Edge computing-based access network selection for heterogeneous wireless networks

Yue Li

▶ To cite this version:

ANNÉE (2017)

UNIVERSITÉ DE RENNES 1

UNIVERSITE BRETAGNE LOIRE

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Bretagne Loire

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : (Informatique)

Ecole doctorale Mathématiques et STIC (MathSTIC)

présentée par

Yue Li

préparée à IRISA (UMR 6074)
Institut de Recherche en Informatique et Système Aléatoire

Edge Computing-based Access Network Selection for Heterogeneous Wireless Networks

**Thèse soutenue à Rennes
le 29 Septembre 2017**

devant le jury composé de :

M. André-Luc BEYLOT
Professeur, ENSEEIHT / rapporteur

Mme. Hakima CHAOUCHI
Professeur, Telecom Sud Paris / rapporteur

M. Farid NAÏT-ABDESSELAM
Professeur, Université de Paris-Descartes / examinateur

Mme. Isabel AMIGO
Enseignant chercheur, IMT Atlantique / examinatrice

M. Mustapha BOUHTOU
Ingénieur de recherche, Orange Labs, Châtillon / examinateur

M. Yassine HADJADJ-AOUL
Maître de conférences, Université de Rennes 1 / examinateur

M. Philippe BERTIN
Ingénieur de recherche, Orange Labs, Cesson-Sévigné / co-directeur de thèse

M. Gerardo RUBINO
Directeur de recherche, INRIA / directeur de thèse

To my parents

Acknowledgements

my friends from Orange and INRIA. You leave me many beautiful memories. Thanks to my girlfriend XinYu LIANG for her accompany and support during these years.

Last but not the least, I would like to thank my family, my parents DeChun LI and Yan LIU. No matter what decisions I have made, thank you for your endless love, support, full confidence and encouraged me during the most difficult times.

Finally, I would like to thank everybody who helped me for the successful realization of the thesis.

Abstract

Telecommunication network has evolved from First Generation (1G) to Fourth Generation (4G) in the past decades. One of the typical characteristics of the 4G network is the coexistence of heterogeneous radio access technologies, which offers end-users the capability to connect them and to switch between them with their mobile devices of the new generation. However, selecting the right network is not an easy task for mobile users since access network condition changes rapidly. Moreover, video streaming is becoming the major data service over the mobile network where content providers and network operators should cooperate to guarantee the quality of video delivery. In order to cope with this context, the thesis concerns the design of a novel approach for making an optimal network selection decision and architecture for improving the performance of adaptive streaming in the context of a heterogeneous network.

Firstly, we introduce an analytic model (i.e. linear discrete-time system) to describe the network selection procedure considering one traffic class. Then, we consider the design of a selection strategy based on foundations from linear optimal control theory (i.e. Linear–Quadratic Regulator (LQR)), with the objective to maximize network resource utilization while meeting the constraints of the supported services. Computer simulations with MATLAB are carried out to validate the efficiency of the proposed mechanism. Based on the same principal we extend this model with a general analytic model describing the network selection procedures in heterogeneous network environments with multiple traffic classes. The proposed model was, then, used to derive a scalable mechanism based on control theory, which allows not only to assist in steering dynamically the traffic to the most appropriate network access but also helps in blocking the residual traffic dynamically when the network is congested by adjusting dynamically the access probabilities. We discuss the advantages of a seamless integration with the Access Network Discovery and Selection Function (ANDSF). A prototype is also implemented into Network Simulator 3 (ns-3). Simulation results sort out that the proposed scheme prevents the network congestion and demonstrates the effectiveness of the controller design, which can maximize the network resources allocation by converging the network workload to the targeted network occupancy.

Thereafter, we focus on enhancing the performance of Dynamic Adaptive Streaming over HTTP (DASH) in a mobile network environment for the users which has one access

network. We introduce a novel architecture based on Multi-access Edge Computing (MEC). The proposed adaptation mechanism, running as an MEC service, can modify the manifest files in real time, responding to network congestion and dynamic demand, thus driving clients towards selecting more appropriate quality/bitrate video representations. We have developed a virtualized testbed to run the experiment with our proposed scheme. The simulation results demonstrate its Quality of Experience (QoE) benefits compared to traditional, purely client-driven, bitrate adaptation approaches since our scheme notably improves both on the achieved Mean Opinion Score (MOS) and on fairness in the face of congestion.

Finally, we extend the proposed the MEC-based architecture to support the DASH service in a multi-access heterogeneous network in order to maximize the QoE and fairness of mobile users. In this scenario, our scheme should help users select both video quality and access network and we formulate it as an optimization problem. This optimization problem can be solved by IBM CPLEX tool. However, this tool is time-consuming and not scalable. Therefore, we introduce a heuristic algorithm to make a sub-optimal solution with less complexity. Then we implement a testbed to conduct the experiment and the result demonstrates that our proposed algorithm notably can achieve similar performance on overall achieved QoE and fairness with much more time-saving compared to the IBM CPLEX tool.

Keywords: Heterogeneous Network; Network Selection; Control Theory; Dynamic Adaptive Streaming over HTTP; Multi-access Edge Computing; Quality of Experience.

Résumé

Au cours de ces dernières décennies, les réseaux de télécommunications mobiles ont évolué de la première génération (1G) à la quatrième génération (4G). Une des caractéristiques du réseau 4G est de permettre la coexistence de différents réseaux d'accès. Ainsi, les utilisateurs ont la capacité de se connecter à un réseau hétérogène, constitué de plusieurs réseaux d'accès, tout en ayant la possibilité de basculer entre eux. Toutefois, la sélection du réseau approprié n'est pas une tâche facile pour les utilisateurs mobiles puisque les conditions de chaque réseau d'accès changent rapidement. Par ailleurs, en termes d'usage, le streaming vidéo devient le service principal de transfert de données sur les réseaux mobiles, ce qui amène les fournisseurs de contenu et les opérateurs de réseau à coopérer pour garantir la qualité de la diffusion. Dans ce contexte, la thèse propose la conception d'une approche novatrice pour la prise de décision optimale de sélection de réseau et une architecture améliorant les performances des services de streaming adaptatif dans un réseau hétérogène.

En premier lieu, nous introduisons un modèle analytique décrivant la procédure de sélection de réseau en ne considérant déjà qu'une seule classe de trafic. Nous concevons ensuite une stratégie de sélection basée sur des fondements de la théorie du contrôle optimal linéaire (c'est-à-dire régulateur quadratique linéaire ou Linear–Quadratic Regulator (LQR)), avec l'objectif de maximiser l'utilisation des ressources réseau tout en répondant aux contraintes des services supportés. Des simulations sous MATLAB sont effectuées pour valider l'efficacité du mécanisme proposé. Sur ce même principe, nous étendons ce modèle avec un modèle analytique général décrivant les procédures de sélection de réseau dans des environnements de réseaux hétérogènes avec de multiples classes de trafic. Le modèle proposé est ensuite utilisé pour dériver un mécanisme adaptatif basé sur la théorie du contrôle, qui permet non seulement d'aider à piloter dynamiquement le trafic vers l'accès réseau le plus approprié mais aussi de bloquer dynamiquement le trafic résiduel lorsque le réseau est congestionné en ajustant les probabilités d'accès optimales. Nous discutons aussi les avantages d'une intégration transparente du mécanisme proposé avec l'ANDSF (Access Network Discovery and Selection Function), solution fonctionnelle normalisée pour la sélection de réseau. Un prototype est également implémenté dans ns-3 (Network Simulator 3). Les résultats de simulation permettent de vérifier que le système proposé empêche la congestion

du réseau et de démontrer l'efficacité du contrôleur pour maximiser l'allocation des ressources.

En second lieu, nous nous concentrons sur l'amélioration des performances de DASH (Dynamic Adaptive Streaming over HTTP) pour les utilisateurs mobiles dans un environnement de réseau d'accès 4G uniquement. Nous introduisons une nouvelle architecture basée sur l'utilisation de serveurs distribués en périphérie de réseau suivant le standard Multi-access Edge Computing (MEC). Le mécanisme d'adaptation proposé, fonctionnant en tant que service MEC, peut modifier les fichiers de manifeste (description de la structure selon laquelle est organisée la diffusion du contenu) en temps réel, en réponse à la congestion du réseau et à la demande dynamique de flux. Ces modifications conduisent ainsi les clients à sélectionner des représentations vidéo de débit / qualité plus appropriées. Nous avons développé une plateforme de tests virtualisée pour l'expérimentation de notre proposition. Les résultats obtenus démontrent ses avantages en terme de Qualité d'Expérience (QoE) comparés aux approches d'adaptation traditionnelles, purement pilotées par les clients, car notre approche améliore non seulement le score d'opinion moyen (MOS) mais aussi l'équité face à la congestion.

Enfin, nous étendons l'architecture proposée basée sur MEC pour supporter le DASH dans un réseau hétérogène afin de maximiser la QoE et l'équité des utilisateurs mobiles. Dans ce scénario, notre mécanisme doit aider les utilisateurs à sélectionner la qualité vidéo et le réseau et nous le formulons comme un problème d'optimisation. Ce problème d'optimisation peut être résolu par l'outil IBM CPLEX, mais cela prend du temps et ne peut être envisagé à grande échelle. Par conséquent, nous introduisons une approche heuristique pour aborder la solution optimale avec moins de complexité. Ensuite, nous mettons en œuvre une expérimentation sur notre plateforme de tests. Le résultat démontre que, par rapport à l'outil IBM CPLEX, notre algorithme permet d'obtenir des performances similaires sur la QoE globale et l'équité, avec un gain de temps significatif.

Mots Clés: Réseau Hétérogène ; Sélection de Réseaux ; Théorie du Contrôle ; Streaming adaptatif dynamique sur HTTP ; Multi-access Edge Computing ; Qualité d'expérience.

Résumé de la thèse

Introduction

Avec l'évolution des technologies des réseaux de télécommunication, nous assistons ces dernières années à l'explosion des usages des services de données mobiles à large bande. Ainsi, selon Cisco Visual Networking Index (VNI) [1], le trafic de données mobiles a augmenté de 4 000 fois au cours des 10 dernières années. On prévoit qu'il augmentera encore presque huit fois entre 2015 et 2020, atteignant $30,6$ exaoctets par mois d'ici 2020. Cependant, la capacité des réseaux n'augmente pas à la vitesse de la croissance des données mobiles en raison de la limite théorique des canaux radio [2], ce qui représente un grand défi pour les opérateurs de télécommunications. Afin d'y faire face, l'une des solutions consiste à densifier le réseau cellulaire en multipliant les nœuds d'accès. Des micro stations de base BS (Base Station), comme les picocell et femtocell, qui ont une puissance d'émission réduite, sont déployées en combinaison avec le déploiement de macro BS pour améliorer la couverture du réseau et la capacité. Le Wi-Fi est également intégré au déploiement en complément des points d'accès intérieurs et extérieurs pour offrir une expérience de haut débit. Ce type de déploiement qui implique un mélange de technologies radio et de types de cellules travaillant ensemble de façon coordonnée constitue un réseau hétérogène [3]. Avec la diversité des technologies d'accès radio, les dispositifs mobiles évoluent vers des terminaux multi-modes équipés d'interfaces radio multiples ou d'une seule interface radio re-configurable leur permettant de se connecter à différentes technologies de réseaux sans fil. Les utilisateurs sont amenés à tirer parti de cette diversité des technologies d'accès et passer facilement d'un système d'accès à l'autre pour optimiser leur rentabilité ou améliorer la perception de la QoS (qualité de service). Du point de vue des utilisateurs et des terminaux, l'enjeu est de parvenir à obtenir à tout moment la meilleure connexion avec le réseau d'accès optimal (i.e. ABC "Always Best Connected") [4]. Du point de vue des opérateurs de réseau, l'assistance à la sélection de réseaux pour les terminaux mobiles peut améliorer l'utilisation des ressources du réseau et fournir une QoE (qualité d'expérience) la meilleure et la

plus équitable pour l'ensemble des utilisateurs. Le réseau hétérogène étant basé sur différentes technologies, y compris les technologies 3GPP (HSPA LTE, etc.) et non 3GPP (Wi-Fi et WiMAX) avec des caractéristiques différentes, celles-ci doivent être gérées conjointement et coordonnées pour optimiser la capacité globale. Au cours des dernières années, la sélection des réseaux a été largement étudiée par l'industrie et les chercheurs. Cependant, les approches existantes ne peuvent que partiellement résoudre ce problème. La limitation de la transparence envers les utilisateurs mobiles et la complexité des modèles sont un frein pour une mise en place dans un réseau réel. Par conséquent, dans cette thèse, un de nos objectifs est d'étudier le problème de sélection de réseaux dans le contexte de réseaux hétérogènes et de proposer de nouvelles approches pour aborder ce problème.

D'autre part, le déploiement rapide des réseaux LTE et LTE-A permettent d'ores et déjà aux opérateurs de fournir des services vidéo HD (haute définition) mobiles, services considérés comme fondamentaux également pour les futurs réseaux mobiles 5G. Selon Cisco [1], le trafic vidéo mobile, qui représentait 55% du trafic mobile total en 2015, représentera plus de 70% en 2020. Lorsque les utilisateurs regardent des vidéos avec leur appareils mobiles tels que les smartphones, ils ne stockent pas une vidéo entière, de sorte que celle-ci est progressivement téléchargée, en plusieurs segments. Pour implémenter ce mécanisme, DASH (Dynamic Adaptive Streaming over HTTP) a été introduit comme une norme de diffusion de flux adaptative. DASH permet aux clients d'ajuster le débit du flux vidéo en surveillant les conditions réseaux. Cependant, les conditions du réseau sont dynamiques et varient beaucoup plus rapidement dans un environnement sans fil que sur un réseau filaire. Les applications de terminaux mobiles estiment souvent de façon imprécise la bande passante du réseau et le débit, ce qui rend difficile de promettre aux utilisateurs une QoE acceptable. En outre, le problème devient plus complexe dans un environnement de réseau hétérogène où les utilisateurs mobiles accèdent de façon concurrente aux différentes ressources sans fil. Ainsi, de nouvelles architectures réseaux prenant en compte ces facteurs émergent. L'exploitation du Multi-access Edge Computing (MEC) est certainement une approche intéressante afin d'atténuer ce problème en tirant partie de la proximité entre les utilisateurs et les réseaux d'accès. Cependant, le standard MEC ne tient pas compte pour le moment de cas d'utilisation dans un environnement de réseau hétérogène. Par conséquent, nous menons une étude plus approfondie dans cette thèse sur la diffusion vidéo basée sur des flux adaptatifs dans le cadre de ce contexte.

Contribution

Selon les discussions ci-dessus sur les objectifs et les défis, nous constatons que pour faire face à la demande toujours croissante de trafic de données sur Internet, le réseau mobile évolue vers un environnement de réseau hétérogène où différentes technologies d'accès telles que LTE, 3G, Wi-Fi et WiMAX coexistent avec des couvertures qui se chevauchent. Une sélection de réseau appropriée peut bénéficier à la fois aux utilisateurs mobiles et aux opérateurs de réseau. Dans le même temps, le trafic multimédia est le principal consommateur de bande passante réseau et domine le trafic Internet. Délivrer efficacement la vidéo est une préoccupation majeure pour les opérateurs de réseaux mobiles et les fournisseurs de contenu. En conséquence, dans cette thèse, notre effort se concentre sur deux problèmes difficiles dans les environnements réseaux hétérogènes actuels: la sélection de réseaux et l'amélioration du streaming de flux video basé sur DASH dans le contexte des réseaux mobiles.

Nos premiers travaux de recherche se concentrent sur la sélection de réseau LTE et Wi-Fi avant d'introduire la gestion de multiples classes de service.

1. **Mécanisme de sélection de réseau basé sur la théorie de contrôle pour un inter-fonctionnement LTE et WiFi**. Dans la première contribution, nous abordons le problème de sélection de réseau dans LTE et Wi-Fi en utilisant la théorie de contrôle. En premier lieu, nous proposons un modèle analytique pour décrire la procédure de sélection du réseau dans LTE et Wi-Fi pour les utilisateurs avec une seule classe de trafic. Ensuite, nous analysons ce modèle en utilisant la théorie de contrôle et concevons un contrôleur LQR (Linear–quadratic regulator) pour contrôler la charge dans chaque réseau. Nous mettons en œuvre le contrôleur et validons notre conception avec MATLAB. Les résultats numériques démontrent que le nombre d'appareils mobiles dans chaque réseau d'accès peut converger vers l'occupation cible du réseau. Cette partie du travail a été publiée dans HPSR 2015 [6].

2. **Mécanisme de sélection de réseaux basé sur la théorie de contrôle dans un réseau hétérogène multi-classes**. Dans la deuxième contribution, nous étendons le modèle de sélection de réseau développé dans [6] à un modèle général avec un trafic multi-classes dans un réseau multi-accès. Ensuite, nous appliquons la théorie du contrôle pour concevoir un contrôleur LQR pour calculer les probabilités d'accès pour chaque réseau. Nous discutons des détails de faisabilité d'intégration avec l'ANDSF (Access Network Discovery and Selection Function) pour notre proposition de contrôleur. Enfin, nous mettons en œuvre un prototype avec ns-3

afin de vérifier que le contrôleur peut être appliqué au réseau réel pour aider les utilisateurs mobiles à prendre des décisions de sélection de réseau optimale. Cette partie du travail a été publiée dans CCNC 2016 [7].

Ensuite, nous dirigeons nos travaux de recherche pour aborder le problème de performance du service de streaming basé sur DASH dans un réseau mobile MEC puis son élargissement au réseau hétérogène.

1. **Une architecture basée sur Multi-access Edge Computing (MEC) pour améliorer la performance DASH dans le réseau mobile**. Dans cette contribution, nous proposons une architecture novatrice basée sur MEC pour améliorer le service de streaming adaptatif DASH dans un contexte de réseau mobile à accès unique et nous décrivons deux possibilités d'implémentation: le mode redirection et le mode proxy pour les flux de communication. Ensuite, nous développons un algorithme d'adaptation qui s'exécute sur le serveur MEC pour contrôler dynamiquement les représentations vidéo disponibles que les clients peuvent télécharger en fonction de l'état actuel du réseau, disponible via une API MEC exposée par le l'opérateur mobile (MNO). Nous mettons en œuvre un testbed notre expérimentation révèle que l'algorithme proposé améliore la qualité d'expérience (QoE) et l'équité pour les utilisateurs mobiles en cas de congestion du réseau. De plus, notre mécanisme est transparent pour les clients (players DASH), conforme aux standards et compatible avec les algorithmes de livraison vidéo adaptative pilotés par le récepteur. Ce travail a été publiée dans CSCN 2016 [8].

