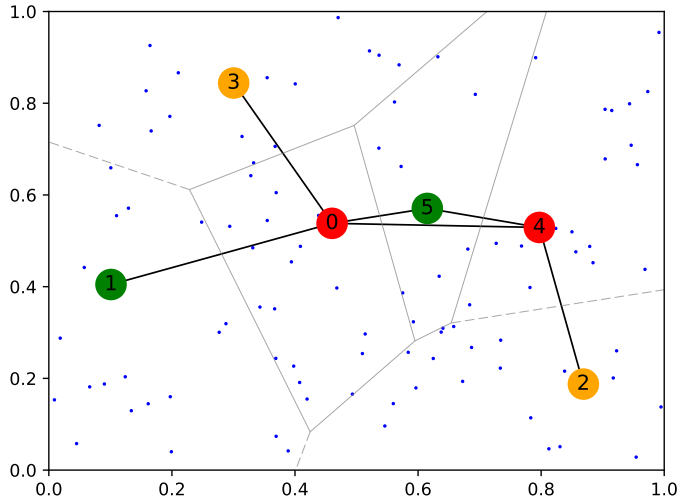


Figure 6.2: A network topology.

average flow arrival rate to $\lambda_f = 0.01$ s^{-1}. Results are shown in Fig. 6.3, for $C_l = 1$ Mb/s, and $C_l = 5$ Mb/s.

From Fig. 6.3, we can deduce that the NAS policy largely outperforms a best SINR policy, achieving significantly lower flow blocking probabilities, for all values of β and C_l. With best SINR, when the backhaul is limited ($C_l = 1$ Mb/s), the blocking probability is very high, ranging between 30% and 40 % for the different values of β. This confirms that a best SINR policy, oblivious to the backhaul state, is not suitable at all for backhaul-limited networks. The blocking is reduced by a factor of 10 with NAS. For $C_l = 5$ Mb/s, the blocking probability is lower in both cases since the backhaul is not a bottleneck anymore. However, it remains high with a best SINR association. Detailed observations on the flow blocking causes and how they are mitigated with NAS, in each scenario, are presented in the upcoming analysis in Section 6.7.2.3.

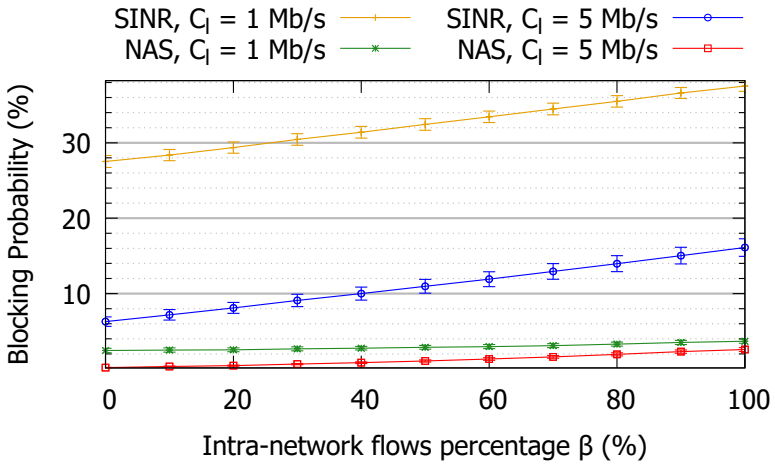


Figure 6.3: Blocking probability function of the percentage of intra-network flows out of all flow requests β, for different inter-BS link capacities C_l.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

For all values of C_l, the flow blocking probability increases with the increase of β. When all flows are inter-network, i.e., $\beta = 0$, the blocking probability is lower, because each flow is only consuming RAN resources on one BS, as we are only associating one user per flow. This is not the case for $\beta > 0$, when there are intra-network flows, since we associate two users per flow, both consuming RAN resources, which increases the load in the network, and consequently the blocking probability. The increase of the blocking probability with respect to β seems more significant with best SINR than with NAS. These results indicate that NAS can adapt to limited and well-provisioned backhaul capacities, and both inter and intra-network flows, and reduce the blocking probability in the different scenarios.

6.7.2.2 Impact of the flow arrival rate

Since the maximum blocking probability is attained for $\beta = 100\%$ in Fig. 6.3, we focus hereafter on this worst case scenario, by assuming that all flows are intra-network. In this section, we fix $\beta = 100\%$, and study the effect of the average flow arrival rate λ_f on the association policy performance. We show in Fig. 6.4 the blocking probability of the association policies, function of the flow arrival rate λ_f, for different backhaul link capacities C_l.

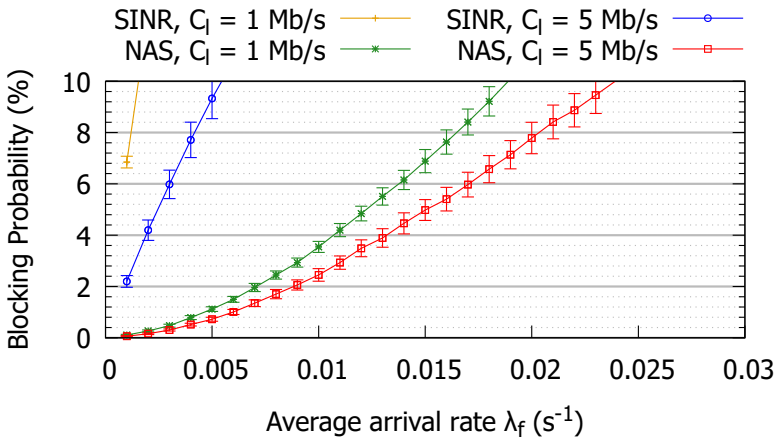


Figure 6.4: Blocking probability function of the average flow arrival rate λ_f, for different inter-BS link capacities C_l.

We set the maximal accepted flow blocking probability to 10%, as any higher value is superfluous and not tolerable in practice. As shown in Fig. 6.4, the gain achieved by the NAS policy with respect to best SINR is significant, for all values of λ_f, and for the different backhaul links capacities. By comparing the values of the arrival rate λ_f beyond which the blocking probability surpasses 10%, we find that, with a best SINR policy, this value is reached for smaller arrival rates. For example, for $C_l = 1$ Mb/s, this arrival rate is smaller than 2×10^{-3} s^{-1}, and for $C_l = 5$ Mb/s, it is as small as 5×10^{-3} s^{-1}. This is highly impractical since it corresponds to a very low traffic. However, with the NAS policy, the 10% blocking is not reached until higher values of λ_f, such as $\lambda_f = 0.02$ s^{-1} for $C_l = 5$ Mb/s, and $\lambda_f = 0.025$ s^{-1} for $C_l = 1$ Mb/s. Indeed, in a best SINR policy, if a UE receives the best SINR from a BS which does not have enough capacity to accept a new flow, then the flow is dropped. In NAS, the UE is given a wider choice in terms of BSs, and the association decision takes into consideration the remaining resources and the bottlenecks in order to avoid early saturation of a BS and/or a backhaul link.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

6.7.2.3 Flow blocking causes

In the following, we set the average arrival rate value to $\lambda_f = 0.01$ s^{-1}, and $\beta = 100\%$. In order to investigate what is causing the flow blocking in each of the association policies, we consider different representative scenarios, that correspond to different backhaul link capacities values C_l, and evaluate the blocking causes. The possible flow blocking causes are:

- **BH = $\emptyset$:** there is no feasible path on the backhaul;
- **UL = $\emptyset$:** there is no candidate BS that has enough resources on the UL;
- **DL = $\emptyset$:** there is no candidate BS that has enough resources on the DL;
- **UL$\cap$DL = $\emptyset$:** there is no candidate BS that has enough resources on both the UL and DL, simultaneously (but UL $\neq \emptyset$ and DL $\neq \emptyset$);
- **UL$\cup$DL = $\emptyset$:** there is no candidate BS that has enough resources on the UL, nor a candidate BS with enough resources on the DL.

Scenario 1: In this scenario, we consider a relatively well-provisioned backhaul, such that all backhaul links have a capacity $C_l = 5$ Mb/s ; $\forall\ l \in \mathcal{L}$.

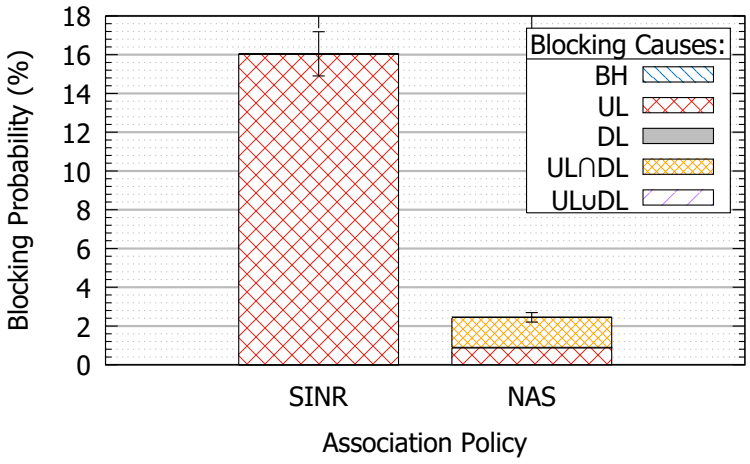


Figure 6.5: Blocking causes for best SINR and NAS, for $C_l = 5$ Mb/s.

Fig. 6.5 shows the blocking probability and causes for the best SINR and the proposed NAS policies. NAS reduces the flow blocking probability by a factor of 6. By investigating the causes that lead to the flow blocking, we notice that, in a best SINR policy, the major blocking cause is the UL. Thus, in this scenario, it is the RAN, and specifically the UL, that creates a bottleneck. This could be a consequence of the assumption we made in the input data that the UL traffic is symmetric to the DL traffic, which is reasonable in many applications, such as video calls and video broadcast [109]. However, even with the asymmetric traffic scenarios, presented in the next Section, observations are the same. Indeed, interference being more aggressive on the UL than on the DL, for the same throughput, more resources are needed on the UL than on the DL, causing the UL to saturate faster. On the other hand, looking at the blocking causes in the NAS policy, we notice that the bottleneck effect is mitigated, and the UL is not a major problem anymore. This shows that NAS, by trying to minimize the remaining UL resources at each association decision, allows a better utilization of available resources.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Scenario 2: We decrease the capacity of the backhaul links, such that $C_l = 1$ Mb/s ; $\forall\ l \in \mathcal{L}$. The network is backhaul-limited, since, practically, a link can only carry one flow at a time.

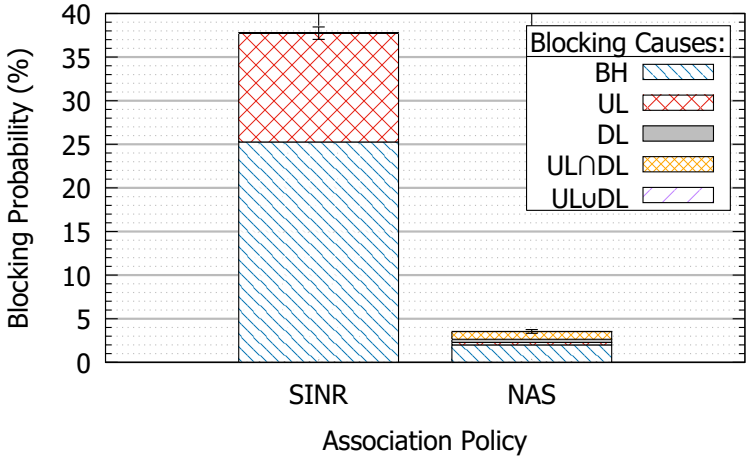


Figure 6.6: Blocking causes for best SINR and NAS, for $C_l = 1$ Mb/s.