2. **La sélection de qualité vidéo et de réseaux pour l'optimisation de la performance DASH basée sur MEC dans un réseau hétérogène**. Dans cette dernière contribution, nous étendons l'architecture MEC présentée dans [8] pour optimiser la QoE des utilisateurs mobiles du service de flux basé sur DASH dans un réseau hétérogène. La qualité vidéo et la sélection du réseau sont les deux facteurs que nous considérons dans ce travail. Premièrement, nous formulons le problème de la sélection de qualité vidéo et de réseau comme un problème d'optimisation. Ce modèle vise à maximiser la qualité d'expérience globale perçue par tous les clients en considérant la capacité de différents réseaux. Le problème d'optimisation peut être résolu avec l'outil IBM CPLEX. Cependant, cet outil ne peut être mis à l'échelle lorsque le nombre de clients augmente. Par conséquent, nous concevons une heuristique évolutive pour trouver la solution optimale. Pour intégrer l'algorithme dans le système réel, nous décrivons l'architecture basée

sur MEC et les flux de communication. Ensuite, nous mesurons l'optimale de l'heuristique proposée par rapport à l'outil CPLEX et les résultats montrent que la solution heuristique permet un gain de temps conséquent par rapport à la solution CPLEX avec une perte d'optimalité ne dépassant pas 20 %. Enfin, nous émulons le réseau hétérogène réel en implémentant le testbed et évaluons l'algorithme heuristique. Les résultats d'expérimentation montrent que notre système a notamment amélioré le MOS (score d'opinion moyen) et l'équité en cas de congestion du réseau. Cette partie du travail a été acceptée pour être publiée à ISCC 2017 [9]

Contenu de la thèse

Les travaux de thèse sont sunthétisés dans les 7 chapitres de ce rapport:

Le Chapitre 1 décrit brièvement l'évolution des réseaux de télécommunications et identifie deux grands défis actuels auxquels la thèse s'attaque: la sélection d'accès et la diffusion de flux adaptatif dans un réseau hétérogène. Ensuite le projet européen dans lequel s'inscrit une partie des travaux est introduit. Enfin les contributions majeures sont présentées.

Le Chapitre 2 passe en revue le contexte du problème de sélection de réseau et étudie l'état de l'art sur les aspects théoriques et pratiques. Dans l'aspect théorique, nous présentons quatre approches mathématiques pour décrire le problème de sélection de réseau: la théorie de l'utilité, la logique floue, le processus de décision de Markov et la théorie des jeux. Ensuite, la limitation de chaque approche est discutée. Pour l'aspect pratique, nous présentons la solution ANDSF (Assisted Network Selection) et le mobile offloading, puis nous analysons les lacunes de ces solutions.

Le Chapitre 3 présente l'histoire et le principe du streaming adaptatif et son enjeu actuel dans un véritable environnement de réseau. Ensuite, nous examinons le travail connexe d'amélioration de la performance de streaming à partir de trois aspects: les approches basés client, proxy et SDN. Les inconvénients de chaque approche sont étudiés.

Le Chapitre 4 propose un schéma de sélection de réseaux basé sur la théorie de contrôle dans un réseau hétérogène. Nous commençons à décrire la procédure de sélection de réseau en LTE et Wi-Fi avec une classe de trafic. Nous présentons ensuite la conception d'un contrôleur optimal. Après avoir validé le contrôleur dans Matlab, nous étendons la procédure à un scénario de réseaux hétérogènes multi-classes de traffic.

Puis un contrôleur optimal est introduit avec le même principe et nous évaluons notre approche dans un environnement réseau simulé avec ns3.

Le Chapitre 5 présente une architecture basée sur Multi-access Edge Computing (MEC) pour l'amélioration de la livraison vidéo adaptative sur HTTP. Nous proposons un algorithme intelligent s'exécutant sur un serveur MEC et nous proposons ensuite une architecture que nous évaluons dans un environnement virtualisé.

Le Chapitre 6 étend l'architecture basée sur MEC dans le contexte de réseaux hétérogènes en permettant une sélection dynamique de la qualité vidéo et du réseau en temps réel. Nous formulons le problème sous la forme d'un problème d'optimisation et proposons une heuristique pour trouver une solution optimale. Une expérimentation est ensuite conduite afin de comparer le mécanisme proposé avec d'autres approches.

Le Chapitre 7 conclut la thèse et discute quelques orientations possibles pour de futurs travaux de recherche.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

With the evolution of mobile telecommunication networks technologies, the supported services went from analog telephony to mobile multimedia broadband data. Therefore, the telecommunication industry has experienced a significant mobile data growth in recent years. Indeed, according to the Cisco Visual Networking Index (VNI) [1], mobile data traffic has grown 4000-fold over the 10 past years. It has been predicted that it will still increase nearly eight-fold between 2015 and 2020 to reach 30.6 exabytes per month by 2020 as illustrated in Figure 1.1. However, the increase of network capacity doesn't keep up with the speed of mobile data growth due to the theoretic radio channel limit [2], which brings big challenges for telecommunication operators to face the ever increased data challenge with limited network resources. Therefore, mobile operators have investigated in innovating network technologies to support the requirements of future wireless networks. For this purpose, one of the solutions is to increase the access node density. Micro Base Station (BS)s, such as picocell BS and femtocell BSs, which have lower transmission range, are deployed in combination with the deployment of macro BSs to improve the network coverage and capacity. Wi-Fi is also integrated into the deployment as a complement in small indoor/outdoor hotspots to deliver a high-quality mobile broadband experience. This kind of network, which involves a mix of radio technologies and cell types working together seamlessly is referred to as heterogeneous network [3]. A typical heterogeneous network deployment is illustrated in Figure 1.2. With the diversity of radio access technologies, mobile devices are evolving towards multi-mode terminals equipped with various radio interfaces allowing them to connect to any existing wireless networks. Users will need to take benefit of this diversity of the access technologies and seamlessly switch between different access

systems to maximize their profitability or improve the perceived Quality of Service (QoS). From the users' perspective, they manage to get the Always Best Connection (ABC) [4] at any time anywhere with the optimized network selection. From the network operators' perspective, assisting network selection for mobile users can improve network resource utilization and provide better Quality of Experience (QoE). Since the heterogeneous network is based on different Radio Access Technologies (RAT) including 3GPP technologies (High Speed Packet Access (HSPA) LTE, etc.) and non-3GPP (Wi-Fi and WiMAX) with different characteristics, they should be jointly managed and coordinated to deliver high network performance. In recent years, network selection has been widely studied by industry and research communities. However, existing approaches can only partially solve this issue. The limitation of the transparency to mobile users and the model's complexity, which presents some difficulties when implemented in real networks. Therefore, in this thesis, one of our work is to study the network selection problem in the heterogeneous network environment and propose novel approaches to address this problem.
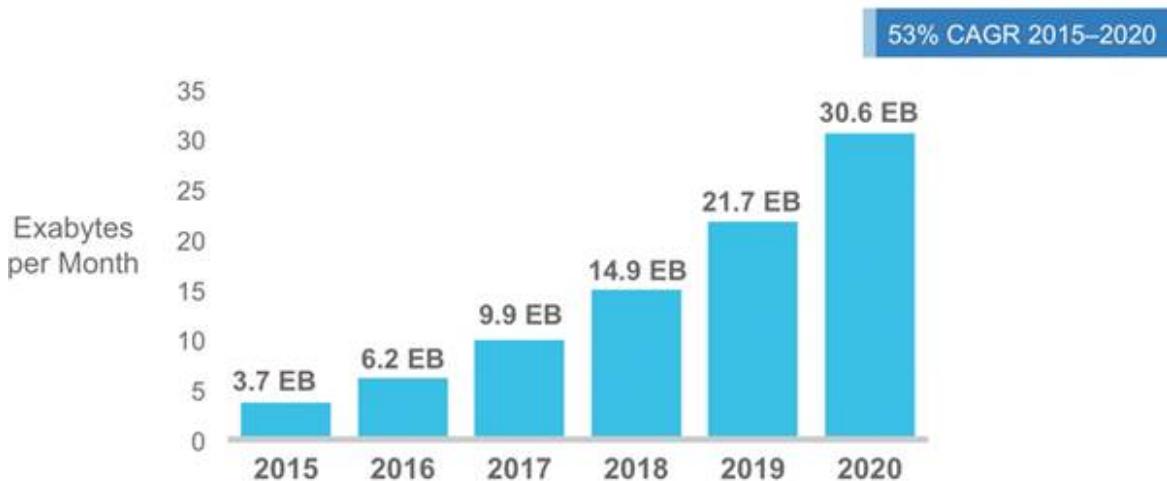
Figure 1.1 – Data Growth Prediction, source [1]

On the other hand, the fast deployment of LTE networks will enable the operators to provide mobile broadband applications, such as mobile High Definition (HD) video services, which will be fundamental in future mobile networks. In fact, according to Cisco [1], mobile video traffic, which accounted for 55% of the total mobile traffic in 2015, will represent more than 70% in 2020. When mobile users watch the video with their mobile devices, such as smartphones and tablets, they do not buffer an entire video, so the video is progressively downloaded in multiple chunks. To implement this mechanism, Dynamic Adaptive Streaming over HTTP (DASH) has been introduced
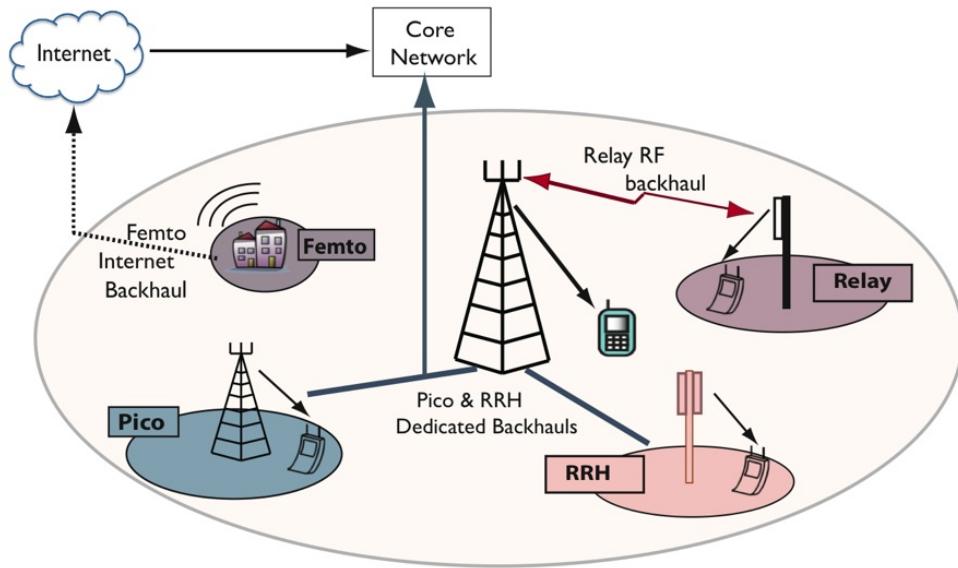
Figure 1.2 – Heterogeneous Network Deployment

as an adaptive streaming standard. DASH allows the client to adjust the rate of the video stream by monitoring the network conditions perceived by this stream. However, the network conditions change much more rapidly in a mobile environment than a fixed network. Mobile users often imprecisely estimate the network bandwidth and the throughput, which makes it difficult to promise end users acceptable QoE. Moreover, the problem becomes more complicated in heterogeneous network environments when mobile users compete for wireless network resources. The new architecture, which has a quick awareness of network information should be designed. Multi-access Edge Computing (MEC) is a good candidate to mitigate this issue with its advantage of the proximity of users and access networks. Moreover, the MEC standard does not consider the use case for video streaming delivery in a heterogeneous network environment. Therefore, we make a deeper study on the video delivery based on the adaptive streaming in the MEC in this thesis.

1.2   Context of the Thesis

Our work is partly performed within the COMBO (COnvergence of fixed and Mobile BrOadband access/aggregation network) [5] project. COMBO is a European Union FP7 ICT Integrated Project, undertaken by 16 partners that started on January 1st,

2013. To achieve the convergence of fixed and mobile networks, COMBO define, develop and technically assess network scenarios organized around the concept of Next Next Generation Point of Presence (NG-POP) and which embody the most promising directions for Fixed–Mobile Convergence (FMC) at the network level. These proposals lead to networks with a more effective distribution of essential functions and improved utilization of equipment and infrastructures. The project aims to define and develop FMC architectures for future networks, demonstrate experimentally key FMC network features to show the feasibility of proposed architecture and influence standardization bodies with respect to FMC architectures to push COMBO concepts.

Our work in this project is to investigate the intelligent network selection in a heterogeneous network which is an FMC scenario while optimizing video delivery over a mobile network using the Multi-access Edge Computing (MEC) architecture, an enabler for implementing FMC architecture.

1.3 Contributions of the Thesis

According to the above discussions on objectives and challenges, we note that to face the ever increasing demand for data traffic over the Internet, the mobile network is evolving towards a heterogeneous network environment where different access technologies such as LTE, 3G, WiFi, and WiMAX co-exist with overlapping coverage. An appropriate network selection can benefit for both mobile users and network operators. At the same time, multimedia applications, which dominate the Internet traffic, are the major consumer of heterogeneous network bandwidth. How to deliver the video becomes the major concern for mobile network operators and content providers. As a consequence, in this thesis, our effort focuses on two challenging issues in the current heterogeneous network environments: network selection and DASH-based streaming service improvement over the mobile network.

Therefore, the major contributions are divided into two axis: firstly our research addresses network selection.

1. **Control theory-based network selection framework in LTE and WiFi interworking**. In the first contribution, we address the network selection problem in LTE and Wi-Fi interworking by leveraging control theory. Firstly we propose an analytical model to describe the user's network selection procedure in LTE and Wi-Fi interworking with one traffic class. Then we analyze this model using control theory and design a Linear–Quadratic Regulator (LQR) controller to control the workload in each network. We implement the controller and validate

our design with MATLAB. Numerical results demonstrate that the number of mobile devices in each access network can converge to the target network occupation. This part of the work has been published in HPSR 2015 [6].

2. **Control theory-based network selection framework in a multi-access heterogeneous network**. In the second contribution, we extend the network selection model developed in [6] to a general model with multi-class traffic in a multi-access network environment. Then we apply the control theory to conceive an LQR controller to calculate the access probabilities' for each network. We discuss the integration feasibility details with the Access Network Discovery and Selection Function (ANDSF) for our proposed controller. Finally we implement a prototype with Network Simulator 3 (ns-3) to prove that the controller can be applied in real network conditions by helping mobile users making optimized network selection decisions. This part of the work has been published in CCNC 2016 [7].

Then, we direct our research work to tackle the performance problem of Dynamic Adaptive Streaming over HTTP (DASH)-based streaming service in heterogeneous network

1. **Multi-access Edge Computing (MEC) based architecture for improving the DASH performance in mobile network**. In the first contribution of this part, we propose a novel architecture based on MEC for improving the performance of DASH-based video streaming in a single access mobile network context and we describe two implementation possibilities: redirect mode and proxy mode by using call flow. Then we develop an adaptation algorithm running in the MEC server to dynamically control the video representations available that clients can download based on the current network status collected via a MEC API exposed by the Mobile Network Operator (MNO). We implement a testbed and the experiment reveals that our proposed algorithm can improve the Quality of Experience (QoE) and fairness for mobile users in case of network congestion. Moreover, our mechanism is transparent to clients and it is standards-compliant and compatible with receiver-driven adaptive video delivery algorithms. This part of the work has been published in CSCN 2016 [8].

2. **Video quality and network selection for optimizing the DASH performance based on MEC in a heterogeneous network environment**. In the second contribution, we extend the MEC-based architecture presented in [8] to

optimize mobile users' QoE of DASH-based streaming service in a multi-access heterogeneous network. The video quality and network selection are the two factors we consider in this work. Firstly we formulate the video quality and network selection problem as an optimization problem. This model aims at maximizing the overall QoE of all the users when considering the capacity of different networks. The optimization problem can be solved with IBM CPLEX tool. However, this tool is not scalable when the number of users increases as demonstrated in the thesis. Therefore we design a scalable heuristic algorithm to find the optimized solution. For integrating the algorithm into the real system, we describe the MEC-based architecture and call flow. Then, we measure the optimality of the proposed heuristic algorithm compared with the CPLEX tool and the results show that the heuristic solution is much more time-saving than CPLEX solution without losing more than 20% optimality. Finally, we emulate the real heterogeneous network by implementing a testbed and evaluate the proposed heuristic algorithm. The simulation results show that our scheme notably improved the achieved Mean Opinion Score (MOS) and fairness in case of network congestion. This part of the work has been accepted for publication in ISCC 2017 [9].

1.4 Organization of the Manuscript

The thesis consists of 7 chapters and the rest of this thesis is organized as follows:

Chapter 2 reviews the background of network selection problem and studies the state of the art from theoretical and practical ways. In the theoretical aspect, we present four mathematical approaches to describe the network selection problem: utility theory, fuzzy logic, Markov decision process and game theory. Then the limitation of each approach is discussed. On the practical side, we present the ANDSF-assisted network selection and the mobile offloading solution then we analyze the shortcoming of these solutions.

Chapter 3 presents the history and principle of HTTP Adaptive Streaming (HAS) and its current issue in a real network environment. Then we survey the relate work of improving the HAS performance from three aspects: client-based, proxy-based and SDN-based approaches. The disadvantages of each approach are investigated.

Chapter 4 proposes a control theory-based network selection scheme in a heterogeneous network environment. We start describing the network selection procedure in LTE and Wi-Fi interworking with one traffic class and designs an LQR controller.

After validating the designed controller with MATLAB, we extend the procedure in the multi-access network scenario with multi-class traffic. Then the LQR controller is introduced with the same principle and we evaluate our approach in a real network environment with Network Simulator 3 (ns-3).

Chapter 5 introduces a Multi-access Edge Computing (MEC) based architecture DASH-based video delivery improvement. We propose an intelligent algorithm running in the MEC server and then the proposed architecture is evaluated in a virtualized environment.