Fig. 6.6 shows the blocking probability for the two association policies, as well as the corresponding blocking causes. Results show that the high blocking probability in the best SINR policy is mainly due to the backhaul. By reducing the link capacities, the number of feasible paths between BSs is also reduced, saturating the backhaul and creating a supplementary bottleneck. Nevertheless, the NAS policy overcomes these limitations by taking into account the remaining backhaul resources. Indeed, the flow blocking probability is reduced by a factor of 9. The UL and backhaul bottlenecks present with the best SINR policy are also mitigated. This shows that NAS can adapt and mitigate bottlenecks, whether on the backhaul or on the RAN.

6.7.2.4 Impact of UL/DL traffic symmetry

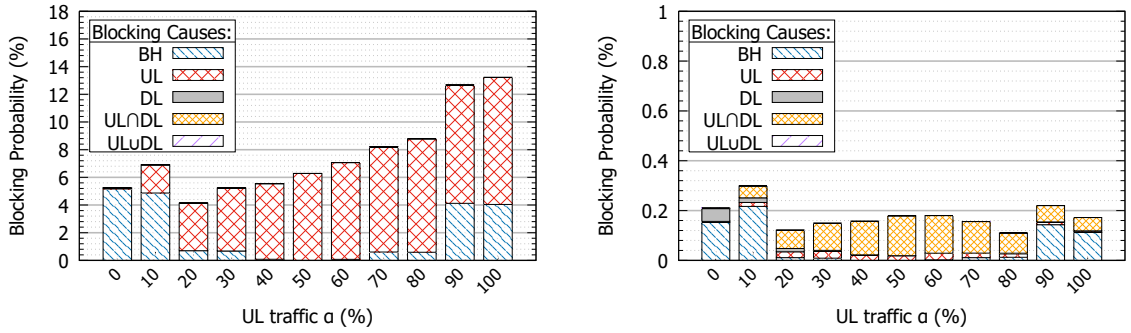
In the following, we evaluate scenarios with asymmetry between the DL and the UL traffic. We consider that all bi-directional flows between u and v are inter-network, i.e., $\beta = 0\%$, such that u is the UE that belongs to the self-deployable network and v is the external party. Moreover, we consider that all bi-directional flows are asymmetrical, such that $\alpha = \frac{d_{u,v}}{d_{u,v}+d_{v,u}}$, with α a percentage in $[0, 100]$. In other words, $d_{u,v}$ is the UL flow, $d_{v,u}$ is the DL flow, and α represents the ratio of UL traffic out of the total traffic in the network.

Fig. 6.7a and Fig. 6.7b show the flow blocking probability and the corresponding blocking causes function of the ratio α, for both best SINR and NAS, respectively. When the majority of the traffic is DL ($\alpha = 0\%, 10\%$) or UL ($\alpha = 90\%, 100\%$), the backhaul is loaded in only one direction, which leads to its saturation, making it the main blocking cause. For the other values of α, the results confirm the previous observations: the blocking is significantly reduced with NAS with respect to best SINR, UL is always the main blocking cause with best SINR, and UL∩DL the main cause in NAS. Furthermore, we notice that the increase in the UL traffic does not impact the blocking probability with NAS as much as it does with best SINR. This proves that NAS uses the resources efficiently and mitigates possible UL bottlenecks, regardless of the UL traffic volume.

6.7.2.5 Impact of the network topology

Up to this point, all the shown results correspond to the topology in Fig. 6.2. In order to validate the NAS policy performance, and demonstrate that the previous results can be generalized, we

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés



(a) SINR (b) NAS

Figure 6.7: Blocking probability and blocking causes function of the UL traffic ratio out of the total total traffic α. *To improve readability, please note that the figures have different scales.*

conducted the same study on a multitude of network topologies. For readability, we only show in Fig. 6.8 a comparison of the flow blocking probability when using best SINR and NAS policies, for 10 of the studied network topologies (denoted t1 to t10). These results correspond to $\lambda_f = 0.01$ s^{-1}, $\beta = 100\%$ and $C_l = 5$ Mb/s. We notice consistent results, with the proposed NAS policy always reducing the flow blocking probability by a factor ranging between 6 and 9.

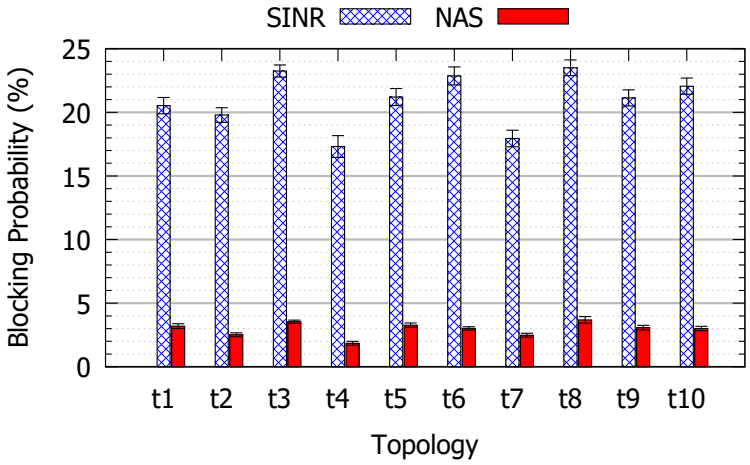


Figure 6.8: Blocking probability for different random topologies.

6.7.2.6 Impact of the backhaul bandwidth

In this section, we study the effect of the backhaul bandwidth on the performance of the NAS policy. For the topology in Fig. 6.2, we set $\beta = 100\%$, and vary the link capacities C_l. We show in Fig. 6.9 the blocking probability function of C_l, for different values of λ_f.

A first observation in Fig. 6.9, corroborating the previous results, is that NAS always outperforms best SINR, regardless of the links capacities, and λ_f. Indeed, as previously stated, the blocking probability with best SINR surpasses 10% for small values of λ_f, which explains why only $\lambda_f = 0.005$ s^{-1} appears in the figure for the best SINR case. Even then, the blocking probability is still higher than that with NAS at $\lambda_f = 0.02$ s^{-1}. Hence, with the gain with respect to the best SINR policy already proved and established, for all values of C_l, we solely focus in the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

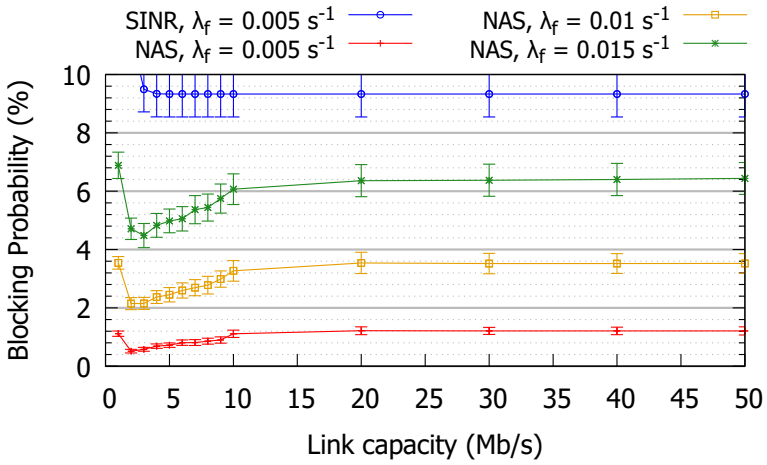


Figure 6.9: Blocking probability function of the inter-BS link capacities C_l, for different average arrival rates λ_f.

following discussion on the performance of the NAS policy with respect to the backhaul state.

Intuitively, the blocking probability should decrease with the increase of C_l, at least when the backhaul is causing a bottleneck. At a certain threshold, when C_l values are sufficiently high such that the backhaul stops being a potential blocking cause, one would expect that the blocking probability would become stable (for the same RAN conditions), i.e, independent of the value of C_l. Indeed, this is the case with best SINR, as seen in Fig. 6.9, with C_l ceasing to affect the blocking probability from $C_l = 6$ Mb/s onward. However, we notice in Fig. 6.9 that is not the case with NAS. In fact, the blocking probability curve with NAS can be split into three distinct behaviors: **(i)** a decrease in the blocking probability for $C_l \in [1, 3]$ Mb/s, **(ii)** an increase in the blocking probability for $C_l \in (3, 10]$ Mb/s, and **(iii)** a relatively stable blocking probability for $C_l > 10$ Mb/s. The same applies for all values of λ_f.

The behavior for $C_l \leq 3$ Mb/s is expected: when the backhaul is limited, the blocking probability decreases with the increase of C_l. This happens because the higher the value of C_l, the lower the probability of having insufficient resources on the backhaul causing blocking. However, the decreasing blocking probability reaches a minimum value at a certain threshold C_l. Beyond this point, we notice that the blocking probability gradually increases, before eventually stabilizing. It seems that, when the backhaul is not limited, accounting for the backhaul when deciding on the association can, counter-intuitively, slightly degrade the performance of NAS. Why does C_l have this impact?

Recall that the association decision in NAS is based on maximizing a product of several metrics (Eq. 6.12), representing the normalized remaining resources on the backhaul, the DL and the UL, as explained in details in Section 6.5. The metrics $L^{BH}_{P(i,j)}$ (Eq. 6.29), $L^{BH}_{P(j,i)}$ (Eq. 6.29), $L^{DL}_{i,j}$ (Eq. 6.9) and $L^{UL}_{i,j}$ (Eq. 6.10), are computed for all feasible pairs $(i, j) \in \mathbb{F}$ and their corresponding feasible paths in $\mathcal{P}_{i,j}$ and $\mathcal{P}_{j,i}$. We are particularly interested in the backhaul metrics, $L^{BH}_{P(i,j)}$ and $L^{BH}_{P(j,i)}$. We remind that for each path $P(x, y)$, $L^{BH}_{P(x,y)}$ is written as:

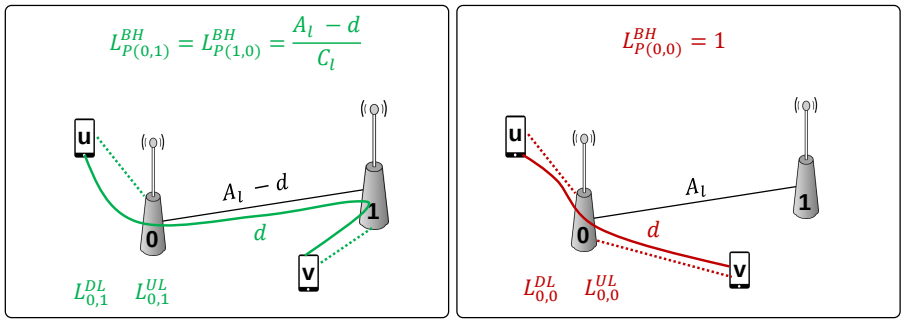
$$L^{BH}_{P(x,y)} = \begin{cases} \min\limits_{l \in P(x,y)} \left(\frac{A_l - d}{C_l} \right), \text{ if } x \neq y \\ 1, \text{ if } x = y \end{cases} \tag{6.29}$$

Let us suppose we are deciding on the association of two UEs, and their routing paths, among

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

a set of pairs (i, j). For the same RAN conditions, $L^{DL}_{i,j}$ and $L^{UL}_{i,j}$ in Eq. 6.12 are not affected by C_l whatsoever. However, $L^{BH}_{P(i,j)}$ and $L^{BH}_{P(j,i)}$ are affected. Let us consider that, at the time of the decision, the backhaul is not sufficiently loaded to cause flow blocking. That is, $A_l \approxeq C_l,\ \forall\ l \in P(i,j),\ \forall\ l \in P(j,i)$. Then, the ratio in Eq. 6.29 becomes $\frac{A_l - d}{C_l} \approxeq 1 - \frac{d}{C_l}$. Recall that we should have $d \leq C_l$ for the path (to which l belongs) to be feasible. Hence, it is the ratio $\frac{d}{C_l}$ that nuances the impact of L^{BH} on the association decision. Three case unfold.