Chapter 6 extends the MEC-based architecture to be applied to the multi-access heterogeneous network to support a dynamic video quality and network selection in real time. We formulate the problem as a Binary Integer Programming optimization problem and propose a heuristic algorithm to find a sub-optimal solution which is more time-saving. The experiment is conducted to compare the algorithm with other approaches.

Chapter 7 gives concluding remarks and some possible future research directions are discussed.

Chapter 2

Network Selection in Heterogeneous Network: From Theory to Practice

2.1 Introduction

Network operators have started to deploy different radio access technologies such as Long Term Evolution (LTE), Wireless Local Area Network (WLAN), Worldwide Interoperability for Microwave Access (WiMAX), Universal Mobile Telecommunications System (UMTS) and Global System for Mobile Communications (GSM) to deal with the tremendous data growth and this makes the coexistence of multiple Radio Access Technologies (RAT). Mobile users are under the coverage of the overlapping radio access network and the Figure 2.1 illustrates a heterogeneous wireless network scenario. In this environment, mobile users can use different types of mobile devices to satisfy their various application requirements. The mobile devices can access to diverse network technologies and each user has different preferences. Therefore, selecting the appropriate network to ensure Always Best Connection (ABC) becomes a complex decision problem. This creates the needs for developing new technologies and standards that seek to provide dynamic and automatic network selection decision for different users' current needs. Normally, a network selection decision is made at service setup or handoff execution and its procedure is illustrated in Figure 2.2: Decision criteria is collected as the input of the decision-making module. Then the decision-making module will apply the network selection approaches to generate a ranked list of best networks as an output for mobile users. In this chapter, we focus on discussing the

state of the art of mathematical and practical approaches from academia to industrial solution in the decision-making module.
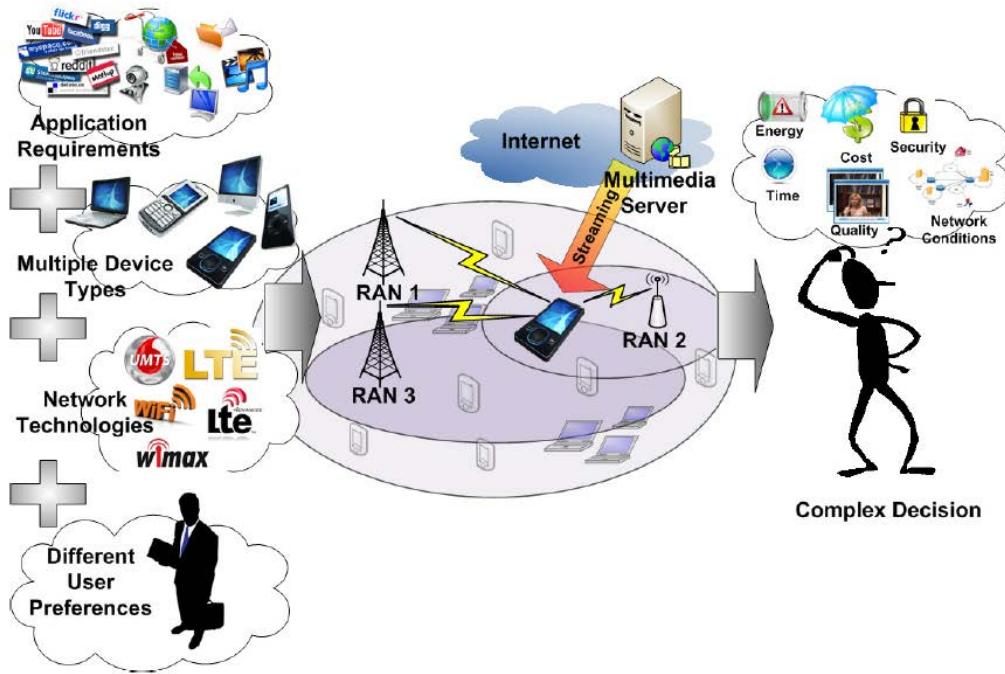
Figure 2.1 – Multiple RAT Scenario, source [10]

2.2 Mathematical Approaches

2.2.1 Utility Theory-Based Approach

Utility theory is originally applied in economic to study the satisfaction that a goods or service can provide to the decision maker [11]. This theory is also suitable to be applied in the network selection decision problem. Applying the utility theory to network selection problem is a user's preference driven approach. Before making a decision, mobile users evaluate their satisfaction about the network attributes (also called criterion) such as bandwidth, delay, packet loss by measuring the utility of the attribute [12] [13]. Then mobile users will then make a decision to select the most suitable network according to the evaluated utility of each network.

The utility of a network attribute is described by its associated utility function. [14] gives some examples of utility functions, which are illustrated in Figure 2.3. To calculate the utility value of a given attribute, it is important to choose the appropriate
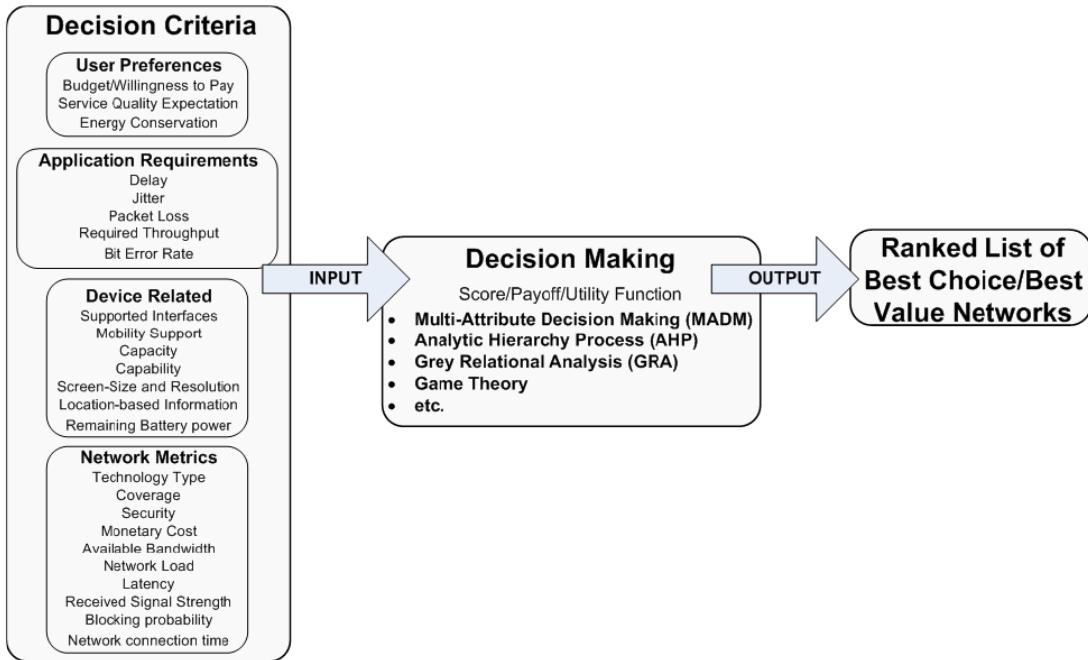
Figure 2.2 – Network Selection Decision Making Procedure, source [10].

utility function which can model its relative utility evolution [15]. For example, [16] pointed out that the sigmoidal function was the most suitable for modeling the utility form in the network selection problem. The form of the sigmoidal function is defined by its various parameters. The parameters should be tuned to adapted to different network attributes features.
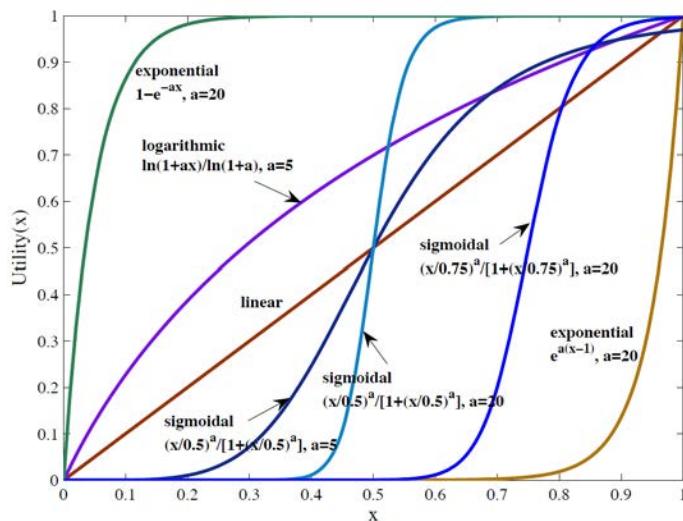
Figure 2.3 – Sigmoide Function Examples, source [16].

When mobile users launch the network selection procedure, multiple network attributes should be considered together to calculate a combined total utility or cost [17]. Since heterogeneous network involves a variety of network attributes: link quality, availability, throughput, network load, file transfer delay, bandwidth, the cost of service, etc. Users can make network selection decision based on single-criterion utility or multiple-criterion utility function. In the area of multi-criteria selection, the problem can be formulated as Multiple Attribute Decision Making (MADM). MADM algorithms which solve the network selection problem include Simple Addicive Weighting (SAW) [16], Multiplicative Exponential Weighting (MEW) [16] [18], Gray Relational Analysis (GRA) [19] [20], Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) [21] [22] [23] [24], etc. For example, in SAW, the overall utility of a candidate network is determined in below:

$$U(x) = \sum_{i=1}^{n} p_i u_i(x_i), \ where \ \sum_{i=1}^{n} p_i = 1 \tag{2.1}$$

where x_i is the value i_{th} criteria, p_i is the weight for criteria i, and $u_i(x_i)$ is the single utility for x_i calculated by its single utility function

In this process, each single utility u_i is scaled by a coefficient p_i according to the users' preference of network attributes and then all the results are added together to get the aggregate utility. This method is very simple and widely used. The selected network is the one which has the highest total utility value. The advantage of additive utility is the independence between different single utility criterion. However, addictive utility approach has its limitation since the absolute independence between criterions will cause a fault in some cases. In the following example, there are two networks and the utility values are shown in Table 2.1. After a simple calculation, the Network B has a higher aggregate utility. However, it is impossible to connect the network B since no bandwidth will be allocated. Moreover, utility theory is a static approach, which is another limit of this solution. Since the network condition is always changing,

Table 2.1 – Network selection with addictive utility approach

Attribute	p_i	A	B
$u(bandwidth)$	0.2	0.4	0
$u(price)$	0.2	0.2	0.4
$u(QoS)$	0.25	0.6	0.8
$u(mobility)$	0.25	0.2	0.6
Aggregate utility		0.4	0.45

the static coefficient and utility form should be adapted dynamically to the changing network condition. However, in the utility theory model, changing these parameters in real time manner are not evident.

2.2.2   Fuzzy Logic-Based Approach

Fuzzy Logic (FL) is a method of reasoning that imitates the way of decision making in humans that involve all intermediate possibilities between digital values 1 and 0. FL can be applied into the network selection problem [25]. The network selection procedure based on FL is depicted in Figure 2.4. The main modules in the procedure are described as followings:
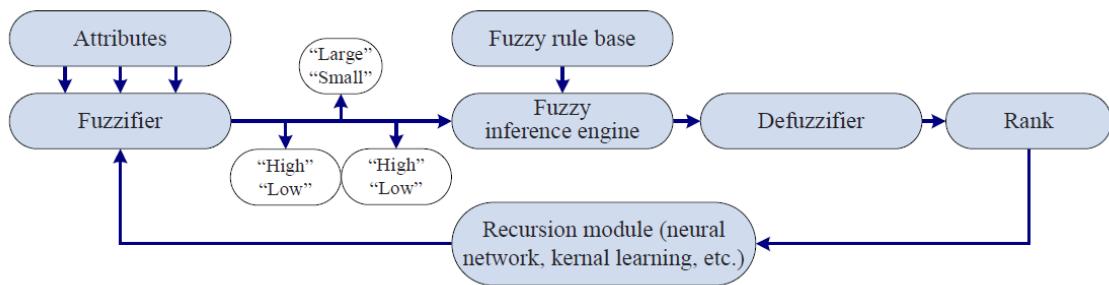
Figure 2.4 – Fuzzy Logic Based Network Selection Procedure, source [14]

- Fuzzy set. Different from classical notion of set, a fuzzy set is a class of objects with a continuum of grades of membership ranging between zero and one. Since in the classical set theory, the membership of elements in a set is assessed in binary terms, which means either belongs or does not belong to the set. By contrast, the fuzzy set theory permits the gradual assessment of membership using a membership function which is valued within $[0, 1]$. The classical set is usually called crisp set in the fuzzy logic theory.

- Fuzzifier. A fuzzifier is a module to map a crisp point into a fuzzy set.

- Fuzzy rule base. The rule base consists of a collection of fuzzy $IF - THEN$ rules.

- Fuzzy inference engine. This module uses fuzzy logic principles to combine the fuzzy $IF - THEN$ rule in the fuzzy rule base.

- Defuzzifier. It is opposite of fuzzifier and this module maps a fuzzy set into a crisp point.

- Membership function. It is used to describe the degree of truth in fuzzy logic theory.

In [26], the authors proposed a fuzzy logic based framework for network selection and the procedure based on the procedure shown in Fig 2.4 by eliminating the recursion part. In their scheme, three input fuzzy variables are considered (i.e., the probability of a short interruption, the failure probability of handover to radio, and the size of unsent messages). In fact, more input fuzzy variables can be considered for other network selection problems. At the beginning of the procedure, the fuzzy variables are fuzzified and converted into fuzzy sets by a singleton fuzzifier. Then, based on the fuzzy rule base, the fuzzy inference engine maps the input fuzzy sets into output fuzzy sets by the algebraic product operation. Finally, the output fuzzy sets are defuzzified into a crisp decision point and generate a network ranking list.

Fuzzy logic can be combined with MADM algorithms; which is called $fuzzy\ MADM$, and this approach is studied in the literature [27] [28] [29] [30]. The idea is to use MADM for the fuzzy interference engine and defuzzifier parts. $fuzzy\ MADM$ is suitable for the case when some attributes are better to be set with fuzziness due to the complex heterogeneous network environment in an MADM scheme.

The complexity is the main limitations of this approach since it is not a trivial task to determine exact fuzzy rules and membership functions. Moreover, extensive testing with hardware is necessary to verify and validate the fuzzy knowledge-based system and stability of the system is another important concern for fuzzy control.

2.2.3   Game Theory-Based Approach

Game theory was originally adopted in economics in order to model the competition between companies which make decisions for maximizing their benefits. It is a mathematical tool for understanding and modeling competitive situations between rational decision makers who have mutual and possibly conflicting interests. Game theory includes the following components:

- Player. A player is the individual who makes the decision. The goal of each player is to maximize his/her own payoff by a choice of strategy.

- Strategy set. It is the set containing all the strategies a player can choose. In each round, the player chooses one strategy from the set.

- Payoff. The payoff represents the utility that a player can receive by playing certain strategy when all the other players' strategies are decided.

- Equilibrium. The players reach an equilibrium means that they find a combination of strategies containing the best strategy for every player.

The principle behind the game theory is that the players seek to maximize their payoffs by choosing strategies depending on the available information at a certain moment. When each player cannot enhance their payoff by changing his strategy while keeping the other players' strategies unchanged, the game is said to have reached a *Nash Equilibrium.*

To apply game theory to network selection problem, mobile users and/or networks become the players in the game. They are seeking to maximize their payoffs by choosing between different strategies such as available bandwidth, subscription plan, or available access points. The payoffs can be evaluated by using utility functions based on various network decision criteria such as monetary cost, energy conservation, network load, availability, etc.

Different categorizations of the various game types are possible. The solutions are classified by the players involved (Users vs. Users, Users vs. Networks, Networks vs. Networks) with a further sub-classification under two broad major game theoretic approaches: cooperative approaches in which implies the joint considerations of the other players and non-cooperative approaches [31] in which each player selects his/her strategy individually.

In this context, game theory is used to model and analyze cooperative or non-cooperative behaviors of users and networks during their interaction in a heterogeneous wireless environment. For example consider a group of users that are located in an area with a number of available networks. Each user is seeking to select the best network that will maximize its utility. Therefore, in [10], the authors pointed out six different game theoretic approaches is proposed:

- Users vs. Users - Non-Cooperative Approach. In this scenarios, users compete against each other while seeking to maximize their own utility. This scenario is presented in [32], in which the behavior of selfish users who compete for access in a WLAN. The authors make use of an evolutionary game-theoretic model to describe the interaction between users. In this game, the mobile users are players and the set of strategies consist of different available transmission rates. To model the payoff for each user, the authors make use of utility function which determines the voice quality received by each user in each state. The wireless characteristics such as delay and loss rate are mapped into the Mean Opinion Score (MOS) which represents a measure for voice quality with the help of utility

function. The authors show that free users with local information can reach an equilibrium which is close to optimal from the system perspective but the equilibrium is very unfair.

- Users vs. Users - Cooperative Approach. In this scenario, users cooperate in order to obtain mutual advantage and maximize the global payoff of the group. In [33] the authors study the problem of bandwidth allocation in wireless networks using the principles of cooperative game theory, such as Nash bargaining and three basic division rules of bankruptcy problems in order to find the optimal solution of the bandwidth allocation problem.

- Networks vs. Users - Non-Cooperative Approach. In this scenario, users compete against networks, each of them is seeking to maximize their own utility without cooperation. On one side, the users try to maximize their benefits from the service for the price they pay. On the other side, the networks try to maximize the profit for the provided services. In [34], the authors consider a Code Dvision Multiple Access (CDMA) based data network. The problem in which the provider and users have a conflicting interest is formulated as s a non-cooperative game. Each of them uses their respective strategies to evaluate the Nash equilibrium.

- Networks vs. Users - Cooperative Approach. In this scenario, users and networks cooperate in order to achieve mutual satisfaction. In [35], the authors studies cooperative interactions between networks and users with their conflicting interest arising due to the network selection problem and seeks appropriate behavior modes for each of the entities. The authors apply the game theory to model, analyze and propose solutions for the network-user interactions to achieve their own satisfaction.

- Networks vs. Networks - Non-Cooperative Approach: In this scenario, the networks compete against each other, seeking to maximize their individual revenues. In [36], a radio resource management framework is proposed to solve the integrated network selection mechanism. The authors model this problem as an interaction game between the service providers and customers in a non-cooperative manner to maximize their rewards. Then the Analytic Hierarchy Process (AHP) approach is integrated with non-cooperative game theory using utility functions into network selection algorithm to decide the best network for the mobile users.