(a) The smaller the difference between d and C_l, the larger the ratio between them, and the smaller is L^{BH}. A small L^{BH} can demote a link (the path it belongs to), and eventually a BS pair, even if, from a RAN point of view, the pair has better conditions.



Case	C_l	A_l	d	$L^{BH}_{P(0,1)}$	$L^{BH}_{P(1,0)}$	$L^{BH}_{P(0,0)}$	$L^{DL}_{0,1} L^{UL}_{0,1} L^{BH}_{P(0,1)} L^{BH}_{P(1,0)}$		$L^{DL}_{0,0} L^{UL}_{0,0} L^{BH}_{P(0,0)}$	BS pair
(a)	3	3	1	0.67	0.67	1	0.45 $L^{DL}_{0,1}$ $L^{UL}_{0,1}$	<	$L^{DL}_{0,0} L^{UL}_{0,0}$	(0, 0)
(b)	5	5	1	0.8	0.8	1	0.64 $L^{DL}_{0,1}$ $L^{UL}_{0,1}$	>	$L^{DL}_{0,0} L^{UL}_{0,0}$	(0, 1)

Figure 6.10: An example with two UEs, u and v, having the possibility of associating to one of two BS pairs, $(0, 0)$ or $(0, 1)$.

To clarify, let us consider a simple example where two UEs have the possibility to associate to two BS pairs, $(0, 0)$ or $(0, 1)$, as shown in Fig. 6.10. On the RAN, the pair $(0, 1)$ is better, such that $L^{DL}_{0,1} \cdot L^{UL}_{0,1} > L^{DL}_{0,0} \cdot L^{UL}_{0,0}$. For case (a) in Fig. 6.10, on the backhaul, with $C_l = A_l = 3$ Mb/s, $\forall\ l \in P(0,1),\ \forall\ l \in P(1,0)$, and $d = 1$ Mb/s, we have $L^{BH}_{P(0,1)} = L^{BH}_{P(1,0)} = 0.67$ and $L^{BH}_{P(0,0)} = 1$. By comparing the products corresponding to each pair, the small values of $L^{BH}_{P(0,1)}$ and $L^{BH}_{P(1,0)}$ do have an impact, such that $L^{DL}_{0,1} \cdot L^{UL}_{0,1} \cdot L^{BH}_{P(0,1)} \cdot L^{BH}_{P(1,0)} < L^{DL}_{0,0} \cdot L^{UL}_{0,0}$. Thus, the pair $(0, 0)$ would be selected, despite the fact that, from a RAN point of view, $(0, 1)$ is better.

(b) When the difference between d and C_l increases, the ratio between them decreases, and L^{BH} increases. With higher L^{BH}, the RAN metrics L^{DL} and L^{UL} have more impact on the decision. In this case, the pair with better conditions from a RAN point of view is more likely to get picked.

Let us get back to our previous example in Fig. 6.10, where we compare BS pairs $(0, 0)$ and $(0, 1)$. Recall that the pair $(0, 1)$ is better from a RAN point view. For case (b) in Fig. 6.10, on the backhaul, with $C_l = A_l = 5$ Mb/s, $\forall\ l \in P(0,1),\ \forall\ l \in P(1,0)$, and $d = 1$ Mb/s, we have $L^{BH}_{P(0,1)} = L^{BH}_{P(1,0)} = 0.8$ and $L^{BH}_{P(0,0)} = 1$. The values of $L^{BH}_{P(0,1)}$ and $L^{BH}_{P(1,0)}$ are sufficiently high in this case to obtain the following: $L^{DL}_{0,1} \cdot L^{UL}_{0,1} \cdot L^{BH}_{P(0,1)} \cdot L^{BH}_{P(1,0)} > L^{DL}_{0,0} \cdot L^{UL}_{0,0}$. Thus, the pair $(0, 1)$ would be selected.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

The only difference between the aforementioned examples in (a) and (b) is the value of C_l. In both cases the backhaul is not a bottleneck. Nevertheless, a different BS pair is selected for the association. Therefore, C_l does have an impact on the decision, even if the backhaul does not particularly pose any threat, and the RAN conditions are the same. The consequence is that, in some cases, the association decisions are more RAN-oriented. BSs with better RAN conditions are prioritized. This creates an impact on the user distribution over BSs, and on the load balance. More UEs associate to the RAN favored BSs, which risks increasing the probability of saturating the UL (or increasing UL interference). Eventually, this results in the slight increase in the blocking probability, seen in Fig. 6.9.

(c) Finally, when d becomes much smaller than C_l, the ratio $\frac{d}{C_l}$ tends towards 0, and L^{BH} tends towards 1. In this case, no matter the value of C_l, $L^{BH}_{P(i,j)}$ and $L^{BH}_{P(j,i)}$ are high enough to have no impact on the result when multiplied with $L^{DL}_{i,j} \cdot L^{UL}_{i,j}$ in Eq. 6.12. This is what we observe on the curves in Fig. 6.9: for the highest values of C_l the blocking probability stabilizes. The association in this case is only RAN-oriented.

Note that the numerical values of the C_l thresholds, in this example, and all the other numerical results, depend on the considered numerical values of d and the other input parameters defining the UL and DL models in Section 6.7.1. A different set of input data might lead to different numerical values, but with no impact on the previous qualitative conclusions.

6.8 Revisiting assumptions on user association

In this section, we assess the impact of two assumptions we previously made in our model. More precisely, we relax the assumption that an associated UE with ongoing flows cannot change its association, by allowing the re-association of already associated UEs. Furthermore, we evaluate a split UL/DL association, where one UE is allowed to associate to two different BSs simultaneously, one for the UL and one for the DL. We keep the same simulation settings presented in Section. 6.7.1, with the topology in Fig. 6.2, $\beta = 100\%$, and $C_l = 5$ Mb/s.

6.8.1 Re-association

One of the assumptions stated in Section 6.2 is that an already associated UE does not re-associate upon the arrival of a new flow, but stays with the same BS as long as it has an ongoing flow. We made this assumption for practical reasons, since a UE re-association would result in a handover, with an important impact on the existing flow(s), and to simplify the problem presentation. In the following, we relax this assumption, and propose a version of the NAS policy, denoted NAS-Re, in which re-association of the UEs involved in a new flow is allowed. When re-associated to a new BS, all the ongoing flows of the UE are moved to this new BS (Fig. 6.11).

A UE u that is part of a new flow request and already has ongoing flows, is treated the same way as a UE with no ongoing flows. Consequently, the same flow admission procedure described in Section 6.4, and the association decision procedure described in Section 6.5, apply to this UE as well. However, there are differences in the computation of the number of channels needed by u on the DL and on the UL from each BS j. That is, in the computation of $m^{DL}_{u,j}$ and $m^{UL}_{u,j}$ used in the flow admission procedure in Section 6.4.1, and defined in Eq. 6.6. If UE u is re-associated, all of its flows are re-located to the new BS. Hence, all of these flows must be taken into account when computing the consumption on the RAN. Instead of computing the number of needed channels to grant u only the throughput it asks for in the new flow request, the number of needed channels is

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

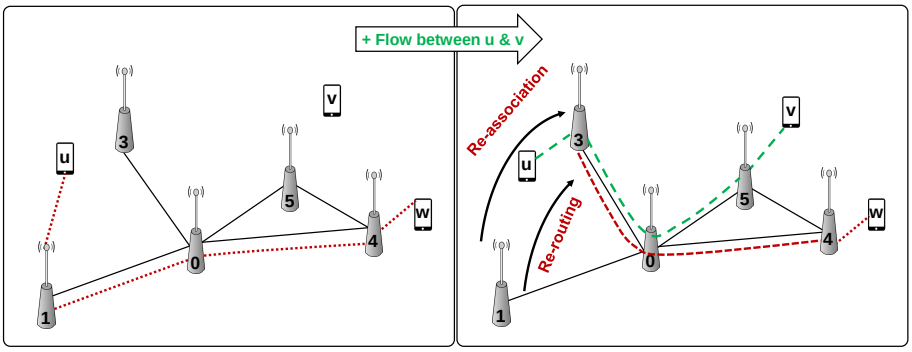
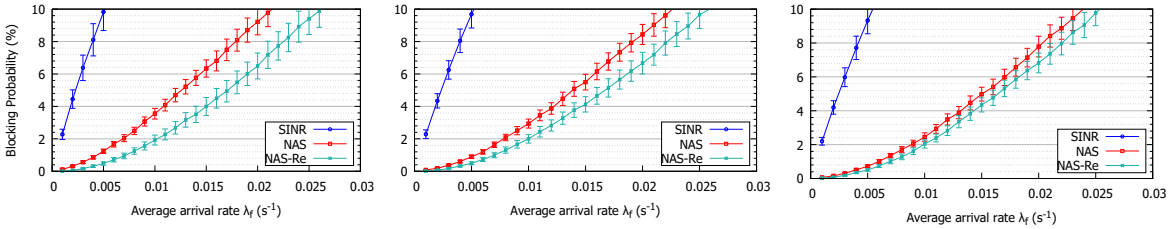


Figure 6.11: Re-association of a UE with an ongoing flow and a new flow request (NAS-Re).

computed for all the throughputs of all the ongoing flows of UE u. If we denote by $\mathcal{F}$ the set of all ongoing flows $f(x, y)$ with $x, y \in \mathcal{U}$, then $m^{DL}_{u,j}$ and $m^{UL}_{u,j}$ are computed in the following way:

$$m^{DL}_{u,j} = \frac{\sum_{f(v,u)\in\mathcal{F}} d_{v,u}}{R^{DL}_{u,j}} \qquad (6.30) \qquad\qquad m^{UL}_{u,j} = \frac{\sum_{f(u,v)\in\mathcal{F}} d_{u,v}}{R^{UL}_{u,j}} \qquad (6.31)$$

Furthermore, the re-association entails a re-routing of the ongoing flows. Therefore, in addition to checking the feasible paths $\mathcal{P}_{i,j}$ and $\mathcal{P}_{j,i}$ for each pair $(i, j) \in \mathbb{F}$ candidate for the association of the new flow, the routing paths of the ongoing flows of u must be also found in case of a re-association. That is, we should also find the set of feasible paths $\mathcal{P}_{i,k}$ and $\mathcal{P}_{k,i}$, where i is the BS to which u can re-associate, and k the BS(s) to which the UE(s) of the ongoing flow(s) with u are already associated. All of these paths should also be included in the computation of the metric representing the remaining resources on the backhaul in Eq. 6.11 to reflect the cost of the re-association on the backhaul.