- Networks vs. Networks - Cooperative Approach. In this scenario, networks cooperate in order to achieve global payoff maximization. In [37], the authors make use of bankruptcy game which is a special type of an N-person cooperative game to obtain a bandwidth allocation and admission control algorithm. In this algorithm, different wireless networks cooperate to offer bandwidth to a new connection form a coalition among the networks.

The limitation of the game theory approach is the complexity implementation. Since in the current literature, the proposed solutions were only tested through intensive numerical analysis or simulations in a simplified wireless environment. There is no test-bed implementation in a real system scenario and the efficiency of this approach should be discussed.

2.2.4 Markov Decision Process-Based Approach

Markov Decision Process (MDP) is a common tool for making multi-objective dynamic decisions. MDP model can solve network selection and Vertical Handover (VHO) decision at one time. To apply the MDP into the network selection problem, the mobile users have to choose an action based on its current state at each decision epoch. With this state and action, they evolve to a new state according to a transition probability function. This new state lasts for a period of time until the next decision epoch comes, and then the mobile users make a new decision again. For any action that they choose at each state, there is a reward and a cost associated with it. The goal of each mobile user is to maximize the expected total reward it can obtain during the connection lifetime.

The components in an MDP procedure are described in [38]: (i) A step space $\mathbb{S}$. (ii) Sets $\mathcal{A}(s)$ of available actions at state $s \in \mathbb{S}$. (iii) Transition probabilities $\rho(Y|s,a)$ and reward functions $r(r,a)$ denoting the one-step reward using action a in state s. The above objects indicate a stochastic system with a state space $\mathbb{S}$.

When the system is at state $s \in \mathbb{S}$, a decision maker selects an action a from the set of actions $\mathcal{A}(s)$. After an action a is selected, the system moves to the next states according to the probability distribution $\rho(Y|s,a)$ and the decision-maker collects a one-step reward $r(s,a)$. The selection of an action a may depend on the current state of the system, the current time, and the available information about the history of the system. At each step, the decision maker may select a particular action or, in a more general way, a probability distribution on the set of available $\mathbb{A}(s)$, which are called

non-randomized and randomized decisions, respectively. An MDP is called *discrete* if the state and action sets are discrete, which is the case for network selection.

In the literature, several network selection schemes based on MDP theory have been proposed. In [39], the authors model the network selection problem into an MDP to determine the conditions under which VHO should be performed. The authors formulate the problem as an MDP in order to maximize the total expected reward per connection.

In [40], the authors propose another approach of using MDP for the network selection problem. The states are defined based on the number of users of different services (e.g., voice and data) in different candidate networks. Transitions between states within the Markov chain will occur due to the arrival and departure of voice call or data session. Giving the arrival distributions of voice calls and data sessions, the transition rates between states in the Markov chain will be decided by the network selection policy. The authors showed that this model could be used to evaluate the performance of many types of network selection scheme.

However, MDP approach has its limitations. Since users' handover cost is high, the MDP approach should consider some state information such as the currently used network to avoid to switch the network too often when making the decision. Moreover, the implementation of MDP into the real system is also complicated and the speed for making a decision should be evaluated.

2.3   Network Selection in Practice

2.3.1   ANDSF-Assisted Network Selection

The widespread development of mobile devices with multiple heterogeneous built-in wireless access networks pushes for supporting a network discovery mechanism to select the best network to connect to the Internet. For the theoretical approaches presented above, there is a lack of discussion on how to integrate these algorithms into the standard network systems. Thus, the 3GPP standards introduce a new entity called Access Network Discovery and Selection Function (ANDSF) [41] [42] to assist mobile devices in searching non-3GPP access networks. Its goal is to offer the mobile device information about the neighboring access networks through discovery information. ANDSF can also assist the device in the handover process through rules-based network selection policies. ANDSF enables the interworking of 3GPP (e.g., GSM, UMTS, LTE) and non-3GPP networks (e.g., CDMA, WiFi, WiMAX) so that 3GPP-compliant User

Equipment (UE)s (e.g. smartphones, tablets) can also access non-3GPP data networks. It is also a key enabler for operators to offload data traffic from the mobile network by treating the congestion information of different networks. The deployment of ANDSF within Evolved Packet Core (EPC) network is showed in Figure 2.5. In this architecture, ANDSF can provide two functionalities: (i) Network discovery and selection, which transmits to the UE a list of available access nodes with the information such as access type, network ID, etc; (ii) Handover decision, which can transmit Handover indications and trigger HO for UEs.
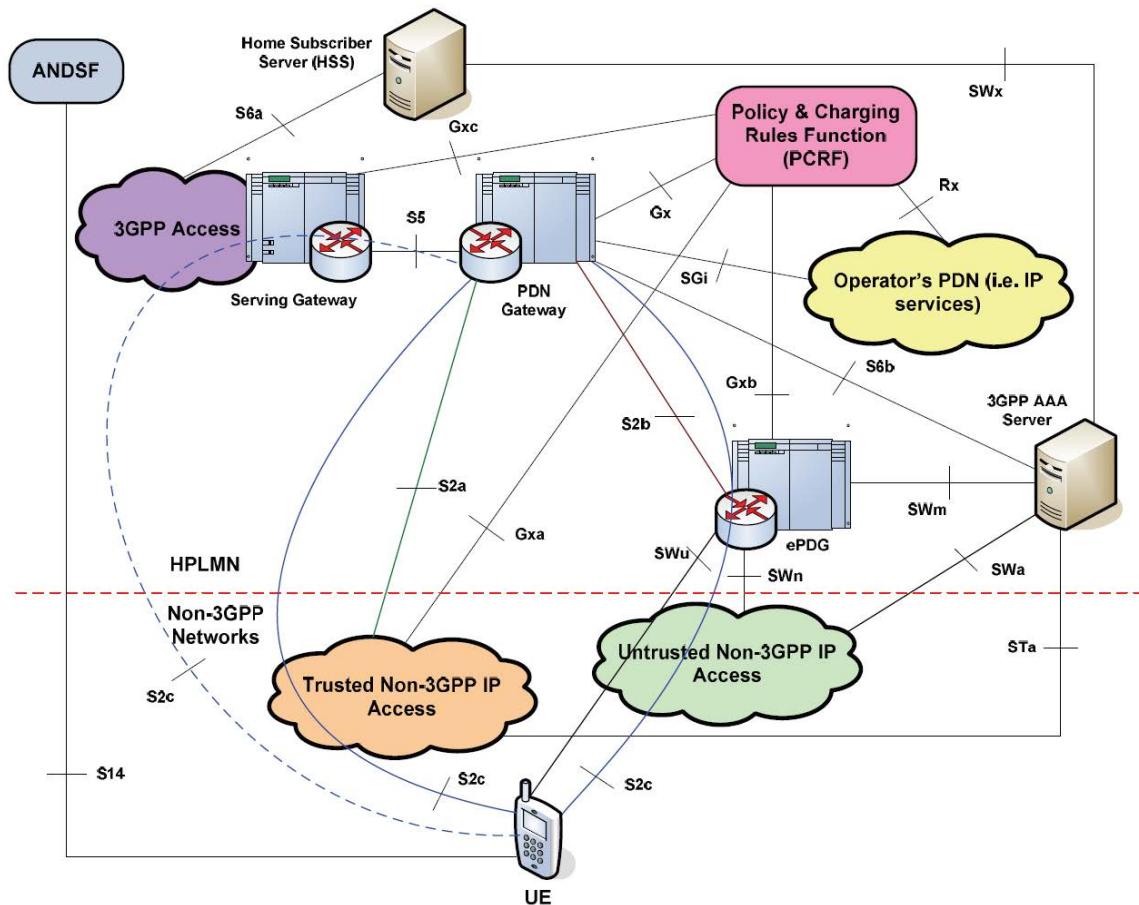
Figure 2.5 – ANDSF integration in EPC architecture. Source [43]

Figure 2.6 illustrates how the UE can communicate with the ANDSF entity based on a simple client-server architecture. The ANDSF is located in either the home Public Land Mobile Network (PLMN) (H-ANDSF) or the visited PLMN of the UE (V-ANDSF) when the UE is in roaming. UEs can interact with the ANDSF in two modes over the $S14$ interface:
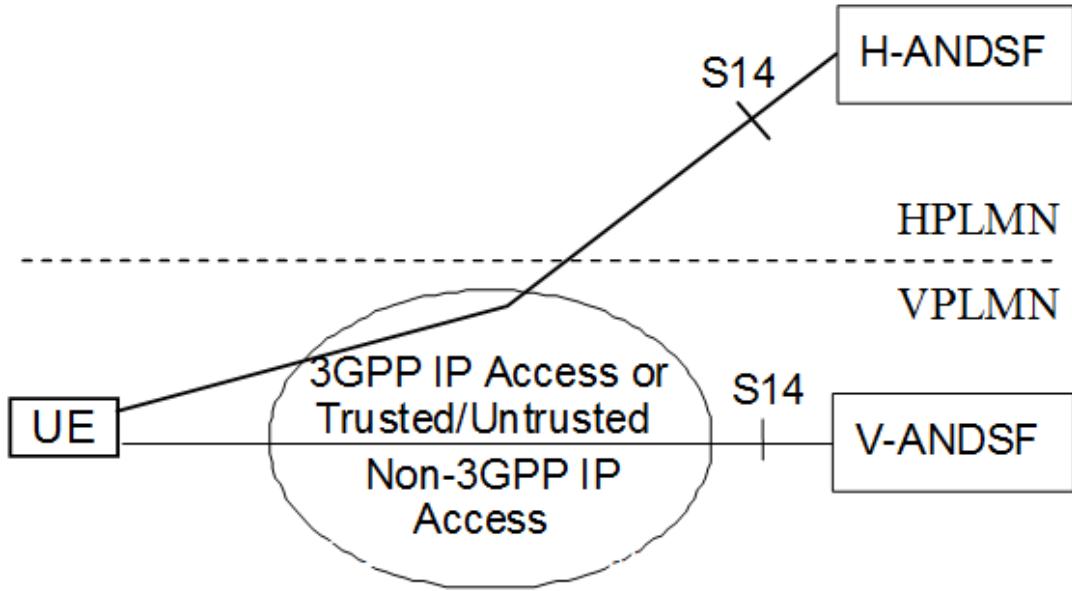
Figure 2.6 – ANDSF architecture

- Pull Mode. In this mode, UE contacts ANDSF to request policy information UE interprets and acts on the policy.

- Push mode. In this mode, Policy is pushed to UE (SMS is the most common method). Alternatively, it may result in UE contacting ANDSF for more information (i.e. Push may be just a Pull trigger).

An ANDSF can provide the following information to a UE, which are also called policies, based on operators' configuration:

- Inter-System Mobility Policy (ISMP): It guides the selection decision for devices with single links network selections rules for a UE with no more than one active access network connection (e.g., either LTE or Wi-Fi).

- Inter-System Routing Policy (ISRP): It directs the distribution of traffic for devices with multiple simultaneous links network selection rules for a UE which has potentially more than one active access network connection (e.g., both LTE and Wi-Fi).

Some research work focuses on improving network selection procedure based on ANDSF. In [44], the authors propose an ANDSF-assisted energy-efficient vertical handover decision algorithm for heterogeneous networks. The proposed algorithm makes mobile devices select the network and minimize the power consumption while

maintaining the Quality of Service (QoS) of ongoing sessions. A novel ANDSF-assisted WiFi control method is presented in [45] to avoid unnecessary WiFi scanning and connections when performing the data offloading from the mobile network. In [46], the authors propose messages and procedures for ANDSF to collect the congestion information among heterogeneous networks and to decide inter-system mobility and routing policy. However, these propositions have little control on the network so that they cannot respond to the frequent change of network conditions. In [47] the authors propose an enhancement scheme to the ANDSF by incorporating a Network Event Reporting Function (NERF) at access network level. The scheme can provide the best access network for data offloading. A multi-attribute decision Multiple Attribute Decision Making (MADM) module is implemented in the ANDSF by taking the dynamic characteristics of the access networks as input in order to update the statically configured mobility and routing policies provided to the UE for access selection.

However, the information sent by ANDSF to mobile devices is static and does not state any information such as the network utilization in the current discovery. Therefore, the existing ANDSF deployment cannot guarantee that mobile devices can always connect to the best available network, even they will select the unavailable or inappropriate access network.

2.3.2 Mobile Data Offloading Solutions

Mobile Data Offloading, often called Wi-Fi offload is a network selection solution to enable transfer of some traffic from the cellular network to complementary network technologies such as Wi-Fi network. As both mobile subscribers and devices have the intention to connect to Wi-Fi whenever it is its coverage, mobile operators need to follow their subscribers into the Wi-Fi environment about their quality of service. Hence the business case for Wi-Fi offload is both about saving on costs and reducing QoS instability. Therefore, commercial Wi-Fi offloading solutions are developed to respond to this need of operators. Openet's Network Selection Intelligence solution [48] enables operators to automatically connect their customers to defined mobile and Wi-Fi networks and make intelligent offload decisions through integration with policy and charging control systems. The solution is based on client interaction with a fully standards compliant ANDSF server contained within Openet's Interaction Gateway. The Aptilo Wi-Fi Offload solution [49] offloads the whole 3G/4G network including the mobile core as well as the different 3GPP standards for backhauling Wi-Fi data back to the mobile core. The solution supports the integration with the mobile core for policy and charging. Aruba's carrier-class Wi-Fi solution [50] provides the operators

complete visibility into Wi-Fi network and the solution offers control capability over Wi-Fi network for successfully deploying and operating public Wi-Fi hotspots and hot zones to offload their 3G/4G networks.

Another notable development in the context of Wi-Fi offloading is the IEEE 802.11u standard [51] that enables mobile devices to select an appropriate Wi-Fi access point before actually associating with it. When the mobile device has no way of determining the best access point in locations where several Wi-Fi access points exist, the standard can provide a much richer information to assist mobile devices in making the network selection. The information includes the type of network (private, public, or paid public, etc.), the current load (which can give a hint about the level of congestion on the access point), or details about the roaming agreements with other service providers etc. Moreover, Access network Query Protocol (ANQP) is introduced in 802.11u to allow a mobile device to run a query/response session with the access point before associating with it. The Wi-Fi Certified Passpoint program, also known as Hotspot 2.0 is based on the IEEE 802.11u standard. Hotspot2.0 allows mobile devices to automatically join a Wi-Fi subscriber service whenever the user enters a Hotspot 2.0 area for providing better bandwidth and services-on-demand to end-users and reliving carrier infrastructure of some traffic.

2.4   Summary

Network selection is critical for mobile users in the heterogeneous network environment to get an Always Best Connection experience. However, this is not a trivial task since mobile users should consider various network statistics and the implementation complexity. In this chapter, the mathematical models and practical solutions for solving the network selection issue were discussed. Firstly we presented four mathematical approaches: utility theory, fuzzy logic, game theory, Markov Decision Process (MDP), which described the network selection decision-making model. However, these models were not dynamic to the changing network conditions and often very complex to implement. Then we discussed the industrial solution and standardization effort to tackle the problem from a practical point of view. In this thesis, we will propose a novel network selection model based on the control theory and discuss its implementation in a real network architecture.

Chapter 3

HTTP-based Adaptive Streaming: History, Issue and Related Work

3.1 Background of HTTP-based Adaptive Streaming

Video content delivery over the Internet was firstly introduced in the 1990s. At that time, the Internet Engineering Task Force (IETF) designed Real-Time Transport Protocol (RTP) to define packets formats of audio and video content for delivering over IP networks. However, RTP is often blocked by firewalls. With the evolution of network architecture, RTP is not suitable anymore for multimedia content delivery over today's Internet. With the emergence of Content Delivery Networks (CDN), Hypertext Transfer Protocol (HTTP) becomes commonly the used protocol to deliver multimedia packets in larger segments, which is called HTTP-based streaming since HTTP could easily pass firewalls or Network Address Translation (NAT) gateways. Moreover, HTTP-based streaming is built on simple, commodity servers and compatible with content caching capability. Therefore, HTTP streaming can provision a large number of streaming users at the same time without having to maintain a sessions state on the HTTP server.

Initially, HTTP-based video protocols required downloading the complete video before it could be played. However, this approach has many shortcomings. One of the issues is that all clients receive the same encoding of the video, they are not flexible to face the congestion in the network, which leads to frame freezes and stuttered playback. Therefore, a novel approach called HTTP Adaptive Streaming (HAS) is developed to

cope with the changing underlying network conditions to provide good performance in terms of Quality of Experience (QoE).

HAS has different versions of implementations. Popular proprietary video streaming applications such as Apple's HTTP Live Streaming [53], Microsoft's Silverlight Smooth Streaming [52], Adobe's HTTP Dynamic Streaming [54] develop their own protocols based on HAS principle. However, since these platforms are proprietary and the implementation is different with each other, the receiver of devices should support its corresponding client protocol, which adds the complexity on devices. Therefore, Moving Picture Experts Group (MPEG) issued a call for proposal for an HTTP streaming standard and finally Dynamic Adaptive Streaming over HTTP (DASH), also known as MPEG-DASH [55], is adopted and the specification is published as ISO/IEC 23009-1 to define a converged format for video streaming. DASH is also adopted by 3GPP Release 10 (3GP-DASH [56]) for the use over wireless networks. Although differences exist between these implementations they are all based on the same basic HAS principles: a video is split up into several segments which are encoded at different quality rates, then the intelligent video clients dynamically adapt the quality, based on metrics such as average throughput, delay, and jitter, etc. A manifest file (often referred to as a playlist) provides a list of available bitrates and for users to choose. In this thesis, we focus on the standard MPEG-DASH since it is a vendor-agnostic design, and its implementations will be adopted on a wide range of devices and operating systems in the future.

In DASH standard, video content is divided into segments, also called chunks. Each segment is encoded with different quality levels (i.e., bitrates) and has fixed duration. All segments are stored at servers of content providers. Each video user can request the segment at the most appropriate quality level according to the estimated bandwidth. Thus, when a network degradation (e.g., reduced available bandwidth, increased latency) is identified by the video client, the latter has the ability to switch to a lower video quality level (and, thus, lower bitrate) in order to reduce the risk of playout interruptions, which are known to be the most harmful for the perceived video quality [58]. Similarly, improved conditions allow switching to higher-quality levels. In this way, video playback dynamically changes its quality depending on the available resources, which can guarantee optimized resource management and a smoother video streaming experience for users in the presence of variable network conditions. The ability for bitrate/quality switching is given by the Media Presentation Description (MPD) file, which is downloaded by the users from the DASH server.