(a) 25 UEs (b) 50 UEs (c) 100 UEs

Figure 6.12: Blocking probability function of the average flow arrival rate λ_f, when re-association is allowed (NAS-Re), for different number of UEs in the network.

We compare the performance of the NAS-Re policy with the original NAS, with no re-association by showing, in Fig. 6.12, their respective blocking probabilities as a function of the average arrival rate λ_f, for different number of UEs in the network. We notice that, for all values of λ_f, NAS-Re is better than NAS, but the gain in blocking probability is not very significant. We decrease the number of UEs in the network while keeping the same flow arrival rate. This would increase the probability that a UE gets a new flow request while it already has ongoing flows. In this case, re-association is more beneficial. Indeed, we remark that the gain by NAS-Re with respect to NAS

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

increases with the decrease of the number of UEs. The impact of the signaling traffic caused by frequent re-associations should be further investigated.

6.8.2 Split DL/UL association

It has long been considered that a UE is associated to one and only one BS that handles all of its traffic, on the DL and the UL. Nevertheless, split DL/UL association schemes, in which a UE can be associated to different BSs on the DL and the UL, are currently being studied for 5G networks [99, 109]. While discussions on the feasibility of such a split are out of the scope of this chapter, we are interested in its potential benefits.

Therefore, we relax the assumption that a UE is associated to one BS at a time, and implement the NAS policy with split DL/UL association, denoted NAS-Split. In NAS-Split, as shown in Fig. 6.13, for each bi-directional flow request between UEs u and v, of data rate requirements $d_{u,v}$ and $d_{v,u}$, the goal is to find two pairs of BSs, one for each UE, i.e., (i_{DL}, i_{UL}) for UE u, and (j_{DL}, j_{UL}) for UE v, and the corresponding routing paths between them, i.e., $P(i_{UL}, j_{DL})$, and $P(j_{UL}, i_{DL})$.

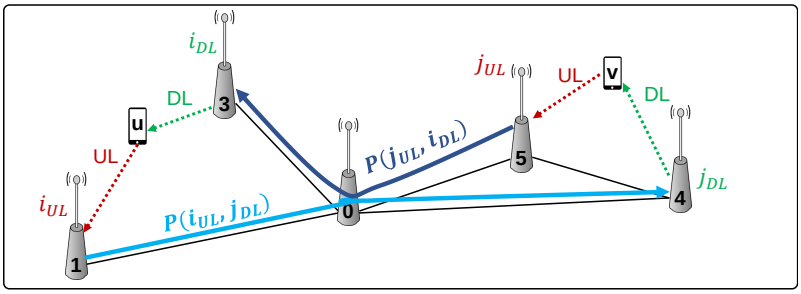


Figure 6.13: Example of split DL/UL association (NAS-Split).

The same flow admission procedure as detailed in Section 6.4 applies in this case. The difference is that a BS does not have to be considered candidate for both DL and UL at the same time to be considered as feasible, as previously stated. That is, on the RAN, there are two sets of feasible candidate BSs pairs, one for the DL, $\mathbb{F}^{DL}$, and for the UL, $\mathbb{F}^{UL}$. The set $\mathbb{F}^{DL}$ consists of the BS pairs feasible on the DL (one BS for each UE), and $\mathbb{F}^{UL}$ consists of the BS pairs feasible on the UL, such that $(i_{DL}, j_{DL}) \in \mathbb{F}^{DL}$ and $(i_{UL}, j_{UL}) \in \mathbb{F}^{UL}$. There should be at least one pair in $\mathbb{F}^{UL}$ and at least one pair in $\mathbb{F}^{DL}$ for the flow to be accepted. The sets $\mathbb{F}^{UL}$ and $\mathbb{F}^{DL}$ are determined according to the same rules detailed in Section 6.4.1. Similarly, the set of feasible routing paths on the backhaul, $\mathcal{P}_{i_{UL},j_{DL}}$ and $\mathcal{P}_{j_{UL},i_{DL}}$, are also determined according to the same rules detailed in Section 6.4.2.

Based on the same Equations detailed in Section 6.5, $L^{BH}_{P(i_{UL},j_{DL})}$, $L^{BH}_{P(j_{UL},i_{DL})}$, $L^{DL}_{i_{DL},j_{DL}}$, and $L^{UL}_{i_{UL},j_{UL}}$ are computed for all feasible pairs in $\mathbb{F}^{DL}$ and $\mathbb{F}^{UL}$, and their corresponding feasible paths in $\mathcal{P}_{i_{UL},j_{DL}}$ and $\mathcal{P}_{j_{UL},i_{DL}}$. Therefore, the association decision is based on a modified version of the metric explained in Eq. 6.12, such that the selected BSs for association (j^{DL}_u, j^{UL}_u, j^{DL}_v, and j^{UL}_v), and paths for routing $\left(P(j^{UL}_u, j^{DL}_v), P(j^{UL}_v, j^{DL}_u)\right)$ are the ones that maximize the following product:

$$\arg \max_{\substack{P(i_{UL},j_{DL})\in\mathcal{P}_{i_{UL},j_{DL}} \\ P(j_{UL},i_{DL})\in\mathcal{P}_{j_{UL},i_{DL}} \\ (i_{DL},j_{DL})\in\mathbb{F}^{DL} \\ (i_{UL},j_{UL})\in\mathbb{F}^{UL}}} L^{BH}_{P(i_{UL},j_{DL})} \cdot L^{BH}_{P(j_{UL},i_{DL})} \cdot L^{DL}_{i_{DL},j_{DL}} \cdot L^{UL}_{i_{UL},j_{UL}} \tag{6.32}$$

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

We compare the performance of NAS-Split with the previous policies in Fig. 6.14. We notice that the NAS-Split significantly outperforms the original NAS with joint DL/UL association, and further reduces the blocking probability, notably for higher values of the average arrival rate λ_f. In fact, we notice that, with the increase of λ_f, i.e., the increase of traffic in the network, the gain in blocking probability with NAS-Split increases significantly. The blocking probability remains lower than 5% for NAS-Split, whereas it surpasses 10% for the same values of λ_f with NAS.

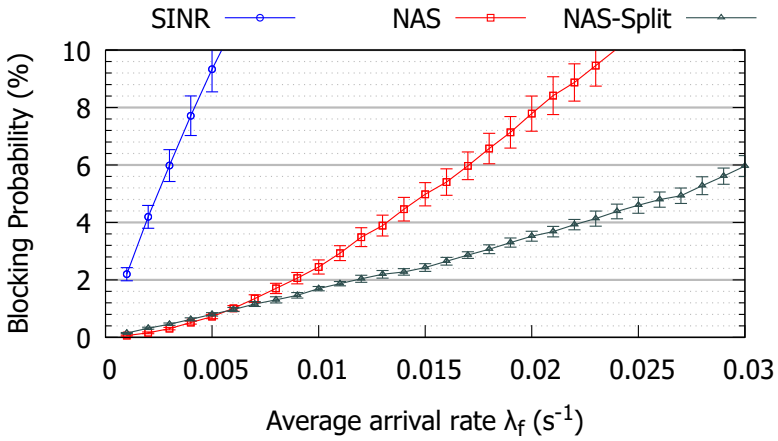


Figure 6.14: Blocking probability function of the average flow arrival rate λ_f, when split DL/UL association is allowed (NAS-Split).

To further investigate these results, we zoom in on the blocking causes in both NAS and NAS-Split policies, for $\lambda_f = 0.02$ s^{-1}, in Fig. 6.15. We notice that the main blocking cause in the NAS policy is in fact UL∩DL, meaning that there is no BS with sufficient resources on both UL and DL to accept the flow. In other words, there might exist a BS i that has enough DL resources to accommodate the flow of a UE u but not enough UL resources, and another BS $j \neq i$ with enough UL resources for u, but not enough DL. Consequently, when split association is allowed, u can associate to BS i on the DL and BS j on the UL. In this case, the flow would not be blocked. This is why, with NAS-Split, we notice that the UL∩DL blocking cause disappears, causing the blocking probability to drop significantly. Hence, allowing multiple associations to different BSs for one UE seems to be a promising technique.

6.9 SFR-based power map on the downlink

In Chapter 3, we proposed an SFR-based frequency reuse scheme on the DL, in which the frequency band is divided into sets of sub-channels, referred to as sub-bands. In such a scheme, all BSs transmit on all sub-bands, with a different transmission power on each sub-band. We proposed an offline algorithm that determines the BSs power allocation on each sub-band (power map), given the network topology. The power map is computed prior to the network operation, and then adopted by the BSs. Extensive simulations on the DL showed that the proposed scheme allows users to have higher throughput in comparison with a conventional reuse scheme with reuse factor r, where all BSs have equal power on all channels (EP-R$_r$).

Thus far, we evaluated the proposed NAS policy under the assumption of a conventional EP-R3 scheme. This is because NAS takes into account both the DL and the UL, whereas the SFR-based power map only applies on the DL. Different reuse schemes on the DL and the UL risk being

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

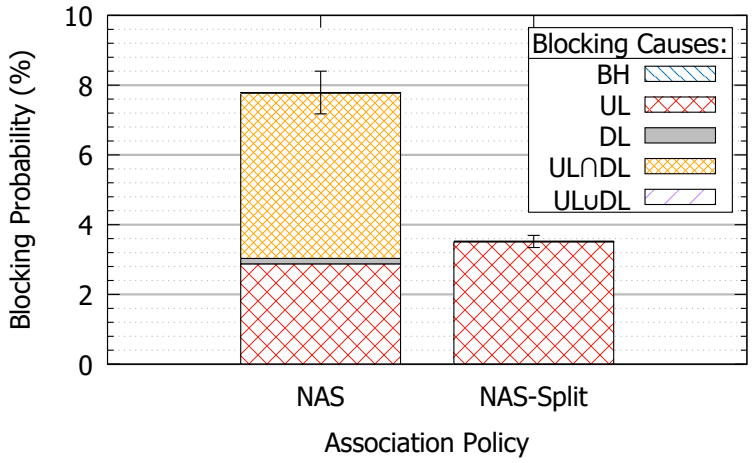


Figure 6.15: Blocking causes for NAS with joint DL/UL association, and NAS-Split with split DL/UL association.

incompatible, which could negatively impact the overall outcome. We evaluate, in this section, the performance of the NAS policy if the BSs frequency and power allocation on the DL were to follow a pre-computed SFR-based power map, as per the algorithm presented in Chapter 3.

6.9.1 Frequency and power allocation model for the downlink

Let $\mathcal{S}$ be the set of sub-bands. Each sub-band has $k = M/|\mathcal{S}|$ sub-channels. We denote by P_j^s the power of a BS j on a sub-band s. P_j^s is the same on all the k sub-channels of sub-band s. The set of all P_j^s values, indicating with which power each BS transmits on each sub-band, constitute what we refer to as power map. To get a throughput d when associated to a BS j, a UE must be allocated a total number of channels, denoted $m_{u,j}$. However, those channels can belong to different sub-bands. If we denote by $m_{u,j}^s$ the number of channels allocated to a UE u, from BS j, on a sub-band s, then the total number of channels allocated to u, from BS j, is:

$$m_{u,j} = \sum_{s \in \mathcal{S}} m_{u,j}^s \tag{6.33}$$

On the other hand, the number of needed channels on a BS depends on the rate received from that BS. We denote by $R_{u,j}^s$ the rate UE u gets from BS j on a sub-band s. With a different transmission power $P_{u,j}^s$ on each sub-band, the corresponding SINR $\Gamma_{u,j}^s$ can be different, and, consequently, $R_{u,j}^s$ can also be different on each sub-band s. Hence, the throughput d is obtained by allocating a number of channels $m_{u,j}^s$ on each sub-band s, depending on the rate $R_{u,j}^s$ that can be perceived on that sub-band, such that:

$$d = \sum_{s \in \mathcal{S}} m_{u,j}^s R_{u,j}^s \tag{6.34}$$

The raised question here is: how are the allocated channels to u on BS j partitioned over the sub-bands? In other words, how are the values of $m_{u,j}^s \ \forall s \in \mathcal{S}$ determined. In an effort to test the scheme performance under optimized settings, we propose an optimization problem that outputs $m_{u,j}^s$ with the objective of minimizing the total number of allocated channels to all UEs on all the sub-bands on a BS j (Eq. 6.35).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

$$\min \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}} m_{u,j}^{s} \tag{6.35}$$

$$\sum_{s \in \mathcal{S}} m_{u,j}^{s} R_{u,j}^{s} = d, \ \forall u \in \mathcal{U} \tag{6.36}$$

$$\sum_{u \in \mathcal{U}} m_{u}^{s} \leq k, \ \forall s \in \mathcal{S} \tag{6.37}$$

The constraint 6.36 ensures that each UE is granted the required throughput. The constraint 6.37 states that the sum of channels allocated to all UEs on a sub-band does not exceed the number of channels in that sub-band.

This channel allocation within the sub-bands of a BS depends on the total number of UEs associated to that BS and their required throughputs. Hence, it must be re-computed at each change in the number of associated UEs, i.e., each UE arrival and departure from the BS. While, in real-life deployments, such an optimal allocation may not be easily implemented (due to computation time), our main objective here is to evaluate the power map performance, under a given assumption on channel allocation. We do not tackle the practical implementation of this scheme or any other.

The same flow admission procedure, as described in Section 6.4, followed by the same association decision process, as described in Section 6.5, applies here. At each flow request between u and v, the number of channels needed by each of them from each BS, i.e., $m_{u,j}^{DL}$, $m_{u,j}^{UL}$ and $m_{v,j}^{DL}$, and $m_{v,j}^{UL}$ are computed, and compared to the number of remaining channels on this BS j, i.e., M_j^{DL} and M_j^{UL}. The difference, in this case, lies in how $m_{u,j}^{DL}$ and $m_{v,j}^{DL}$ are computed. Indeed, as explained in Eq. 6.33, the number of channels on each sub-band must be first determined. This can be done by solving the channel allocation optimization problem (Eq. 6.35-Eq. 6.37).

6.9.2 Numerical results

In the following, we compare the association policies performance under two different frequency and power allocation schemes on the DL: EP-R3 and SFR-based power map (PM). In both cases, we maintain the EP-R3 scheme on the UL. The rest of the simulation settings are the same as those described in Section 6.7.1.

For the topology in Fig. 6.2, we compute the power map on the DL, according to the algorithm presented in Chapter 3. We obtain the allocation scheme shown in Fig. 6.16. This power map is fixed and followed by the BSs during the network operation.

Fig. 6.17 shows a comparison of the blocking probability between the best SINR policy and NAS, for the two schemes. The shown results correspond to $\lambda_f = 0.01$ s^{-1} and $C_l = 5$ Mb/s.

The blocking probability with the PM scheme on the DL is higher in both best SINR and NAS. However, we already demonstrated in Chapter 3 the benefits of the PM scheme in comparison with the conventional EP-R3. We showed that it significantly increased the users' throughputs on the DL. Hence, one would expect that with the better DL conditions with PM, the blocking probability would decrease. So why does that scheme actually under-perform in this case?

To start with, even with EP-R3, the DL was not a bottleneck, nor a major blocking cause. The impact seen on the blocking probability is mainly due to the UL. More precisely, on the incompatibility between the frequency allocation schemes on the DL and the UL. When both interference schemes are the same on the UL and the DL, both based on EP-R3 for example, the disadvantage of a BS is mutual on the DL and on the UL. For a particular BS, the interfering BSs (i.e., those sharing the same channels) on the DL cause limitations to the DL SINR obtained from

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

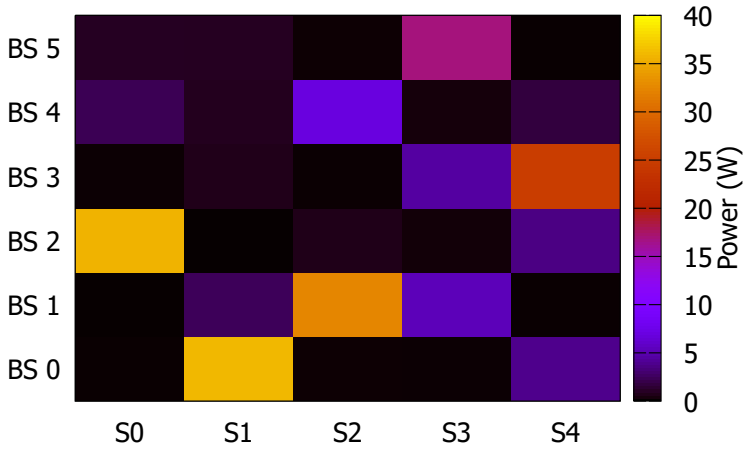


Figure 6.16: Computed power map for the studied topology.

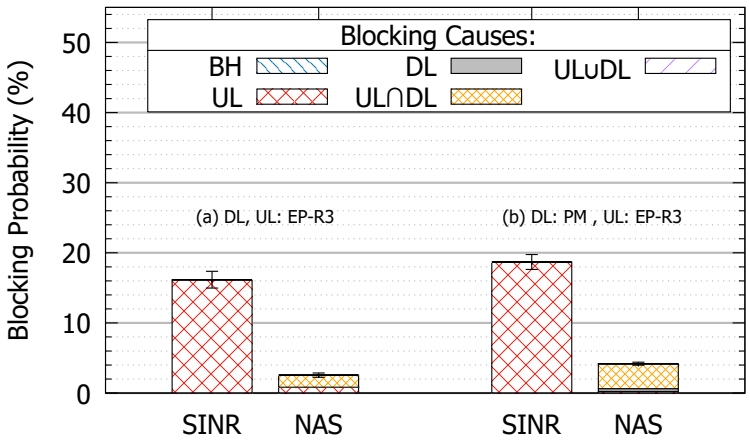


Figure 6.17: Blocking probability for best SINR and NAS, under two allocation schemes: (a) EP-R3 on DL and UL, (b) PM on DL and EP-R3 on UL.

that BS. For the same BS, on the UL, the UEs associated to those same interfering BSs are the potential interferers, causing limitations to its UL SINR. This is not the case, however, when reuse schemes are different on the DL and the UL.

With PM on the DL, all BSs are using the same DL channels, with a carefully computed power allocation scheme on different sub-bands. The interference on the DL is per sub-band, and caused by all the BSs, but with different degrees, depending on their transmission power on that sub-band. For example, for a BS j, another BS i can be heavily interfering on one (or multiple) sub-band(s), but has a negligible interference on the remaining sub-bands. Whereas, on the UL, BS i and BS j might or might not be sharing the same channels. This difference in the "conflicting" BSs can have impacts on how UEs are associated. If, for example, a BS presents strong advantages on the DL, due to how power is allocated on its sub-bands, an increase in the number of UEs associated to it can be expected. However, more UEs associating to that BS leads to an increase in the UL interference caused by its UEs, negatively impacting the BSs sharing the same UL channels, and limiting the UL rate they can offer to UEs. This increases the UL outage on those BSs.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

These observations open a much-needed and interesting discussion on the UL in cellular networks, more specifically on the UL frequency allocation and scheduling problems, and their correlation with the DL problems. While the problem of resource allocation on the UL has been studied [61], it is usually decoupled from the DL problem. Nevertheless, the two problems seem to be intricate. From a simple observation in our study, we can deduce that a certain reuse scheme on the DL can impact how UEs are associated, which itself impacts the UL interference, and all the resulting consequences. Indeed, a "smarter" UL interference scheme, correlated to that on the DL, could have prevented the resulting conflicts we observe. Likewise, a "smarter" user scheduling algorithm on the UL, that takes the system state and interference into account, is also an option.

6.10 Conclusion

With self-deployable networks having a specific backhaul architecture, with potentially limited bandwidth, on one hand, and users having increased throughput requirements, on the other hand, the user association problem deserves to be revisited. In this chapter, we proposed NAS, a novel network-aware user association policy for self-deployable networks. It takes into account the resources of both the DL and UL of the access network, the resources on the backhaul, as well as the requested flows throughputs. With guaranteed throughput requirements, we adopted flow blocking probability as the metric of interest when associating users. Our results showed that this network-aware strategy largely outperforms a classical user association policy based solely on access layer information. We evaluated the policy performance under different network settings, by varying key parameters, such as inter-network and intra-network flows, the backhaul bandwidth, and the network load. The obtained results confirmed that NAS performs well in varying conditions, and is capable of mitigating both RAN and backhaul bottlenecks. Furthermore, we investigated two variants of the proposed policy. We showed that allowing re-association of already associated users with ongoing flows can slightly improve performance, notably when users have multiple simultaneous flows. On the other hand, we showed that splitting the UL and DL association is an advantageous paradigm, that is worthy of future study.

The association policy, as presented, takes a decision in real-time, based on the instantaneous value of the decision metric, which accounts for the actual network state. Whether re-association is allowed or not, the decision only concerns the UEs involved in the arriving flow. However, since each new association has the potential to modify the network state, the consecutive association decisions are dependent. A decision at time t_x affects the upcoming decision at t_{x+1}. Hence, a limitation of this scheme is that each decision is solely based on the present network state, with complete disregard for the upcoming events. An advantageous decision concerning one or two UEs at instant t_x does not necessarily lead to an optimal outcome on the long run. Indeed, the optimal performance can only be achieved if, at each arrival, the association of all UEs in the network is re-evaluated. Such a scheme is, however, costly in practice, due to the significant amount of signaling overhead it generates to manage all the handovers, in addition to its computational cost. Another possible enhancement would be to include some kind of prediction of future events, or learning opportunities, in the association decision metric. In other words, make the instantaneous association decision smarter, and future-proof.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Chapter 7

Conclusion and Perspectives

7.1 Summary

In parallel to the rapid evolution of cellular networks, grew the need to provide coverage and reliable data services anywhere, anytime. Nevertheless, in a variety of use cases, a classical network fails, is not suitable, or simply does not exist. This happens, for example, when the network saturates during crowded events, when first responders need private broadband communication in mission-critical situations, or due to the lack of infrastructure in areas with low population density. Self-deployable mobile networks are the envisioned solution to provide network services in such cases. They are rapidly deployable, easily transportable, autonomous, and small-sized mobile networks, that can be deployed and operated on demand. In this context, innovative deployment platforms on land or in the air are developed.