Figure 3.1 shows the MPD hierarchical data model. The MPD file, also called the manifest file, is an eXtensible Markup Language (XML) document which includes information on the available video representations. The MPD file firstly consists of one or multiple periods which indicate the starting time and duration of the video. Then each period consists of one or multiple adaptation sets, which might contain the different bitrates of the video or audio component of the same multimedia content. Each adaptation set provides multiple representations, which presents an encoded alternative of the same media component. Each representation indicates the specific bandwidth to transmit the video of the corresponding bitrate, resolution, or other characteristics. video chunks are contained in the representation and each chunk has an addressable location on a server presented by a Uniform Resource Identifier (URI), Thus, users can downloaded video chunks from the video servers via HTTP according to its corresponding URI.
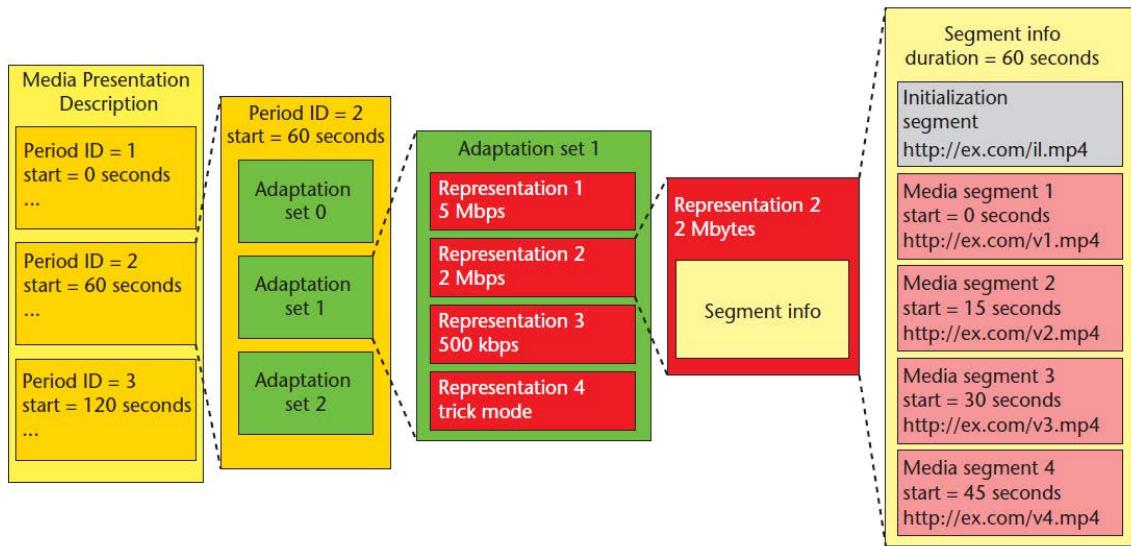
Figure 3.1 – MPD Hierarchical Data Model, source [57].

Figure 3.2 illustrates a DASH streaming scenario with the use of the MPD file between a DASH client and an HTTP server. Before playing the content, the DASH client fetches firstly the MPD file then parses it with the MPD parser installed in the DASH client's player to get the information about the video characteristics. Then the DASH client selects the appropriate encoded alternative among the set of representation proposed in MPD file. The video player then starts streaming the content by downloading the segments stored on an HTTP server using HTTP GET requests. The client starts to buffer the video chunks into its player and at the same

time, the client monitors the network bandwidth and throughput variations to decide how to adapt to the available bandwidth by requesting segments of different bitrate alternatives to maintain an adequate buffer.
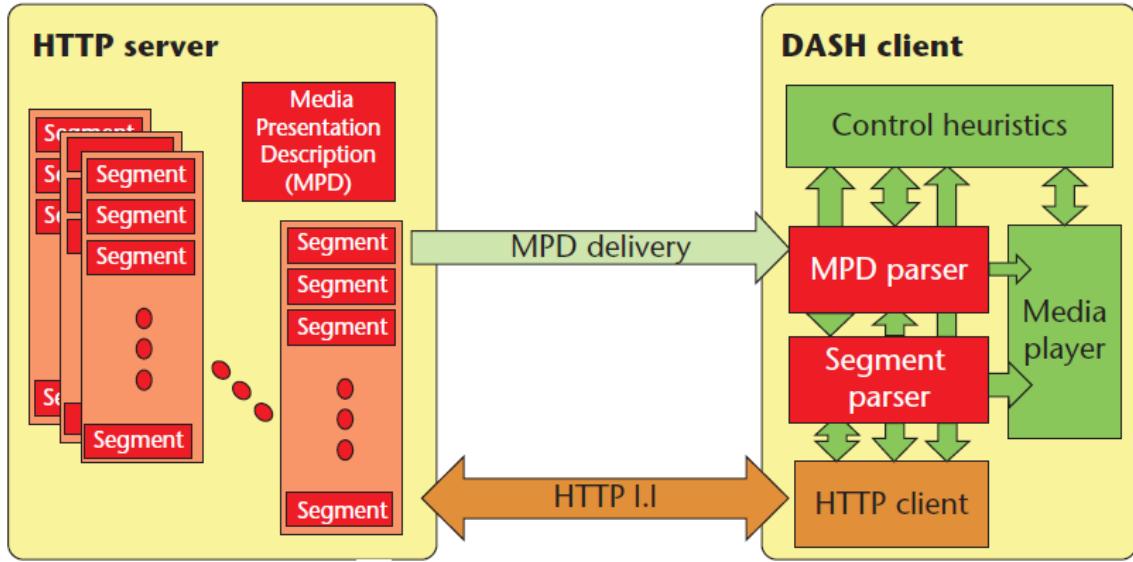
Figure 3.2 – A Simple Streaming Scenario Between an HTTP Server and a DASH Client, source [57].

3.2 Current Issues in HTTP-based Adaptive Streaming

The drawback of HAS approach is that the quality assignment process is fully dominated by the clients. The MPEG-DASH specification only defines formats of the MPD file and the segment, however, the client behavior for fetching the video segment and the rate adaptation schemes are not considered by MPEG-DASH. Therefore, current HAS implementations have its intrinsic shortcomings due to the greedy nature of clients.

In [59], the authors present an analysis of the performance and drawbacks on three commercially adaptive HTTP streaming players: Microsoft's Smooth Streaming, Netflix, and Adobe players. The experiments show that current heuristics rate adaptation schemes perform the sub-optimally quality selection. Drops in the client playout buffer and unnecessary quality switches often occur since they fail to adapt to rapid bandwidth changes. They also analyze the performance of two competing adaptive

HTTP streaming clients sharing the same bottleneck and point out that the players do not aim to reduce unfairness in bandwidth sharing. In [60], the authors conducted experiments on three popular video streaming service: Hulu, Netflix, and Vudu. The results indicate it is hard to estimate accurately client-side bandwidth above the HTTP layer. As a result, rate selection based on inaccurate estimates can trigger a feedback loop and the clients will get undesirably variable and low-quality video. In [61], the authors discuss the problems of instability, unfairness, and bandwidth under-utilization caused by the competition for available bandwidth between multiple adaptive streaming players in adaptive HTTP streaming architecture. Then they analyze the underlying root cause of the performance problems that take place when multiple adaptive streaming players compete. In [62], the author show that when multiple adaptive HTTP streaming clients compete at a network bottleneck, the clients can hardly perceive its fair shared bandwidth. The conducted experiments indicate that this fundamental limitation will cause video bitrate oscillation, which degrades the video viewing experience. The streaming service problem is even more complex in a wireless and mobile environments due to the high dynamicity caused by time-varying fading, shadowing interference and hand-off delays. In [63], the authors evaluated and analyze the performance of adaptive HTTP streaming commercial solutions in a real-world mobile network condition. In [64] the authors investigate the effect fluctuating bandwidth on adaptive HTTP streaming clients over a mobile 3G network. The experiment shows that different players schedulers have its own strategy, which makes significant differences in performance and optimization goals. Therefore, optimizing adaptive HTTP streaming in a mobile environment become a challenging task in research in industry community in the coming 5G era. In [65], the authors analyze the scenario in which multiple concurrent video streaming traffic of two or more video players competing for the shared bandwidth. The experiments show the unfairness in sharing the bandwidth among the players, and its impact on the QoE.

3.3 HTTP-based Adaptive Performance Improvement: the State of the Art

From the previous sections, it is obvious that providing seamless multimedia experience for the HTTP-based adaptive streaming service is still challenging. The improvement strategy should be further investigated under the time-varying network bandwidth condition, especially in a mobile environment in order to optimally use the available end-to-end network resources. In the literature, the current research work can be

mainly categorized into three approaches: client-based approach, proxy-based approach and SDN-based approaches.

3.3.1   Client-based Approach

The main idea behind the client-based approach is that the rate adaptation algorithm runs at the client's side. The algorithm is designed based on the information such as the clients' player buffer state, estimation of the throughput and the available bandwidth to choose the appropriate downloaded video quality.

In [67], the authors present a receiver-driven rate adaptation algorithm for HTTP adaptive streaming. A smooth HTTP/Transmission Control Protocol (TCP) throughput measurement method is proposed to decide switch-up or switch-down operations between different representations. In [68], the authors propose a novel client rate adaptation algorithm based on a Proportional Integral Controller (PIC) in order to stabilize the playout buffer at a reference level. As a result, the algorithm can implicitly match the video rate by making clients share the bandwidth fairly in the long run of video streaming service. In [69], the authors proposed a client-side controller to optimize the users' perceived quality. The controller computes a set of strategies by employing the Markov Decision Process (MDP) approach for each bandwidth segment in order to maximize QoE of clients. In [70], the authors propose a novel cross-layer media-buffer aware optimization framework for wireless resource allocation by constraining rebuffering probability for adaptive streaming clients. Re-buffering Aware Gradient Algorithm (RAGA) is used to solve this optimization problem, which relies on a simple periodic feedback of media buffer levels returned by adaptive streaming clients. In [71], the authors formulate the video bit rate adaptation as a stochastic optimal control problem. Then a Model Predictive Control (MPC) scheme is designed by combining buffer occupancy and throughput predictions to maximize the clients' QoE. In [72], the authors propose a novel video rate control scheme that balance the needs for the responsiveness and video rate smoothness in Dynamic Adaptive Streaming over HTTP (DASH). The algorithm uses client-side buffered video time as the feedback signal to design a controller. With the controller, the DASH system can smoothly increase video rate as the available network bandwidth increases, and promptly reduce video rate in response to sudden congestion level shift-ups. In [73], the authors propose a novel receiver buffer based rate adaptation scheme for HTTP streaming using control theory. This scheme is formulated as a proportional controller problem and a Proportional Derivative (PD) controller is designed to prevent the receiver buffer from either overflow or underflow.

The client-based side approach has its own limitations. The main problem is that in a purely client-based optimization strategy, there is no coordination among the clients, which will entail the fairness problem since different clients negatively influence each other as they compete for the same bandwidth. Moreover, the providers and operators has no control over the provided video quality with this approach, which is essential when offering a managed service.

3.3.2   Proxy-Based Approach

The principle of the proxy-based approach is to place intermediary nodes between HAS clients and servers to monitor and collect network information such as the available bandwidth, network bottleneck, link conditions etc. The proxy can be combined with client-side rate adaptation scheme to deliver optimal video service in a multi-client environment. The proxy functionality can be implemented via software so that it can be placed liberally on network nodes. For example, a coordination proxy function can be automatically activated when the traffic exceeds a certain percentage of the link capacity to execute the optimization algorithm.

In [74], the authors propose a coordination proxy between HAS server and clients to collect information on the overall system performance. This information is then provided to HAS clients to run a reinforcement learning-based quality selection algorithm to achieve fairness in a multi-client environment. In [75], the authors propose a QoE management framework for Video on Demand (VoD) HTTP adaptive streaming. Rate adaptation algorithm such as centralized, distributed and heuristic algorithms can run on the intermediary proxies to enable the network providers to control the quality selection at the client's side. In [76], the authors propose to deploy an intelligent node between HAS servers and clients to periodically solves an optimization problem to determine the maximum quality the clients can download, based on the current network status. In [77] the authors propose a wireless DASH (WiDASH) proxy located between the Internet and the wireless networks to enhance the QoE of wireless DASH clients. This proxy is transparent to HTTP streaming server on the Internet and it is responsible for rate adaptation algorithm based on multiple-input-multiple-output adaptive optimal control. In [78], the authors implement an experimental framework using open source Squid proxies [79] and the most recent Open Source Media Framework (OSMF). With this framework, a range of different policy classes may be employed in the proxy and the performance perceived by clients can be evaluated while controlling the network conditions and protocol parameters. In [80], the authors implement a network-based bandwidth manager to determine the quality levels to assign to the clients based on the

available bandwidth and their subscription level. The congestion-aware rate adaptation scheme over the top of existing segmented delivery protocol can ensure uniform QoE across multiple clients. In [81], the authors propose a QoE-aware DASH system, named QDASH to improve the user-perceived quality of video watching. QDASH consists of two building blocks: QDASH-abw and QDASH-qoe. The QDASH-abw module measures the network available bandwidth by integrating available bandwidth measurement into the video data probes with a measurement proxy architecture, which facilitates the selection of video quality levels. The QDASH-qoe module determines the video quality levels with a QoE-aware quality adaptation algorithm. In [82], the authors propose a proactive QoE-driven multi-user DASH scheme that enhances the adaptive HTTP media delivery to multiple clients in a wireless cell. A QoE proxy is developed to intercept and rewrite the client HTTP requests with the optimal rate based on the wireless channel conditions. In [83], the authors proposed a novel client-based rate adaptation algorithm called FINEAS (Fair In-Network Enhanced Adaptive Streaming) to optimize the QoE perceived by clients and achieve fairness in a multi-client setting. The algorithm is together with a system of coordination proxies which is in charge of collecting measurement on the network conditions and provide information to clients in order to facilitate fair resource sharing among clients. In [86], the authors employed feedback control theory to design a QAC (Quality Adaptation Controller) for live adaptive video streaming. With the QAC controller, the video quality is throttled to match the available bandwidth with a transient of less than 30 seconds while ensuring a continuous video reproduction. Moreover, QAC fairly shares the available bandwidth both in the cases of a concurrent TCP greedy connection or a concurrent video streaming. In [84, 85], the authors propose a video proxy manager at the eNodeB level and a controller of resources at the base station node for HTTP Adaptive Streaming. The manager can select the appropriate channel bandwidth and video quality levels at which to transmit the video chunks. However, the bandwidth allocation scheme considers the Channel Quality Indicator (CQI) reports received from the mobile users, which may significantly increase the signaling overhead.

The aforementioned algorithms present the advantages of using proxy node to facility the QoE improvement for HAS service. However, the proxy-based approach presents robustness issues in case of fault or malfunctioning of the proxy elements. If the proxy node fails to work correctly, the rate adaptation algorithm will be run only on the client-side, which brings sub-optimal behavior. Moreover, the proxy-based approach increases the system implementation complexity in the real scenario.

3.3.3 SDN-based Approach

Software Defined Networking (SDN) [87] is a novel networking approach that facilitates the decoupling of the control plane in a network (i.e. the decision making entity) from the data plane (i.e. the underlying forwarding nodes). Typically, the control plane decisions are taken at a central controller that has a global network view by monitoring and collecting the status of the whole network. Monitoring and control traffic is communicated using a secure interface, which known as the southbound interface, between the controller and forwarding nodes. The OpenFlow protocol [88] is considered the most widely supported protocol in the southbound. On top of the controller, a diverse range of networking applications are deployed to perform different functions including, but not limited to, routing, security, load balancing, and network optimization at the run time.

SDN is a promising network management standard in future multimedia networks since its flexibility offers different options for improving the streaming quality. Software Defined Networking (SDN) allows the dynamic adjustment of forwarding tables to reroute different flows so that different flows can be treated differently. For example, the SDN controller can assign different routes for video flows to satisfy the pre-defined quality of service rules. Additionally, SDN also supports QoS functions such as traffic shaping that limits the amount of network resources consumed by different streaming flows in the network to achieve streaming delivery fairness. Therefore, SDN may offer new opportunities for better coordination, and finer granularity, between the video streaming service and resource management control loops.

In [89], the authors propose an OpenFlow-assisted QFF (QoE Fairness Framework) to achieve user-level fairness in an adaptive video streaming environment. In this framework, a centralized OpenFlow controller monitors the status of all the DASH video applications and allocates dynamically the bandwidth for each streaming device by adjusting video flow characteristics to ensure network-wide QoE fairness. In [90], the authors evaluate the performance of DASH considering different metrics, such as fairness, efficiency, and quality with traffic shaping techniques in an SDN environment. The results show that traffic shaping could reduce the number of stalls and quality switches when the clients share the same bottleneck link. In [91], the authors propose an SDN-based architecture for improving the performance of scalable video streaming using DASH. The main idea is to exploit path diversity when streaming the different video layers. For improved scalability, the authors propose a modification of the MPD file to include the port to be used for each video layers. In [92], the authors propose a buffer-aware HTTP live streaming approach for SDN-enabled Fifth Generation (5G)

wireless networks, which evaluates the "weight" of each video chunk to determine its transmitting priority, while arranging the transmission path according to link utilization and router/switch stability. In [93] the authors propose an algorithm to dynamically calculate the optimal delivery paths of Scalable Video Coding (SVC)-based video streaming in real time for the different video layers using SDN paradigm. The forwarding tables of routers and switches are updated based on the monitored information of the network links, which allows transferring the different video layers over distinct network paths to optimize the delivery of video over congested networks.

However, since SDN is based on a centralized controller to manage the network, the scalability of this kind of approach is limited when the number of clients increases. Moreover, the resilience of such centralized architecture and the complexity of integrating these mechanisms into the legacy mobile core network should be evaluated.

3.4 Summary

Today's Internet is dominated by video traffic. Video delivery protocol has evolved from Real-Time Transport Protocol (RTP) to HTTP Adaptive Streaming (HAS). In this chapter, we firstly presented the background of HTTP-based adaption streaming. Then the current issues of HAS were discussed. Due to the intrinsic disadvantages of HAS protocol, clients often suffered QoE degradation under changing network conditions. Lots of research work were made to tackle this issue. We presented from three approaches: Client-based approach, Proxy-based approach, and SDN-based approach. However, these approaches had limitations such as lack of cooperation among clients, implement complexity or scalability issues. In this thesis, we will leverage the emerging architecture, Multi-access Edge Computing (MEC) to improve the DASH-base streaming deliver and guarantee the seamless QoE of clients in the heterogeneous network environments.