Self-deployable networks are entering uncharted territory in cellular networks. In this thesis, we focused on several challenges, building on what distinguishes them from classical networks. Indeed, state of the art cellular architecture is pre-planned, with a strict physical separation between the RAN and the CN, and an over-provisioned backhaul network. In contrast, the self-deployable architecture challenges those inherent properties, with a rapid deployment short-cutting the planning and provisioning phase, irregular deployment topologies, core network functions co-located with BSs, and a limited backhaul network. These differences push for the re-evaluation of classical radio access network problems, such as resource allocation and user association, to adapt them to the self-deployable networks attributes. Furthermore, they create novel challenges related to the virtualization of the core network functions and their placement in the network.

Starting with the RAN, we first revisited the frequency and power allocation problem in an irregular network topology. While efficient power management in schemes such as SFR have been proven to achieve frequency efficiency and reduce interference, their properties (i.e., three sub-bands, two power levels) are usually fixed in accordance with the hexagonal grid topology [67]. We evaluated a more flexible SFR scheme that allows more sub-bands and power levels. We proposed an offline algorithm that takes as input the network topology, and returns a power map which determines the transmission power of each BS on each sub-band. The power map is computed based solely on the network topology, prior to the network operation, when no accurate information on the user density or distribution is available. Then, it is followed by the BSs throughout their operation. We showed that the gain achieved in user throughput when BSs adopt the power map is significant in comparison with a conventional frequency reuse scheme. Furthermore, we highlighted the robustness of the statically computed power map by evaluating its performance facing varying number of concurrent users, user distribution, and association policies. The computation algorithm consists of solving a non-convex, non-linear, complex optimization problem,

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

through transformations based on signomial programming.

Then, we tackled the issue of core network configuration. The latter is co-located with the BSs, and referred to as Local CN. In a self-deployable network, the backhaul consists of the inter-BS links interconnecting the BSs, where all data and signaling traffic exchanged between the BSs and the Local CN are routed. An efficient placement of the Local CN functions is key when the backhaul is limited, in order to avoid backhaul saturation. We focused on the Local CN functions placement and user attachment to the Local CN under the assumption of a limited backhaul bandwidth. For the centralized placement, we proposed a novel metric called flow centrality, defined for a node as the maximum traffic that can be sent simultaneously by all the other nodes in the network towards this node, under certain capacity and load distribution constraints. By benchmarking the proposed metric against state of the art centrality metrics, we highlighted the potential loss in the total traffic received by the Local CN, when the latter is not placed on the BS with the maximum flow centrality in the network. The flow centrality metric computation is based on a linear optimization problem. Furthermore, we deduced analytical expressions for direct computation of the metric in some particular network topologies.

On the other hand, we evaluated the backhaul bandwidth consumption when the Local CN functions are distributed in comparison to when they are centralized. We approached the problem by explicitly taking into account the number of users in the network, their distribution, and their requests, and by considering both data and signaling traffic. We demonstrated that distributing S-GWs (which ensure the routing function), such that each BS is co-located with one S-GW, consumes less backhaul bandwidth than a centralized placement. The Local CN functions placement problem is formulated based on mixed integer linear programming. It determines the optimal placement of the Local CN when the latter is centralized, and the number of instances and their placement when it is distributed, with the objective of minimizing the backhaul bandwidth consumption caused by data and signaling traffic exchanged between the BSs and the Local CN.

Then, we tackled the attachment problem. As opposed to all users attaching by default to the centralized S-GW, we verified that an attachment policy where each user attaches to the S-GW of their BS performs better. Following a comparison between this simplified attachment scheme and more sophisticated optimal attachments, we concluded that optimizing attachment only achieves noteworthy gain when the signaling traffic is significant. The attachment problems are formulated based on mixed integer quadratic programming.

Finally, with both the RAN and the Local CN configured, and the network ready to operate, users start requesting services. Therefore, we tackled the user association problem, by taking into account critical attributes of self-deployable networks. We proposed a novel network-aware association policy that goes beyond state of the art policies, by jointly including key metrics, such as backhaul bandwidth, UL resources, and user demands. By evaluating the flow blocking probability, we showed that the proposed association policy outperforms a conventional RAN-centric policy. It adapts to different networks limitations by eliminating the backhaul and/or RAN bottlenecks causing flow blocking. We evaluated the policy performance, based on dynamic simulations, with respect to different varying network parameters, such as the backhaul bandwidth, the traffic model, and the number of concurrent users.

7.2 Open perspectives

By challenging standard paradigms and introducing novel concepts, self-deployable networks open various and interesting perspectives for future work.

Besides the pressing problems discussed in this thesis, other worthwhile subjects related to the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

self-organization of self-deployable networks deserve to be studied. For example, cognitive radio spectrum sensing, neighboring BSs discovery, wireless backhaul set-up, and connectivity establishment to external networks. The corresponding algorithms, protocols and their implementation should satisfy the rapid and unplanned deployment nature of this type of networks. Furthermore, all of these tasks are to be carried by the network autonomously. This brings self-deployable networks closer to ad-hoc networks, and allows building on the advancements made in that field. On the other hand, user mobility was not considered in the presented works. The resulting implications on mobility management procedures are to be further investigated.

We identified in this thesis a number of challenges regarding self-deployable networks configuration, proposed some insights on the network architecture, and provided solutions to the raised challenges. In a next step, substantial efforts are needed to revisit those challenges and tackle the remaining ones from a practical point of view. That is, investigate in more depth how can everything be implemented in a real-life network deployment by answering the following questions: What are the corresponding protocols to implement the proposed solutions, namely the power map computation, the Local CN placement, and the association policy? Which network entities are involved in each case, how and when do they communicate, and what are the exact messages they exchange? How much of the cellular network existing standard can be reused, and what are the new requirements? What are the implications of practical implementation on the network-configuration time, and the end-to-end delays during network operation?

The light-weight and compact nature of self-deployable BSs makes them easily transportable. While we focused in this work on a static RAN with fixed BSs, having a network with moving BSs is envisioned. This is needed when the users served by the self-deployable network are highly mobile, and a pre-determined coverage area is not suitable (e.g., convoys, BSs in backpacks of army officers, surveillance drone networks). BS mobility adds an additional layer of complexity to all of the aforementioned problems, and creates a novel set of challenges. One of the raised problems is the ability to maintain the inter-BS links to preserve connectivity among BSs, and, when needed, the connectivity with external networks [14]. With the Local CN functions co-located with the BSs, the consequences of losing inter-BS connectivity are severe. Indeed, if a BS is incapable of reaching the Local CN, it cannot serve users, which impacts network availability. The network coverage area is also affected and prone to changes, whether because moving BSs usually have smaller coverage areas than fixed BSs, or because of the topology changes due to the different BSs speeds and trajectories [134]. Users are at greater risk of getting out of the coverage area, notably when their speeds and trajectories are incompatible with those of the BSs. While mobility in classical networks is one-sided, it is two-sided in moving self-deployable networks, requiring further investigation.

Finally, self-deployable networks are eliminating the long-standing physical separation between the RAN and the CN counterparts. The result of such a separation has been the clear distinction between two functional layers: an access stratum (between the UE and the RAN), and a non-access stratum (between the UE and the CN). Physically co-locating RAN and CN equipments calls into question the need for such a rigid strata separation. A re-evaluation of both strata functions, the strict split between them, and the corresponding protocols is needed, in order to simplify their implementation and adapt it to the physical architecture. For example, with the core network functions co-located with the RAN functions on a single physical entity, should association (an access stratum attribute) and attachment (a non-access stratum attribute) still be considered as two different procedures, each with its corresponding set of protocols? Self-deployable networks could be a first step toward single stratum cellular networks, a paradigm-shifting concept worth tackling.

In conclusion, while this thesis sets the base of what could and should be done, a multitude of challenges remain unexplored within the novel concept of self-deployable mobile networks.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Bibliography

[1] M. Sauter, *From GSM to LTE-advanced: An Introduction to Mobile Networks and Mobile Broadband*, John Wiley & Sons, 2014.

[2] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on communications*, vol. 59, no. 11, November 2011.

[3] B. S. Manoj and A.H. Baker, "Communication challenges in emergency response," *Communications of the ACM*, vol. 50, no. 3, March 2007.

[4] Federal Communications Comission, "Communications status report for areas impacted by tropical storm Harvey," August 2017.

[5] Airlynx, *https://www.air-lynx.com*, Last visited June 2018.

[6] Klas Telecom, "Create and deploy a private 4G cellular network anywhere in the world on land, sea or air," *http://klastelecom.com/solutions/deployable-4g-lte-networks/*, Last visited June 2018.

[7] Investissements d'avenir, "Bernard Cazeneuve, Emmanuel Macron, Axelle Lemaire, Louis Gautier et Louis Schweitzer saluent le lancement du démonstrateur radiocommunications sécurisées," *http://proxy-pubminefi.diffusion.finances.gouv.fr/pub/document/18/20812.pdf*, April 2016, Last visited June 2018.

[8] J. Oueis, V. Conan, D. Lavaux, R. Stanica, and F. Valois, "Overview of LTE isolated E-UTRAN operation for public safety," *IEEE Communications Standards Magazine*, vol. 1, no. 2, July 2017.

[9] World Economic Forum, "Internet for all: a framework for accelerating internet access and adoption," April 2016.

[10] Project Loon, *https://x.company/loon/*, Last visited June 2018.

[11] P. Castagno, V. Mancuso, M. Sereno, and M. A. Marsan, "Why your smartphone doesn't work in very crowded environments," in *Proceedings of IEEE WoWMoM*, June 2017.

[12] 3GPP TS 22.346, "Technical specification group services and system aspects; isolated Evolved Universal Terrestrial Radio Access Network (E-UTRAN) operation for public safety (Release 13)," September 2014.

[13] K. Gomez, L. Goratti, T. Rasheed, and L. Reynaud, "Enabling disaster-resilient 4G mobile communication networks," *IEEE Communications Magazine*, vol. 52, no. 12, December 2014.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[14] R. Favraud, A. Apostolaras, N. Nikaein, and T. Korakis, "Toward moving public safety networks," *IEEE Communications Magazine*, vol. 54, no. 3, March 2016.

[15] T. X. Brown, "Cellular performance bounds via shotgun cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, November 2000.

[16] S. Sesia, I. Toufik, and M. Baker, *LTE - the UMTS Long Term Evolution*, chapter 2, pp. 25–31, 2015.

[17] M. Sauter, *From GSM to LTE-advanced: An Introduction to Mobile Networks and Mobile Broadband*, chapter 4, pp. 240–243, John Wiley & Sons, 2014.

[18] S. Sesia, I. Toufik, and M. Baker, *LTE - the UMTS Long Term Evolution*, chapter 5, p. 123, 2015.

[19] TETRAPOL Forum, "TETRAPOL specifications Part-PAS 0001-2, version 3.0.0," November 1999.

[20] ETSI TR 102 580 V1.1.1, "Terrestrial trunked radio (TETRA), Release 2," October 2007.

[21] Codan Radio Communications, "P25 radio systems training guide," September 2013.

[22] T. Doumi, M.F. Dolan, S. Tatesh, A. Casati, G. Tsirtsis, K. Anchan, and D. Flore, "LTE for public safety networks," *IEEE Communications Magazine*, vol. 51, no. 2, February 2013.