Chapter 4

Control Theory-based Network Selection Scheme in Heterogeneous Network

4.1 Introduction

The recent evolution of terminals, with the multiplication of built-in access interfaces to heterogeneous networks, has refocused the interest of researchers from both industry and academia to the problem of optimal selection of access networks. Access network selection schemes can be decided from two sides. From the users' perspective, the focus is mainly to maximize their own Quality of Experience (QoE). This presents, however, several limitations as the users are only aware of the link quality, and have no information on the level of the network usage, which could clearly induce instability in the perceived quality and thus frequent handovers. Accordingly, a simple interface selection mechanism from just the users' side is not sufficient to reach the optimal network resource utilization. From Network Operator (NO)s' perspective, the objective is to maximize the revenues and the overall users' satisfaction by avoiding network congestion and by selecting the optimum interface for each user. This comes, though, with an increased signaling overhead. Network interface selection problems were widely studied from both academia and industry, which is presented in Chapter 2. However, the mathematical models cannot respond to the rapid changing of the network parameters. Operators or tiers developed connection manager applications, running on mobile devices to assist users to choose the authenticated Wi-Fi networks automatically. However, since different access networks have distinctive characteristics

(e.g. spectrum range, network capacity or management systems), the interface selection problem becomes a multi-criteria decision-making problem for users who desire to be Always Best Connection (ABC) considering various network characteristics. Moreover, frequent scanning for Wi-Fi Access Points should be also avoided to reduce the energy consumption of mobile devices when making interface selection decisions. Consequently, a more flexible and intelligent network selection scheme should be conceived to ease the congestion and adapt to variable network workloads.

The purpose of this chapter is to address the network selection problems for different traffic classes in the context of heterogeneous access networks. We advocate thus for cooperative actions in which the NOs provide general rules consisting in network access probabilities, and the users select the best interface based on their own preferences and the policies of NOs. Firstly, we introduce a simple model to describe the network selection procedure for in a Long Term Evolution (LTE) and Wi-Fi interworking for the scenario of one traffic class. Then we design a Linear–Quadratic Regulator (LQR) controller and make analytical validation with MATLAB. Based on the same principal we extend the simple model with a general analytic model describing the network selection procedure in a multi-class traffic context. We consider, then, the design of an optimized controller for the access probabilities' calculation, which is augmented with an admission control functionality. In contrast with existing mechanisms, which generally rely on heuristics, the proposed mechanism allows computing network access probabilities based on linear discrete-time optimal control theory. We discuss the advantages of a seamless integration with the Access Network Discovery and Selection Function (ANDSF). A prototype is also implemented into Network Simulator 3 (ns-3) to prove that our controller can be applied to the real network conditions.

4.2   Problem analysis and methodology

As mentioned in Section 4.1, the goal of the designed controller is to optimize the network resources' utilization by assisting mobile devices to make the access network selection decision. Indeed, the network utilization can have a direct impact on the network performances. It can be indicated by the number of running sessions in this network. Since the number of running sessions in each network is related to the network selection strategy, we implement an LTE and Wi-Fi interworking prototype using the Network Simulator 3 (ns-3) [94] to explore how the network selection schemes impact the network resource utilization and the performances. The simulator uses a single eNodeB for the LTE access network and a Wi-Fi access point.

To measure the congestion level, we considered, in the following, the end-to-end delay for Wi-Fi access networks and the packets' loss ratio for LTE networks. The Wi-Fi link layer represents, indeed, a reliable media for packets' transmission in which the communications may experience increased access delays due to the contention at the access and due to packets' collisions (i.e. retransmission). In case of congestion, the communications in LTE networks don't really experience significantly increased delays, thanks to the deterministic access technology. However, when the congestion occurs, the communications endure packets' losses due to buffer overflows.

(a) Delay in Wi-Fi   (b) Packets' Loss Ratio in LTE
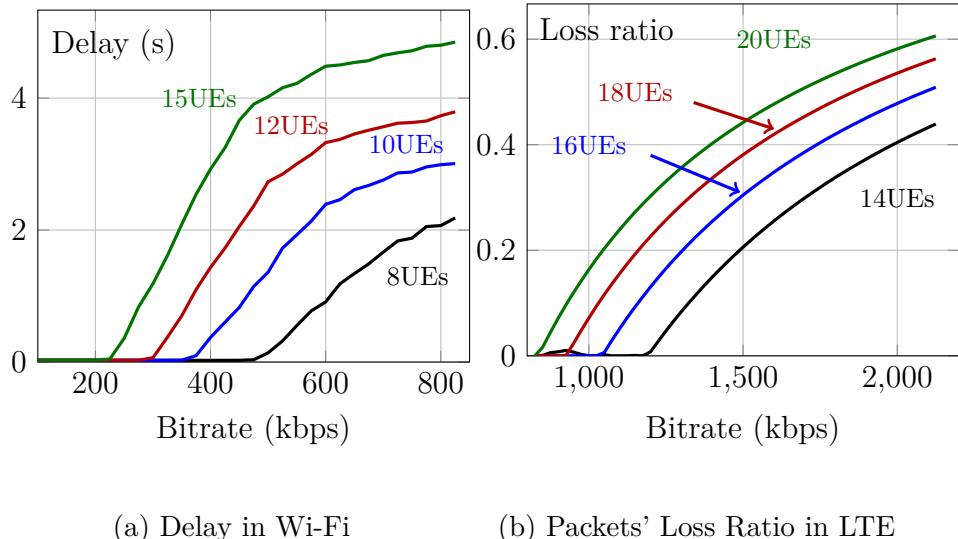
Figure 4.1 – Performance Evaluation for LTE and Wi-Fi Interworking

Figure 4.1 presents the performance evaluation for LTE and Wi-Fi interworking. The interface selection strategy in the simulation is to attach to the access network according to a simple recommendation of a connections' manager. We vary the number of sessions and increase the mobile application bitrate gradually. Figure 4.1a shows that the end-to-end packet transmission delay is becoming intolerable (i.e. $> 1s$) when the number of sessions and the bitrate are increasing in the Wi-Fi network, which means that the network congestion occurs. Similarly, the packets' loss ratio during connections gets higher when there are more sessions, which select to attach to the LTE network, as it can be seen in Figure 4.1b. The results show clearly the impact of an inappropriate network interface selection's mechanism on the network congestion.

These simulations also unveil the importance of avoiding oversimplified mechanisms such as the ones based only on the recommendation of connections manager. Instead, collaborative mechanisms are needed to ensure a good quality for the supported services. We can also observe that each network has an ideal network workload to reach the

optimal network resource utilization. Thus, the proposed method consists in (1) determining the best workload for each network (i.e. the targeted network occupancy) and (2) integrating an admission controller to regulate the number of sessions in each network to converge the workload to that occupancy. For simplicity, the targeted network occupancy is defined by ourselves for performance evaluation purposes. In real operator network environments, the network operators need to make a campaign of measures to define the best target.

Therefore, the methodology consists of the following steps. Firstly, we should select the ideal workloads for each network. Then, we build the model corresponding to the network scenario. Finally, we design an intelligent controller, which can deliver the optimal network selection strategy to terminals.

4.3 Fundamentals of control theory

4.3.1 Control theory fundamentals

Control theory is originally designed to influence the behavior of dynamical systems with the objective of keeping the output of the systems following a desired control signal, also called a reference. The control theory is relied on the feedback, which is described by the difference between actual output and the desired reference. The central idea to achieve this objective is to establish a feedback loop, which is illustrated in Figure 4.2. These elements are described as follows:
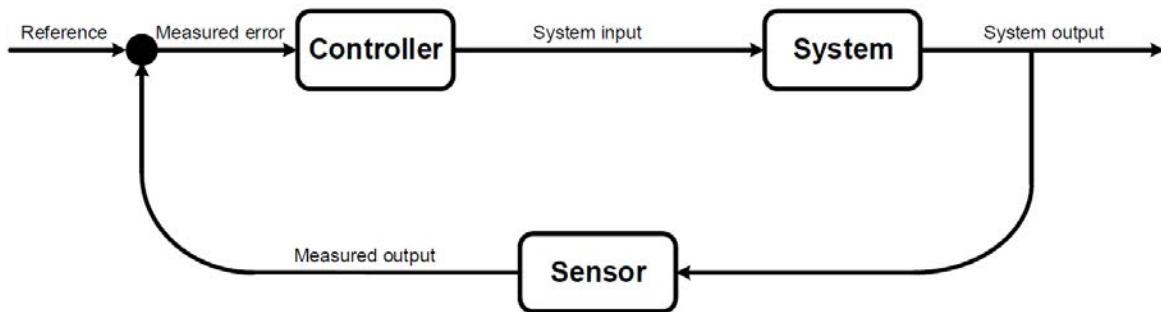
Figure 4.2 – Diagram of a Feedback Control System, source [97].

- The reference is the value of the desired output.

- The measured error which is also called feedback, is the difference between the reference input and the measured output.

- The system input is the setting of one or more parameters, which can be can be adjusted dynamically, to manipulate the behavior of the controlled systems.

- The controller determines the setting of the control input needed to achieve the reference input. The controller computes values of the system input based on current and past values of measured error.

- The system output is a measurable characteristic of the target system.

- The sensor transforms the measured output so that it can be compared with the reference.

The key element in the feedback loop is the controller to achieve the desired output. The controller monitors the feedback of the system, affects the system output and brings the actual output closer to the reference.

When analyzing a dynamic system for designing a controller, the **stability**, the **Controllability** and the **Observability** of the system are the main characteristics which should be analyzed [99]. They are qualitative properties of control systems and are of particular importance in control theory. **Stabality** indicates whether the output will converge to the reference value or oscillate about it. **Controllability** describes the ability of an external input to move the output from any initial condition to any final condition in a finite time interval. **Observability** means that how well the current state of a system can be determined in finite time from its external outputs. If the system is controllable, then we are able to design a controller.

Control theory is widely used in many fields such as mechanical engineering, electrical engineering, and economics to analyze and design feedback loops. Recently, control theory has been adopted in the design of computing systems and networks to achieve control objectives by taking resource actions such as adjusting network bandwidth allocations, memory allocations and scheduling priorities. For example, in data networks control theory has been applied to flow control [95] and to the design of new versions of TCP/IP [96].

To the best our knowledge, little work has been done for the investigation on applying control theory to network selection. This gap drives us to use a control theoretical approach to design a network selection controller. This advantage of this approach is that it which can be not only mathematically analyzed but also experimentally tested.

4.4   Network selection model in LTE and Wi-Fi interworking

4.4.1   Model Description
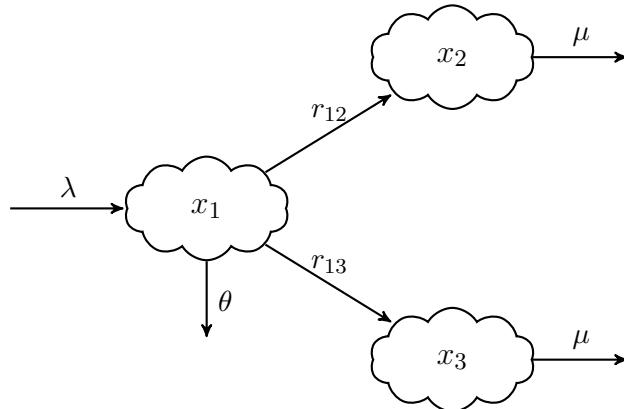
Figure 4.3 – System model

As depicted in Figure 4.3, we describe the network selection procedure of mobile devices in LTE and Wi-Fi interworking. We assume that mobile devices can have both access to LTE and Wi-Fi network. Thus, this model consists of the following three states[1]:

- x_1: represents the state where the mobile devices have already finished the attachment procedure, and they can start the data connection following the strategy defined by our approach. The waiting time in state x_1 corresponds to the time necessary to make the network selection decision to steer the traffic flow of devices to LTE or WiFi network.

- x_2: represents the state where mobile devices are using LTE network for the data communication. When a session of a particular mobile device is terminated, the mobile leaves the system.

- x_3: represents the state where the mobile devices are using WiFi network for the data communication.

Now we use the following variables to capture all the characteristics of our model:

---

[1]The concept proposed considers the mobile devices network selection but can be generalized without any change in the concept of sessions.

- $x_1(k)$ is the number of mobile devices in LTE-WiFi interworking which communicate with the controller to get the network selection information at instant k;

- $x_2(k)$ is the number of mobile devices which are using LTE network at k;

- $x_3(k)$ is the number of mobile devices which are using Wi-Fi network at k;

- $\lambda(k)$ is the number of new mobile devices arriving at k;

- $\mu(k)$ is the number of mobile devices leaving the states x_2 or x_3 at k;

- $r_{12}(k)$ is the number of devices steering to the LTE network at k;

- $r_{13}(k)$ is the number of devices steering to the Wi-Fi network at k;

- $\theta(k)$ is the number of blocked devices decided by the controller to avoid network overload at k.

After defining all the variables of the model, we are ready to describe the evolution of the number of mobile devices in every state $x_i(k)$:

$$
\begin{cases}
x_1(k+1) = x_1(k) + \lambda(k) - r_{12}(k) - r_{13}(k) - \theta(k), \\
x_2(k+1) = x_2(k) + r_{12}(k) - \mu(k), \\
x_3(k+1) = x_3(k) + r_{13}(k) - \mu(k).
\end{cases}
\tag{4.1}
$$

To transforme this model into a dynamic control system representation form, Let us denote $r'_{12}(k) = r_{12}(k) - \mu(k)$, $r'_{13}(k) = r_{13}(k) - \mu(k)$, $\theta'(k) = \lambda(k) - 2\mu(k) - \theta(k)$. Thus, we obtain from (4.1) the following linear system:

$$
\begin{cases}
x_1(k+1) = x_1(k) + \theta'(k) - r'_{12}(k) - r'_{13}(k), \\
x_2(k+1) = x_2(k) + r'_{12}(k), \\
x_3(k+1) = x_3(k) + r'_{13}(k).
\end{cases}
\tag{4.2}
$$

Let's define the following vectors: $X(k) = \left[x_1(k) - x_1^{\text{ref}}, x_2(k) - x_2^{\text{ref}}, x_3(k) - x_3^{\text{ref}}\right]^{\text{T}}$,

$U(k) = [\theta'(k), r'_{12}(k), r'_{13}(k)]^{\text{T}}$,

where $X(k)$ and $U(k)$ represent, respectively, the state vector and the input (*e.g.,* control) vector. The constants x_i^{ref} for $i \in \{1,2,3\}$ represent the target number of mobiles, which optimizes the resource utilization while avoiding network overload, and $[.]^{\text{T}}$ represents the matrix transpose. Therefore, the system (4.2) can be expressed as follows:

$$\begin{cases} X(k+1) = AX(k) + BU(k), \\ Y(k) = CX(k), \end{cases} \qquad (4.3)$$

where $A = C =$ the identity matrix of dimension 3, and $B = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

The system presented in (4.3) models the variation of the number of mobile devices in every state with different transition parameters. We can observe that (4.3) is in discrete-time. This linear system is unstable, which means that the network occupancy cannot converge to the desired occupancy without any action. This is because that all the eigenvalues of matrix A do not satisfy the stability condition (*i.e.,* the eigenvalues should be lower than 1 [99]).

The system described in (4.3) is controllable and observable. Indeed, the controllability matrix $C_M = \begin{bmatrix} B & AB & A^2B \end{bmatrix}$ and the observability matrix $O_M = \begin{bmatrix} C & CA & CA^2 \end{bmatrix}^{\mathrm{T}}$ have both a full row rank of 3. Thus, a network selection strategy can be conceived with the help of a controller in a way to stabilize the system around the target values.

Now we will design a control theory-based network selection mechanism to stabilize the number of mobile devices in LTE and Wi-Fi network at an ideal workload. The goal of this mechanism is to avoid the network congestion while maximizing the network throughput in LTE and Wi-Fi interworking. This ideal workload corresponds to the targeted Quality of Service (QoS) required by the supported services (see Figure 4.1 for more details). To meet the desirable workload, the number of terminals passing to LTE (*i.e., r_{12}*), Wi-Fi (*i.e., r_{13}*) and the terminals to be blocked (*i.e., θ*) should be determined. We consider the Discrete-Time Linear Quadratic Regulator (DT LQR) controller to figure out the optimal values.

4.4.2   LQR Controller-based State Feedback Design

The Discrete-Time Linear Quadratic Regulator (DT LQR) is a well-known design technique to find an appropriate state-feedback controller in an automated way. Given a discrete-time linear dynamic system described in 4.3, the LQR controller design method makes the measured output converge to the desired reference by minimizing performance index J:

$$J = \sum_{k=0}^{\infty} \Big[X(k)^{\mathrm{T}} Q\ X(k) + U(k)^{\mathrm{T}} R\ U(k) \Big], \qquad (4.4)$$

where $X(k)$ and $U(k)$ represent the state vector and the control vector, respectively. The J is quadratic and also called the cost function, which represents the sum of the deviations of the key measurements from their desired values. The weighting matrices Q and R, which hold the penalties on the deviation of the networks workload from their targeted values, should satisfy:

$$Q = Q^{\mathrm{T}} \geq 0, \quad R = R^{\mathrm{T}} > 0.$$

The main idea behind minimizing J for the network selection problem is to minimize the distance between the current system workload and the desired workload in order to converge to the objective while minimizing the controller output action.

The feedback control vector $U(k)$ can be written as:

$$U(k) = -K \ X(k) \tag{4.5}$$

with the matrix $K = \left[R + B^{\mathrm{T}} P(k+1) B \right]^{-1} B^{\mathrm{T}} P(k+1) A$ and the $n \times n$ matrix $P(k)$ is the solution of the following Riccati difference equation [100]:

$$
\begin{aligned}
P(k) = Q + A^{\mathrm{T}} P(k+1) A - \\
A^{\mathrm{T}} P(k+1) B \left[R + B^{\mathrm{T}} P(k+1) B \right]^{-1} B^{\mathrm{T}} P(k+1) A.
\end{aligned}
\tag{4.6}
$$

In steady state, $P(k) = P(k+1) = P$, thus the Riccati equation expressed in (4.6) can be written as:

$$P = Q + A^{\mathrm{T}} P A - A^{\mathrm{T}} P B (R + B^{\mathrm{T}} P B)^{-1} B^{\mathrm{T}} P A. \tag{4.7}$$

The optimal control can thus be described by:

$$U(k) = -\left(R + B^{\mathrm{T}} P B \right)^{-1} B^{\mathrm{T}} P \ A \ X(k). \tag{4.8}$$

The control matrix $U(k)$ depends on the matrices A, B, Q, R. While A and B are defined in the system, Q and R are the principal factors which determine the trade-off between the control effort and the regulation for the controller. We can choose the appropriate Q and R for different control requirements.

4.4.3   Analytical validation

In the previous section, we have described the details of our model and the proposed network selection mechanism, we now direct our focus on evaluating its performance using the computer simulation. We use MATLAB to validate our controller design. We assume that the average arrival rate of new session is represented by λ. Once the mobile device chooses a network, the average time it will stay in this network is denoted by $1/\mu$. We denote that the identity matrix is represented by I and we choose that $Q = I$ and $R = \rho I$, where the ρ is the control effort coefficient. The initial state vector is set to $x_1 = x_2 = x_3 = 0$.

The reference is set to $x_1^{\text{ref}} = 0$, $x_2^{\text{ref}} = 50$, $x_3^{\text{ref}} = 100$. We run the simulation when the network will be congested owing to the heavy arrival of mobile devices. Since the system is time-discrete, the simulation will show how many steps the system will be required to achieve the objective with the controller.
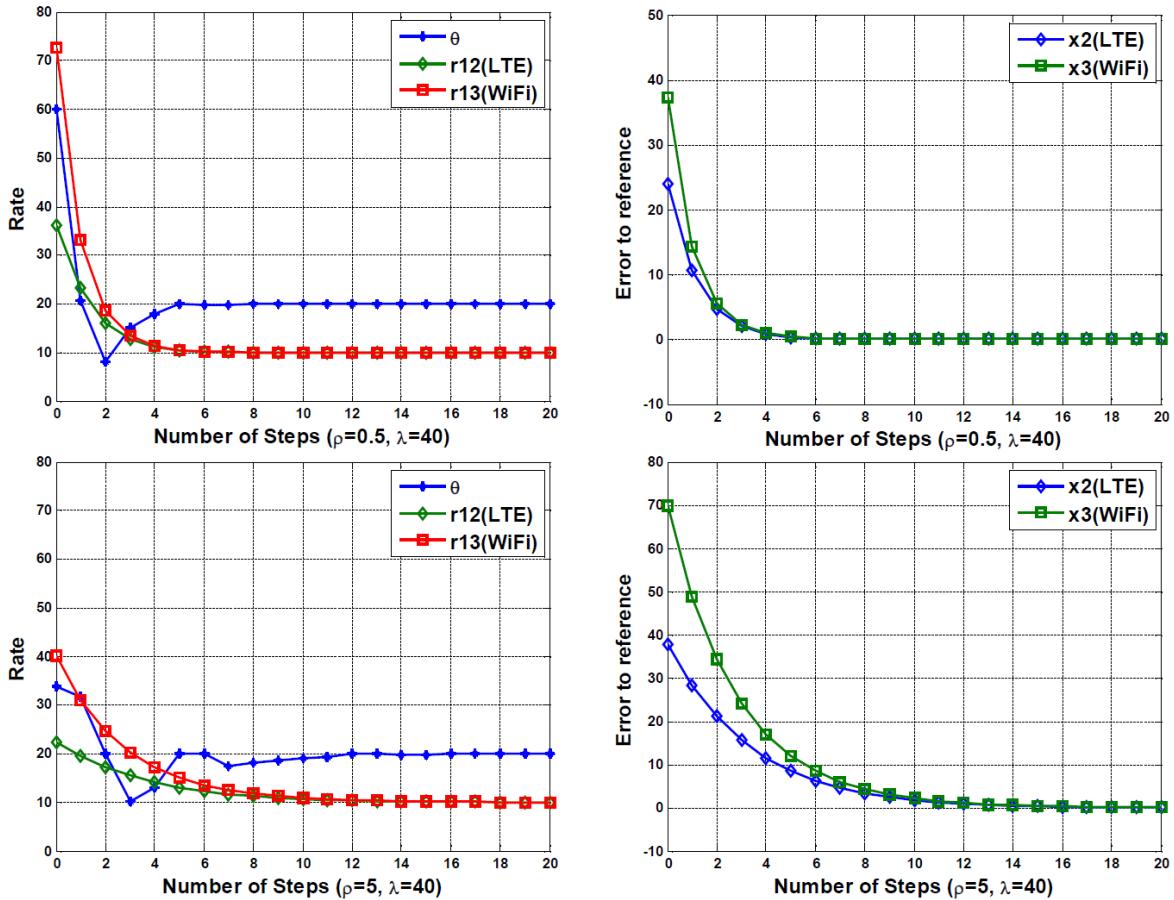
Figure 4.4 – Evaluation with Different Values of ρ

In the first evaluation, the λ is fixed at 40 and μ is fixed at 10. We measure the difference between the actual workload and the ideal workload, which is called *error to reference.* Figure 4.4 showed how the system evolves under the control of controller with different ρ. All of the final states can achieve the state which are around the reference values because the *error to reference* converged to 0. This validates our controller design. We notice that the ρ has impacts on the weight of the controller action compared with the difference between the target workload and the desired workload. The performance of the controller varies for different values of ρ. We can also notice that the number of mobile devices converges faster with smaller ρ. For example, only 6 steps are needed to converge to the reference with $\rho = 0.5$ while nearly 14 steps are needed with $\rho = 5$. These results are consistent with the performance criteria J, which in this case favors the reduction of the control action.

Meanwhile, we can see that faster convergence will require more control effort. For instance, the initial value of r_{13} is 72 for $\rho = 0.5$ while the initial value of r_{13} is 40 for $\rho = 5$. Therefore, the controller performance and the control effort constitute a trade-off, and we can design distinct controllers under different control effort and performance constraints by choosing the appropriate matrix Q and R.

Now we vary one of the external parameter λ which is not in the controller vector. In this evaluation, the value of ρ is fixed to 1. We vary the value of mean arrival rate λ. The results are showed in Figure 4.5. Firstly we notice that the number of steps needed did not change with different λ. However, the number of blocked device is increasing when the λ changes from 30 to 50. This is normal because the capacity of the network is limited so that there should be more blocked devices to guarantee that the network is not overloaded. However, in the real system, it is not practical because too many blocked devices will also degrade the QoE. We can then limit the number of blocked devices according to distinctive demand of QoE with the cost of more steps to get the convergence, for example. Therefore, we should design an adaptive controller which takes the QoE into account to satisfy different requirement considering the different trade-off between the QoE, controller performance, and effort, which is a part of our future work.

4.5 General network selection model in heterogeneous network

In the Section 4.4.1, we have proposed a control theory-based mechanism for network interface selection in an LTE and Wi-Fi interworking context. This section will
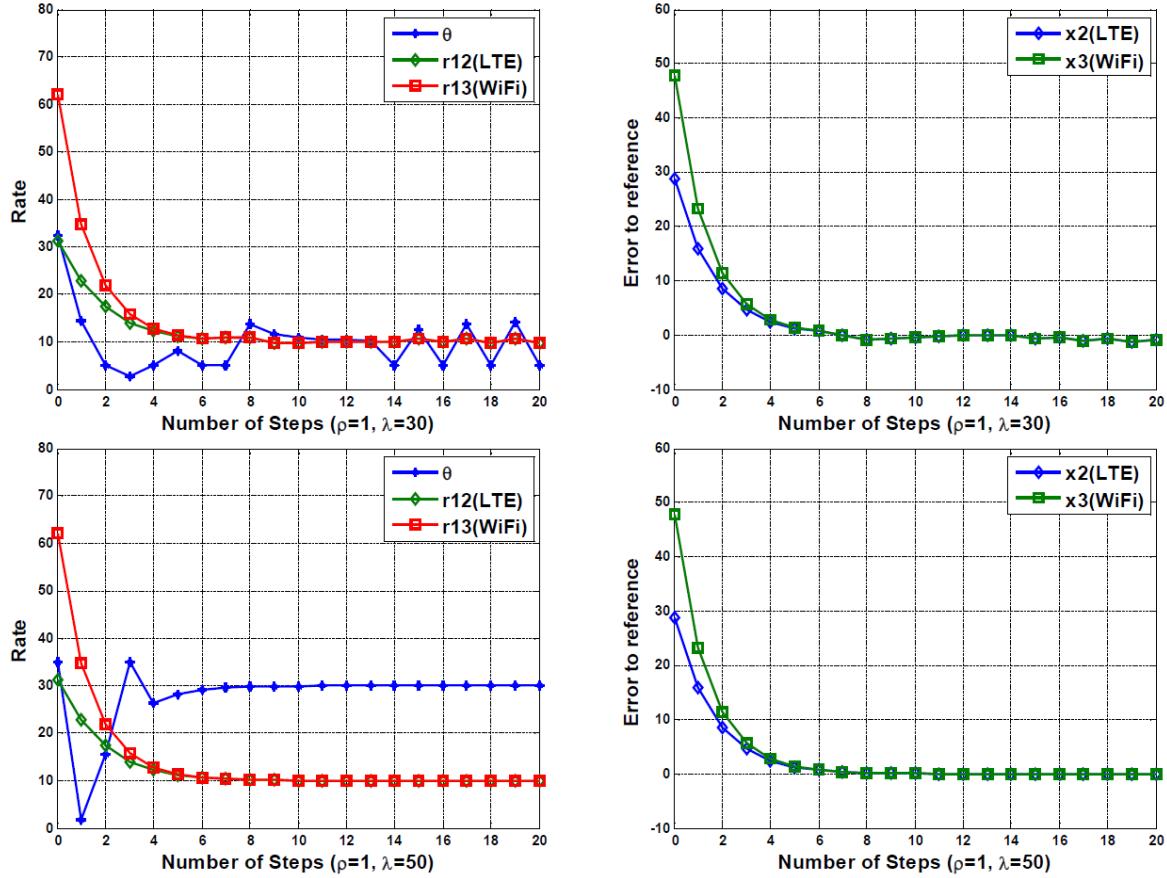
Figure 4.5 – Evaluation with Different Values of λ

introduce a more general model, in which traffic prioritization can be taken into account in the context of any number of heterogeneous networks. To the best of our knowledge, this is the first approach applying the control theory to address the network selection problem in the heterogeneous network. We believe that the designed controller can achieve an optimized network resources' utilization by providing the appropriate interface selection information to mobile devices.

4.5.1   General Model Description

In the following, we introduce a general model for describing the network selection process and the corresponding controller, which can be triggered to keep the actual network workload around the ideal workload while maximizing the resource utilization and avoiding network performance degradation. Before describing the details of the proposed model, we start by specifying the different parameters of the system. We

assume in the following that there are l traffic classes and m access networks, which can be used by the arriving sessions.

- c_i: The number of sessions in class i before network selection, where i is an integer in $[1,l]$.

- n_j: The number of sessions in network j after network interface selection decision, where j is an integer in $[1,m]$.

- r_{ij}: The number of sessions which select the network j from class i.

- θ_i: The number of sessions of class i, which are blocked and retry for connections after a certain period of time (backoff).

- λ_i: The arrival rate of new sessions in class i.

- μ_j: The departure rate of sessions from network j.
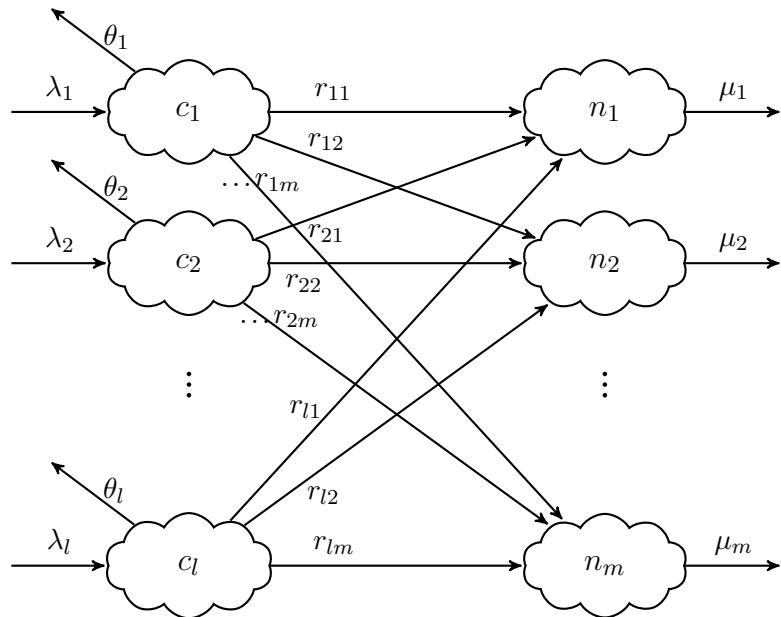
Figure 4.6 – System model

The proposed network selection model, which is described in Figure 4.6, assumes that the average sessions' arrival rate for class i is λ_i. Then, based on a decision mechanism, there will be r_{ij} sessions which connect to one of the access networks j from the class i. At the same time, there will be θ_i blocked sessions depending upon the congestion state of the access networks. After a certain period of time, the sessions leave the system with the average rate μ_i.

Having defined the network selection model and all its parameters, we now direct our focus on describing the evolution of the sessions' number using the following discrete-time system of equations:

$$\begin{cases} c_1(k+1) = c_1(k) + \lambda_1(k) - \sum_{i=1}^{m} r_{1i}(k) - \theta_1(k), \\ c_2(k+1) = c_2(k) + \lambda_2(k) - \sum_{i=1}^{m} r_{2i}(k) - \theta_2(k), \\ \vdots \\ c_l(k+1) = c_l(k) + \lambda_l(k) - \sum_{i=1}^{m} r_{li}(k) - \theta_l(k), \\ n_1(k+1) = n_1(k) + \sum_{i=1}^{l} r_{i1}(k) - \mu_1(k), \\ n_2(k+1) = n_2(k) + \sum_{i=1}^{l} r_{i2}(k) - \mu_2(k), \\ \vdots \\ n_m(k+1) = n_m(k) + \sum_{i=1}^{l} r_{im}(k) - \mu_m(k), \end{cases} \tag{4.9}$$

Let us denote: $r'_{ij}(k) = r_{ij}(k) - \frac{\mu_j(k)}{l}$, and $\theta'_i(k) = \lambda_i(k) - \theta_i(k) - \frac{\sum_{j=1}^{m} \mu_j(k)}{l}$, now the system represented in (4.9) can be reformulated as follows:

$$\begin{cases} c_1(k+1) = c_1(k) + \theta'_1(k) - \sum_{i=1}^{m} r'_{1i}(k), \\ c_2(k+1) = c_2(k) + \theta'_2(k) - \sum_{i=1}^{m} r'_{2i}(k), \\ \vdots \\ c_l(k+1) = c_l(k) + \theta'_l(k) - \sum_{i=1}^{m} r'_{li}(k), \\ n_1(k+1) = n_1(k) + \sum_{i=1}^{l} r'_{i1}(k), \\ n_2(k+1) = n_2(k) + \sum_{i=1}^{l} r'_{i2}(k), \\ \vdots \\ n_m(k+1) = n_m(k) + \sum_{i=1}^{l} r'_{im}(k), \end{cases} \tag{4.10}$$

Now we define the state vector X(k):

$$X(k) = \left[C_l(k) - C_l^{\text{ref}}, N_m(k) - N_m^{\text{ref}} \right]^{\text{T}},$$

where

$$C_l(k) = [c_1(k), c_2(k), \ldots, c_l(k)],$$

$$N_m(k) = [n_1(k), n_2(k), \ldots, n_m(k)].$$

The constant vectors $C_l^{\text{ref}} = \left[c_1^{\text{ref}}, c_2^{\text{ref}}, \ldots, c_l^{\text{ref}} \right]$ represents the reference number of sessions before making selection decisions in each traffic class. $N_m^{\text{ref}} = \left[n_1^{\text{ref}}, n_2^{\text{ref}}, \ldots, n_m^{\text{ref}} \right]$ represents each target network occupancy, which optimizes the resource utilization

while avoiding network overload if the network states converge to the target. A matrix transpose is represented by $[.]^{\mathrm{T}}$.

The control vector $U(k)$ is defined as:

$$U(k) = [-\Theta_l(k), R_{l1}(k), R_{l2}(k), \ldots, R_{lm}(k)]^{\mathrm{T}},$$

where $\Theta_l(k) = [\theta_1'(k), \theta_2'(k), \ldots, \theta_l'(k)]$, $R_{li}(k) = [r_{1i}'(k), r_{2i}'(k), \ldots, r_{li}'(k)]$. Thus, the system (4.10) can be expressed in the following discrete-time linear system form:

$$\begin{cases} X(k+1) = AX(k) + BU(k), \\ Y(k) = CX(k), \end{cases} \tag{4.11}$$

where $A = C = I_{l+m} = diag(\underbrace{1, 1, \ldots, 1}_{n+m})$,

$$B_{(l+m)\times(lm+l)} = \begin{pmatrix} -I_{l\times l} & -I_{l\times l} & -I_{l\times l} & \ldots & -I_{l\times l} \\ 0_{m\times l} & Z_1 & Z_2 & \ldots & Z_m \end{pmatrix}$$

with $Z_i = [z_1, z_2, \ldots, z_m]^{\mathrm{T}} \times [\underbrace{1, 1, \ldots, 1}_{l}]$, where $z_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}$, $j = 1, 2, \ldots, m$. $Y(k)$ is a $(l+m) \times 1$ output vector which represents the workload of each network we want to observe periodically with discrete time of k.

The system expressed in (4.11) models the variation of the session numbers in every state with different transition parameters. We observe that (4.11) is unstable, which means that the network workload cannot converge to the desired workload if no control action is performed. Indeed, all the eigenvalues of the matrix A do not satisfy the stability condition (i.e. the eigenvalues should be lower than 1 to satisfy the stability condition [99]).

When designing a controller for such dynamic system, the controllability is one of the main characteristics, which should be analyzed. The controllability describes the ability of an external input to move the output from any initial state to any targeted final state in a finite time interval. A system is said to be controllable when the controllability matrix which is given by

$$\mathcal{C} = \begin{pmatrix} B & AB & A^2B & \ldots & A^{l-1}B \end{pmatrix}$$

has a full row rank. It can be easily checked that the system (4.10) is controllable. Indeed, its controllability matrix $\mathcal{C}$ has a full row rank of $l+m$. This means that we are able to design a controller to make the system output converge to the target.