[23] C. Gruet, X. Pons-Masbernat, and P. Force, "The LTE evolution: private mobile radio networks," *IEEE Vehicular Technology Magazine*, vol. 8, no. 2, June 2013.

[24] Cisco, "Cisco visual networking index: global mobile data traffic forecast update, 2016-2021 white paper," March 2017.

[25] R. Liebhart, D. Chandramouli, C. Wong, and J. Merkel, *LTE for public safety*, John Wiley & Sons, 2015.

[26] Ramon Ferrus, Oriol Sallent, Gianmarco Baldini, and Leonardo Goratti, "LTE: the technology driver for future public safety communications," *IEEE Communications Magazine*, vol. 51, no. 10, October 2013.

[27] 3GPP TS 22.468, "Group communication system enablers for LTE (Release 14)," March 2017.

[28] 3GPP TS 24.379, "Mission critical push to talk (MCPTT) call control; protocol specification (Release 14)," October 2017.

[29] 3GPP TS 23.303, "Proximity-based services (ProSe); stage 2 (Release 14)," May 2017.

[30] THALES, "Thales creates Fed4PMR consortium for broadband professional mobile radio," *https://www.thalesgroup.com/en/worldwide/press-release/thales-creates-fed4pmr-consortium-broadband-professional-mobile-radio*.

[31] H. Mopidevi, D. Rodrigo, O. Kaynar, L. Jofre, and B. A. Cetiner, "Compact and broadband antenna for LTE and public safety applications," *IEEE Antennas and Wireless Propagation Letters*, vol. 10, 2011.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[32] F. Z. Yousaf, P. Loureiro, F. Zdarsky, T. Taleb, and M. Liebsch, "Cost analysis of initial deployment strategies for virtualized mobile core network functions," *IEEE Communications Magazine*, vol. 53, no. 12, December 2015.

[33] Huwawei, "Forging ahead with core network virtualization," *http://www.huawei.com/minisite/has2016/forging-ahead-with-core-network-virtualization*, 2016, Last visited June 2018.

[34] 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 IEEE NetSoft*, April 2015.

[35] Redline Communications, "Deployable networks, added capacity when and where it counts," *http://rdlcom.com/applications/deployable-networks*, Last visited June 2018.

[36] INC. NTT DOCOMO, "Vehicle-mounted transportable mobile base station and backhaul link for disaster relief operation," *https://www.ituaj.jp/*, Last visited June 2018.

[37] I. Bekmezci, O. Koray Sahingoz, and Ş. Temel, "Flying ad-hoc networks (FANETs): a survey," *Ad Hoc Networks*, vol. 11, no. 3, May 2013.

[38] L. Reynaud and T. Rasheed, "Deployable aerial communication networks: challenges for futuristic applications," in *Proceedings of ACM MSWiM*, October 2012.

[39] I. Bor-Yaliniz and H. Yanikomeroglu, "The new frontier in RAN heterogeneity: multi-tier drone-cells," *IEEE Communications Magazine*, vol. 54, no. 11, November 2016.

[40] A. Dhekne, M. Gowda, and R. Choudhury, "Extending cell tower coverage through drones," in *Proceedings of ACM HotMobile*, February 2017.

[41] 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, May 2016.

[42] J. Hoadley and P. Maveddat, "Enabling small cell deployment with HetNet," *IEEE Wireless Communications*, vol. 19, no. 2, 2012.

[43] 3GPP TR 38.874, "Study on integrated access and backhaul (Release 15)," February 2018.

[44] R. Taori and A. Sridharan, "Point-to-multipoint in-band mmWave backhaul for 5G networks," *IEEE Communications Magazine*, vol. 53, no. 1, January 2015.

[45] C. Dehos, J. L. González, A. De Domenico, D. Ktenas, and L. Dussopt, "Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?," *IEEE Communications Magazine*, vol. 52, no. 9, September 2014.

[46] A. Apostolaras, N. Nikaein, R. Knopp, AM. Cipriano, T. Korakis, I. Koutsopoulos, and L. Tassiulas, "Evolved user equipment for collaborative wireless backhauling in next generation cellular networks," in *Proceedings of IEEE SECON*, June 2015.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[47] J. Huang, F. Qian, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," in *Proceedings of ACM MobiSys*, June 2012.

[48] M. Casoni, CA. Grazia, M. Klapez, N. Patriciello, A. Amditis, and E. Sdongos, "Integration of satellite and LTE for disaster recovery," *IEEE Communications Magazine*, vol. 53, no. 3, March 2015.

[49] B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, "Gateway placement optimization in wireless mesh networks with QoS constraints," *IEEE Journal in Selected Areas of Communication*, vol. 24, no. 11, November 2006.

[50] First Responder Network Authority, *https://www.firstnet.gov/*, Last visited June 2018.

[51] Agene National des Fréquences (ANFR), "La bande 700 Mhz," *https://www.anfr.fr/gestion-des-frequences-sites/bande-700-mhz/*, Last visited June 2018.

[52] Agence Nationale des Fréquences (ANFR), "Le cas particulier des réseaux de sécurité et de secours," *www.anfr.fr/licences-et-autorisations/reseaux-professionnels/les-reseaux-mobiles-professionnels-pmr/*, Last visited June 2018.

[53] European Conference of Postal and Telecommunications Administrations (CEPT), "ECC report 218: harmonised conditions and spectrum bands for the implementation of future European Broadband Public Protection and Disaster relief (BB-PPDR) systems," October 2015.

[54] Autorité de Régulation des Communications Électroniques et des Postes (ARCEP), "L'arcep met en consultation publique les modalités d'attribution des fréquences de la bande 2,6 GHz TDD pour le passage à la 4G des réseaux mobiles professionnels," *https://www.arcep.fr/*, 2018, Last visited June 2018.

[55] H. Cao, W. Jiang, T. Javornik, M. Wiemeler, T. T. Nguyen, and T. Kaiser, "Spectrum awareness scheme of the rapidly deployable eNodeB for unexpected and temporary events," in *Proceedings of IEEE CAMAD*, September 2013.

[56] A. Valcarce, T. Rasheed, K. Gomez, S. Kandeepan, L. Reynaud, R. Hermenier, A. Munari, M. Mohorcic, M. Smolnikar, and I. Bucaille, "Airborne base stations for emergency and temporary events," in *Proceedings of Springer PSATS*, June 2013.

[57] D. Gesbert, S. G. Kiani, A. Gjendemsjo, and G. E. Oien, "Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks," *Proceedings of the IEEE*, vol. 95, no. 12, December 2017.

[58] S.E. Elayoubi, O.B. Haddada, and B. Fourestie, "Performance evaluation of frequency planning schemes in OFDMA-based networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, May 2008.

[59] A. S. Hamza, S. S. Khalifa, H. S. Hamza, and K. Elsayed, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, March 2013.

[60] S. Sesia, I. Toufik, and M. Baker, *LTE - the UMTS Long Term Evolution*, chapter 12, pp. 287–291, 2015.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[61] E. Yaacoub and D. Zaher, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, 2012.

[62] A. Simonsson, "Frequency reuse and inter-cell interference co-ordination in E-UTRA," in *Proceedings of IEEE VTC*, April 2007.

[63] X. Mao, A. Maaref, and K. H. Teo, "Adaptive soft frequency reuse for inter-cell interference coordination in SC-FDMA based 3GPP LTE uplinks," in *Proceedings of IEEE GLOBECOM*, November 2008.

[64] K. Begain, G. I. Rozsa, A. Pfening, and M. Telek, "Performance analysis of GSM networks with intelligent underlay overlay," in *Proceedings of IEEE ISCC*, July 2002.

[65] M. Sternad, T. Ottosson, A. Ahlen, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks," in *Proceedings of IEEE VTC*, October 2003.

[66] R. Kwan and C. Leung, "A survey of scheduling and interference mitigation in LTE," *Journal of Electrical and Computer Engineering*, January 2010.

[67] 3GPP TSG RAN WG1 Meeting #41, "Soft frequency reuse scheme for UTRAN LTE (R1-050507)," May 2005.

[68] M. Maqbool, P. Godlewski, M. Coupechoux, and J.M. Klif, "Analytical performance evaluation of various frequency reuse and scheduling schemes in cellular OFDMA networks," *Performance Evaluation*, vol. 67, no. 4, April 2010.

[69] 3GPP TSG RAN WG1 Meeting #42, "Further analysis of soft frequency reuse scheme (R1-050841)," September 2005.

[70] K. Doppler, C. Wijting, and K. Valkealahti, "Interference aware scheduling for soft frequency reuse," in *Proceedings of IEEE VTC*, April 2009.

[71] L. Chen and D. Yuan, "Soft frequency reuse in large networks with irregular cell pattern: how much gain to expect?," in *Proceedings of IEEE PIMRC*, September 2009.

[72] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, December 2011.

[73] G. Giambene, T. Bourgeau, and H. Chaouchi, "Soft frequency reuse schemes for heterogeneous LTE systems," in *Proceedings of IEEE ICC*, June 2015.

[74] T. Q. S. Quek, Z. Lei, and S. Sun, "Adaptive interference coordination in multi-cell OFDMA systems," in *Proceedings of IEEE PIMRC*, September 2009.

[75] M. Ezzaouia, C. Gueguen, M. Yassin, M. Ammar, X. Lagrange, and A. Bouallegue, "Autonomous and dynamic inter-cell interference coordination techniques for future wireless networks," in *Proceedings of IEEE WiMob*, October 2017.

[76] C. Rosenberg Y. Ozcan, S. Jabeen and H. Rivano, "User scheduling on the uplink of multi-cell OFDMA networks: From a system-wide benchmark to practical schemes," *Submitted to IEEE Transactions on Mobile Computing*, 2018.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[77] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell OFDMA system," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, 2006.

[78] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, April 2010.

[79] I. G. Fraimis, V. D. Papoutsis, and S. A. Kotsopoulos, "A decentralized subchannel allocation scheme with inter-cell interference coordination (ICIC) for multi-cell OFDMA systems," in *Proceedings of IEEE GLOBECOM*, December 2010.

[80] S. Cicalo, V. Tralli, and A.I. Perez-Neira, "Centralized vs distributed resource allocation in multi-cell OFDMA systems.," in *Proceedings of IEEE VTC*, May 2011.

[81] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: challenges and opportunities for innovations," *IEEE Communications Magazine*, vol. 53, no. 2, February 2015.

[82] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Elsevier Computer networks*, vol. 47, no. 4, March 2005.

[83] I. F Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Elsevier Computer networks*, vol. 38, no. 4, March 2002.

[84] T. Taleb and A. Ksentini, "Gateway relocation avoidance-aware network function placement in carrier cloud," in *Proceedings of ACM MSWiM*, November 2013.

[85] M. Bouet, J. Leguay, T. Combe, and V. Conan, "Cost-based placement of vDPI functions in NFV infrastructures," *International Journal of Network Management*, vol. 25, no. 6, November 2015.

[86] L.C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 1, March 1977.

[87] R. Chandra, L. Qiu, K. Jain, and M. Mahdian, "Optimizing the placement of integration points in multi-hop wireless networks," in *Proceedings of IEEE ICNP*, October 2004.

[88] J. Wong, R. Jafari, and M. Potkonjak, "Gateway placement for latency and energy efficient data aggregation," in *Proceedings of IEEE LCN*, November 2004.

[89] Y. Bejerano, "Efficient integration of multihop wireless and wired networks with QoS constraints," *IEEE/ACM Transactions on Networking*, vol. 12, no. 6, December 2004.

[90] S. N. Muthaiah and C. Rosenberg, "Single gateway placement in wireless mesh networks," in *Proceedings of ISCN*, June 2008.

[91] D. Das, Z. Rehena, S. Roy, and N. Mukherjee, "Multiple-sink placement strategies in wireless sensor networks," in *Proceedings of IEEE COMSNETS*, October 2013.

[92] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, January 2013.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[93] H.Kim, G.Veciana, X.Yang, and M.Venkatachalam, "Distributed α-optimal user association and cell load balancing in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, February 2012.

[94] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.Wong, R. Schober, and L. Hanzo, "User association in 5G networks: a survey and an outlook," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, April 2016.

[95] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, June 2014.

[96] X. Chen and R. Q. Hu, "Joint uplink and downlink optimal mobile association in a wireless heterogeneous network," in *Proceedings of IEEE GLOBECOM*, December 2012.

[97] H. Boostanimehr and V. Bhargava, "Joint downlink and uplink aware cell association in HetNets with QoS provisioning," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, October 2015.

[98] Aamod Khandekar, Naga Bhushan, Ji Tingfang, and Vieri Vanghi, "LTE-advanced: heterogeneous networks," in *Proceedigs of IEEE EW*, April 2010.

[99] N. Sapountzis, T. Spyropoulos, N. Nikaein, and U. Salim, "Optimal downlink and uplink user association in backhaul-limited HetNets," in *Proceedings of IEEE INFOCOM*, April 2016.

[100] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, June 2014.

[101] A. Youssef M. Kamel, W. Hamouda, "Ultra-dense networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, May 2016.

[102] N. Sapountzis, T. Spyropoulos, N. Nikaein, and U. Salim, "Joint optimization of user association and dynamic TDD for ultra-dense networks," in *Proceedings of IEEE INFOCOM*, April 2018.

[103] T. Qu, D. Xiao, and D. Yang, "A novel cell selection method in heterogeneous LTE-A systems," in *Proceedings of IEEE IC-BNMT*, October 2010.

[104] 3GPP TSG RAN WG1 Meeting #62, "Potential performance of range expansion in macro-pico deployment (R1-104355)," August 2010.

[105] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, June 2016.

[106] J. Ghimire and C. Rosenberg, "Revisiting scheduling in heterogeneous networks when the backhaul is limited," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, October 2015.

[107] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy efficient context-aware user association for outdoor small cell heterogeneous networks," in *Proceedings of IEEE ICC*, June 2014.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[108] T. Zhou, Y. Huang, and L. Yang, "User association with jointly maximising downlink sum rate and minimising uplink sum power for heterogeneous cellular networks," *IEEE IET Communications*, vol. 9, no. 2, January 2015.

[109] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: a disruptive architectural design for 5G networks," in *Proceedings of IEEE Globecom*, December 2014.

[110] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Communications Magazine*, vol. 54, no. 3, March 2016.

[111] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Load and backhaul aware decoupled downlink/uplink access in 5G systems," in *Proceedings of IEEE ICC*, June 2015.

[112] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, May 2016.

[113] H. Qiaoni, B. Yang, G.Miao, C. Chen, X. Wang, and X. Guan, "Backhaul-aware user association and resource allocation for energy-constrained HetNets," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, January 2017.

[114] A. De Domenico, V. Savin, and D. Ktenas, "A backhaul-aware cell selection algorithm for heterogeneous cellular networks," in *Proceedings of IEEE PIMRC*, September 2013.

[115] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Communications Magazine*, vol. 47, no. 4, April 2009.

[116] 3GPP TR 36.814, "Evolved Universal Terrestrial Radio Access: further advancements for E-UTRA physical layer aspects (Release 9)," January 2015.

[117] D. Lpez-Prez, A. Ladnyi, A. Jttner, H. Rivano, and J. Zhang, "Optimization method for the joint allocation of modulation schemes, coding rates, resource blocks and power in self-organizing LTE networks," in *Proceedings of IEEE INFOCOM*, April 2011.

[118] S. Boyd, S.J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimization and Engineering*, vol. 8, no. 1, March 2007.

[119] M. Chiang, C. Tan, D. Palomar, O. Daniel, and D. Julian, "Power control by geometric programming," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, July 2007.

[120] P. Bonami, J. Forrest, C. Laird, F. M. J. Lee, and A. Wchter, "Bonmin: basic open-source nonlinear mixed integer programming," *http://www.coin-or.org/Bonmin*, Last visited June 2018.

[121] "SimPy, discrete event simulation for python," *https://simpy.readthedocs.io/*, Last visited June 2018.

[122] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*, chapter 25, p. 427, John Wiley & Sons, 1991.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

[123] 3GPP TR 23.797, "Technical specification group services and system aspects; study on architecture enhancements to support isolated Evolved Universal Terrestrial Radio Access Network (E-UTRAN) operation for public safety (Release 13)," June 2015.

[124] J. Oueis, R. Stanica, F. Valois, V. Conan, and D. Lavaux, "Core network function placement in mobile networks," in *Proceesings of IEEE PIMRC*, October 2017.

[125] S. P. Borgatti, "Centrality and network flow," *Social networks*, vol. 27, no. 1, January 2015.

[126] L. C. Freeman, "Centrality in social networks: conceptual clarification," *Social Networks*, vol. 1, no. 3, January 1978.

[127] A. Bavelas, "Communication patterns in task oriented groups," *The Journal of the Acoustical Society of America*, vol. 22, no. 6, November 1950.

[128] IBM Corp., "IBM ILOG CPLEX Optimization Studio v12.7.0 documentation," 2016.

[129] J. Oueis, C. Rosenberg, R. Stanica, and F. Valois, "Network-aware user association in public safety oriented mobile networks," in *ACM CoNEXT Workshop I-TENDER*, December 2017.

[130] Cisco, "LTE design and deployment strategies," 2011.

[131] R. Kreher and K. Gaenger, *LTE Signaling: Troubleshooting and Optimization*, chapter 2, pp. 131–186, John Wiley & Sons, 2010.

[132] D. Lopez-Perez and A. Ladanyi, "Optimization method for the joint allocation of modulation schemes, coding rates, resource blocks and power in self-organizing LTE networks," in *Proceedings of IEEE INFOCOM*, April 2011.

[133] The Online Encyclopedia of Integer Sequences, "Number of connected graphs with n nodes A001349 (formerly M1657 N0649)," *http://oeis.org/A001349*, Last visited June 2018.

[134] Y. Sui, J. Vihriala, A. Papadogiannis, M. Sternad, W. Yang, and T. Svensson, "Moving cells: a promising solution to boost performance for vehicular users," *IEEE Communications Magazine*, vol. 51, no. 6, June 2013.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

List of Publications

International Journal Articles

- J. Oueis, V. Conan, D. Lavaux, R. Stanica, and F. Valois, "Overview of LTE isolated E-UTRAN Operation for public safety", *IEEE Communications Standards Magazine*, vol. 1, no. 2, July 2017.
- 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, January 2019.

International Conference Papers

- J. Oueis, R. Stanica, and F. Valois, "Virtualized local core network functions placement in mobile networks". in *Proceedings of IEEE WCNC*, April 2019.
- J. Oueis, C. Rosenberg, R. Stanica, and F. Valois, "Network-aware user association in public safety oriented mobile networks", in *ACM CoNEXT Workshop I-TENDER*, December 2017.
- J. Oueis, V. Conan, D. Lavaux, R. Stanica, and F. Valois, "Core network function placement in mobile networks", in *Proceedings of IEEE PIMRC*, October 2017.

National Workshop Articles

- J. Oueis, V. Conan, D. Lavaux, R. Stanica, and F. Valois, "Placement du coeur d'un réseau mobile autonome", in *ALGOTEL*, May 2017.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : OUEIS DATE de SOUTENANCE : 27/11/2018

Prénoms : Jad

TITRE : RADIO ACCESS AND CORE FUNCTIONALITIES IN SELF-DEPLOYABLE MOBILE NETWORKS

NATURE : Doctorat Numéro d'ordre : 2018LYSEI095

Ecole doctorale : informatique et Mathématiques de Lyon

Spécialité : Informatique

RESUME :

Les réseaux mobiles auto-déployables sont des réseaux qui peuvent être rapidement déployés, facilement installés, sur demande, n'importe où, n'importe quand. Ils ciblent divers cas d'utilisation, pour fournir des services aux utilisateurs lorsque le réseau cellulaire classique ne peut pas être utilisé, ou n'existe pas.
Ces réseaux font évoluer l'architecture d'un réseau cellulaire classique en éliminant la séparation physique entre le réseau d'accès et le cœur de réseau. Dans un réseau auto-déployable, cette séparation est uniquement fonctionnelle, vu qu'une station de base est colocalisée avec les fonctionnalités du réseau de cœur traditionnel, comme la gestion de session et le routage, à travers la virtualisation de ces dernières. Une station de base, sans connexion à un réseau de cœur externe, est capable de fournir des services aux utilisateurs dans sa zone de couverture. Lorsque plusieurs stations de base s'interconnectent, les liens entre eux forment un réseau d'interconnexion qui risque d'avoir une capacité limitée.
Dans cette thèse, nous partons des propriétés distinguant ces réseaux auto-déployables pour revisiter des problèmes classiques du réseau d'accès, et pour aborder de nouveaux défis créés par l'architecture du cœur de réseau et du réseau d'interconnexion. Nous proposons tout d'abord un algorithme qui retourne un schéma d'allocation de fréquences et de puissances pour les stations de bases du réseau d'accès. Ensuite, nous passons à l'organisation du réseau de cœur et nous traitons le problème de placement de ses fonctionnalités. Pour un placement centralisé, nous définissons une nouvelle métrique de centralité dans le réseau qui permet de maximiser le trafic échangé. Pour le placement distribué, nous étudions le nombre et le placement optimaux des différentes fonctionnalités, et nous évaluons divers politiques d'attachement des utilisateurs. Finalement, nous abordons le problème d'association des utilisateurs, en proposant une nouvelle politique d'association adaptée aux réseaux auto-déployables, qui prend en compte le réseau d'accès, le réseau d'interconnexion et les demandes des utilisateurs.

MOTS-CLÉS : Réseaux mobiles auto-déployables, Réseau d'accès, Cœur de réseau

Laboratoire (s) de recherche : CITI

Directeur de thèse : Fabrice Valois

Composition du jury :

CASETTI, Claudio	Associate Professor	Politecnico di Torino	Rapporteur
CHAOUCHI, Hakima	Professeure des Universités	Telecom Sud Paris	Rapporteure
LAGRANGE, Xavier	Professeur des Universités	IMT Atlantique	Rapporteur
CONAN, Vania	Habilité à Diriger des Recherches	Thales	Examinateur
FDIDA, Serge	Professeur des Universités	UPMC	Examinateur
PERROT, Nancy	Docteure	Orange Labs	Examinatrice
VALOIS, Fabrice	Professeur des Universités	INSA-LYON	Directeur de Thèse
STANICA, Razvan	Maître de Conférences	INSA-LYON	Co-encadrant de Thèse

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2018LYSEI095/these.pdf
© [J. Oueis], [2018], INSA Lyon, tous droits réservés