4.5.2   LQR Controller Design for Optimized Network Selection

After establishing the system model and analyzing its characteristics, we are ready to design an access network interface's selection mechanism, by which we can achieve an optimal network workload, as determined in Section 4.2. Indeed, the main objective of the designed controller is to converge the number of sessions in each network to the desired workload in order to maximize the network resource utilization, while avoiding the network congestion.

To achieve the desired workloads, the number of sessions coming from the various classes should be steered to the different networks (i.e. this requires the calculation of the r_{ij}, for all i and j) or blocked (i.e. this requires the calculation of θ_i, for all i). To figure out these quantities, we still apply LQR controller [100] with the same procedure presented in Section 4.4.2, which allows computing in real-time these values optimally to find the control vector by solving the Ricatti equation 4.6 to get the control vector $U(k)$

It can be seen from equation 4.8 that the control vector $U(k)$ allows to determine the quantities r_{ij} and θ_i for all i and j based on the gap existing with the objective and the matrix P, which is determined dynamically. This gives the access controller the capability to adapt to network parameters variation.

4.5.3   Integration within a 3GPP Standardized Network

In practical use cases, the more natural way to handle the proposed mechanism in mobile networks is to integrate it within the Access Network Discovery and Selection Function (ANDSF) server. In fact, the ANDSF server, which is introduced in the 3GPP standards [41] [42], has the role of assisting mobile devices in discovering and selecting 3GPP and non-3GPP networks. An IP-Based interface of communication between the server and the mobile devices is, indeed, already defined (i.e. the interface *S14*). Moreover, Management Objects (MO) are used by the ANDSF to encapsulate messages carrying out information[2] related to the surrounding access networks.

The different quantities obtained, using the proposed controller, allow to have an efficient scheme for both network selection and admission control. Using these quantities in the scenario described above is not easy and may induce a lot of signaling overheads. Indeed, each terminal willing to connect needs to exchange signaling packets

---

[2]The message in an MO contains information like the validity areas, the position of the UE and the availability of access networks in terms of geographical coordinates.

with the ANDSF server. In this way, we propose to transform these quantities into probabilities of access, which can be broadcasted to the devices through a control channel.

Let us take one of the traffic class (e.g. c_1) when this traffic class has m networks which can be selected. The access probability P_j to access the network j is calculated as follows:

$$P_j(k) = \frac{r_{1j}(k)}{\sum\limits_{j=1}^{m} r_{1j}(k) + \theta_1(k)}, \quad \text{for all } j \text{ in } [1, m] \tag{4.12}$$

and the the probability to be blocked P_{block} is given by:

$$P_{\text{block}}(k) = \frac{\theta_1(k)}{\sum\limits_{j=1}^{m} r_{1j}(k) + \theta_1(k)}. \tag{4.13}$$

These probabilities are broadcasted, for network interface selection's purpose, through the ANDSF, which gives the possibility of broadcasting the network access information to the mobile devices. This reduces significantly the signaling information exchanged over the wireless link and simplify the interface selection decision for the devices.

The access network selection's probability can be added into the MO which can be fetched by mobile devices. The procedure is described as follows:

- Mobile devices need to be attached to the 3GPP network in order to receive the signaling information sent by ANDSF server.

- ANDSF broadcasts periodically the network information to the mobile devices.

- Mobile devices retrieve network information from the MO object and select the network interface according to the access network selection probability.

Each device generates randomly a value in the interval $[0, 1]$ (i.e. a uniformly distributed random value). If the value belongs to the interval $[0, P_1(k)]$, then it connects to the network 1; if the value belongs to the interval $\left[\sum\limits_{z=1}^{j-1} P_z(k), \sum\limits_{z=1}^{j} P_z(k)\right]$ (for j in $[2, m]$), then it connects to the network j; otherwise, it is blocked. If the communication is blocked the mobile device may reattempt a new access using the same process, after a random period of time. Note that $\sum\limits_{z=1}^{m} P_z(k) + P_{\text{block}}(k) = 1$ and this access probability calculation can be applied to any other traffic classes.

4.5.4   Analytical Validation

Having described the details of the proposed approach, which combines interface selection and access control, we direct now our attention to its validation using computer simulation under MATLAB.

We assume in the following three traffic classes (i.e. c_1, c_2, and c_3) and three access networks: LTE, Wi-Fi and Femto, with their corresponding workload 200, 150 and 100, respectively. In real use cases, the ideal workloads can be determined dynamically based on the number of sessions and their average arrival rate or other networks' characteristics (see Figure 4.1 for more details). We also assume that mobile devices are capable of connecting to the three communication networks and they are under the overlapping coverage of the three networks.

The average arrival rates are configured as follows (sessions/step): $\lambda_1 = 10$, $\lambda_2 = 15$, $\lambda_3 = 20$ for the classes c_1, c_2, and c_3, respectively. Similarly, the average departure rates are (sessions/step): $\mu_1 = 15$, $\mu_2 = 20$, and $\mu_3 = 25$. The identity matrix is selected for both matrices Q and R. The value of the initial state vector is set to 0. We run the simulation when the network is congested owing to the heavy arrival of sessions. Since the system is time-discrete, the simulation will show how many steps are needed to reach the objective. In the evaluation, we measure the difference between the actual workloads of the different access networks and their respective references, which is called *Tracking error*.
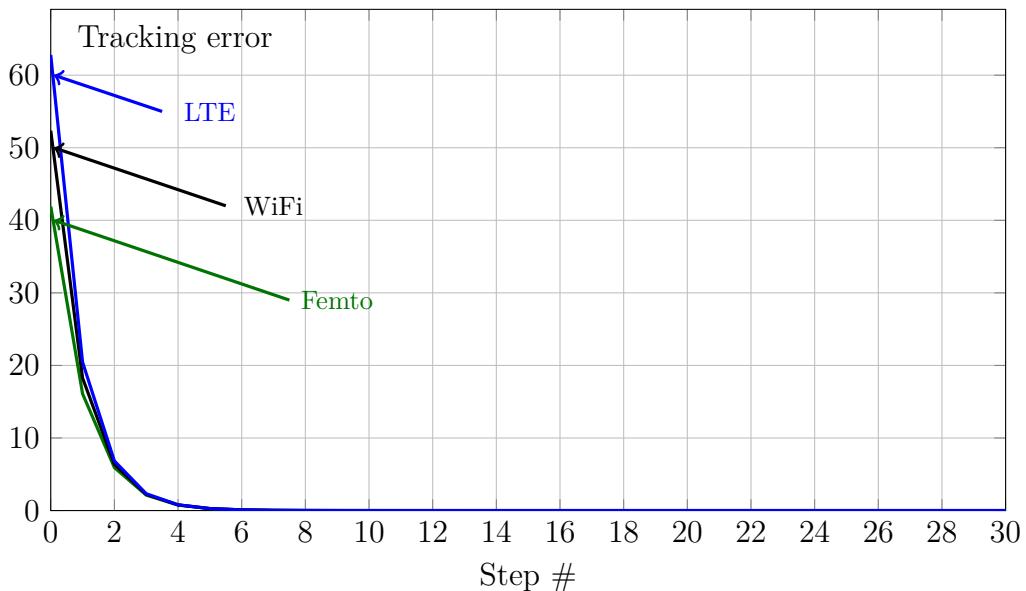
Figure 4.7 – Tracking Error for Different Networks
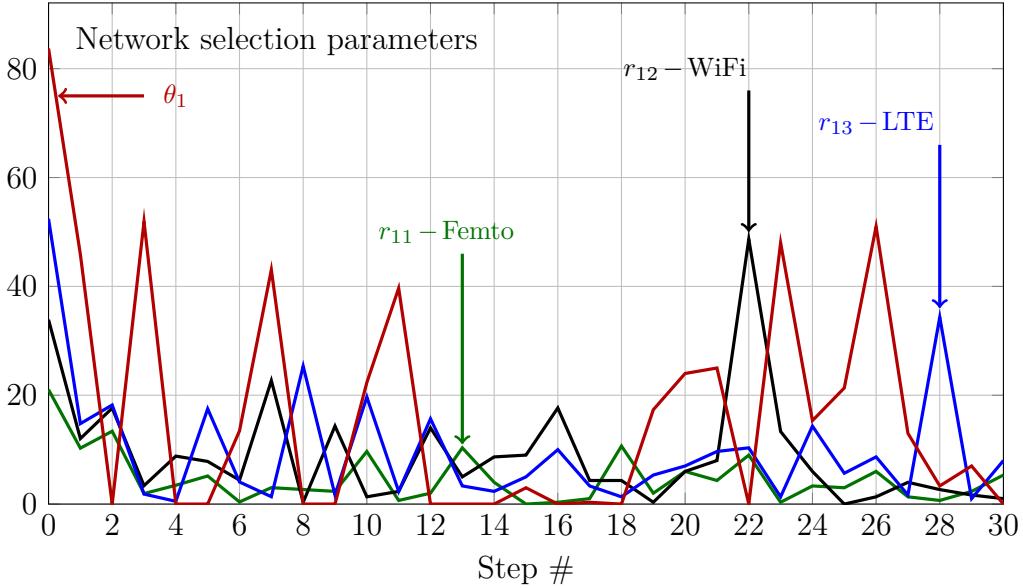
Figure 4.8 – Network Selection Parameters for Traffic Class 1

Figure 4.7 illustrates how the tracking errors vary when using the proposed mechanism. We notice that all the final states in the three networks can converge asymptotically to the targeted values as the tracking errors converge to 0. This demonstrates the stability of the whole system and validates the efficiency of the designed mechanism. Figure 4.8 presented the control parameters' variation for the traffic class c_1. We observe that the controller can tune the transition parameters (i.e. r_{11}, r_{12}, r_{13}) and the blocking parameter (i.e. θ_1) according to the state of the system in real-time, making the system converge to the reference. We obtained similar results for traffic class c_2 and c_3.

4.5.5 Real Scenario Simulation

To validate the proposed mechanism in realistic conditions, when using the access probabilities calculated in Section 4.5.3, we implemented the controller using the NS-3 simulator [94]. We simulate a number of devices with both interfaces LTE and Wi-Fi. After their attachment to the LTE network, they start receiving the access selection probabilities sent periodically by the controller. Then, before initiating the sessions, the mobile devices execute the network selection procedure described in Section 4.5.1. In the simulation, we model the new sessions' arrivals as a Poisson process with $1/\lambda = 0.3$ and the application duration follows an exponential distribution with the average duration of $30s$.

Figure 4.9 shows the variation of new arrivals of mobile devices, or sessions as we considered to have only one session per device. The simulation measures the first 200 mobiles devices arriving at the area of the LTE and Wi-Fi interworking. Figure 4.10 represents how the number of mobile devices is controlled in LTE and Wi-Fi networks. According to Section 4.2, the ideal workload for LTE and Wi-Fi networks is 10 and 15, respectively. We observe that the workload in each network can be always fixed around the targeted workload despite the intense arrival of mobile devices. Figure 4.11 depicts the different access probabilities sent by the controller, which are the function of the workload of the different access networks. It can be seen that when the workload is approaching the reference, the blocking probability calculated by the controller is increasing to maintain the current workload around the reference value. Furthermore, when mobile devices begin to leave the system, the blocking probability decreases as the access networks start being in relaxed conditions, and thus there is no need to block the new arriving mobile devices. This clearly validates our controller design in a real networking environment.
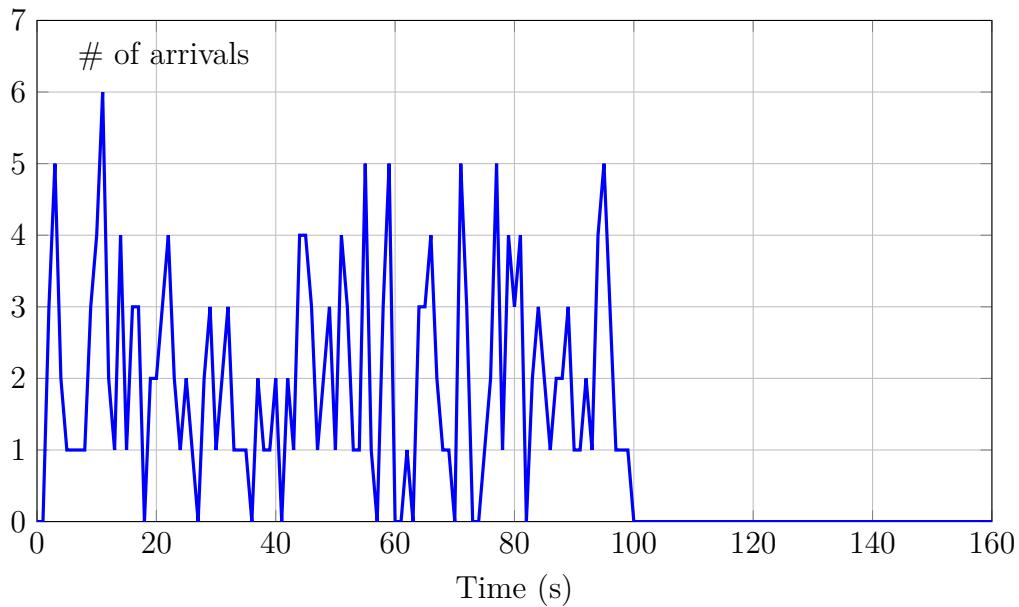
Figure 4.9 – Variation of the Number of Devices' Arrivals

4.6   Summary

In this chapter, we described network selection procedure by proposing a simple model for LTE and Wi-Fi interworking and extended it to a general model when having
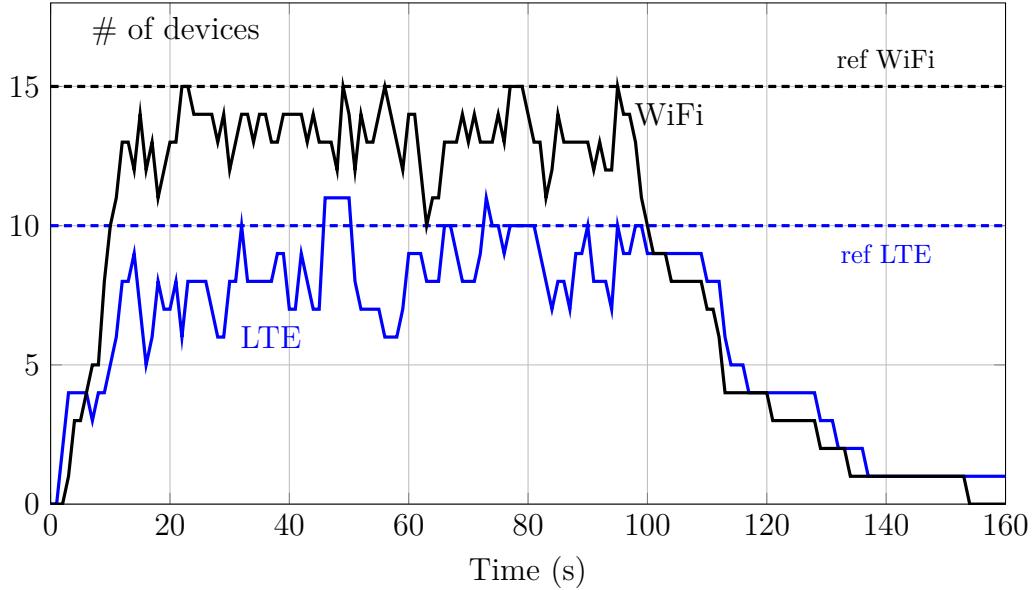
Figure 4.10 – Variation of the Number of Devices in LTE and Wi-Fi Networks
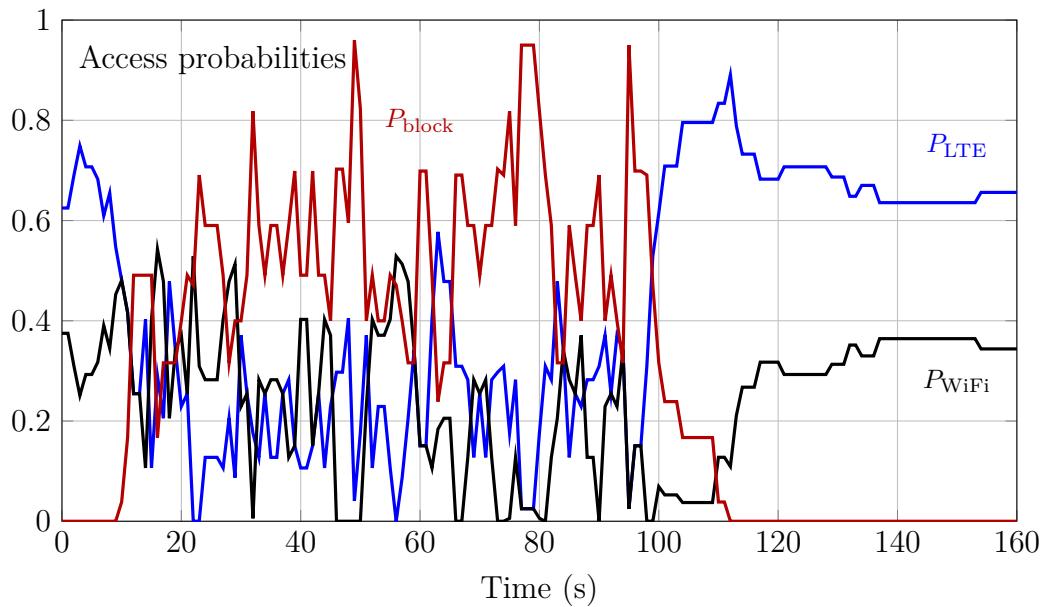
Figure 4.11 – Variation of the Network Interface Selection Probabilities

multiple traffic classes in heterogeneous network environments. The proposed models were, then, used to derive a scalable mechanism based on control theory, which allowed not only to assist in steering dynamically the traffic to the most appropriate network access but also helped in blocking the residual traffic dynamically when the network is congested by adjusting optimal access probabilities. We discussed the integration

of our approach with the current 3GPP standardized network. Compared to existing contributions, the proposed mechanism is (i) scalable as it presents a reduced complexity. (ii) Efficient as it decreases significantly the signaling overhead. Extensive simulations showed that the proposed approach minimizes network congestion while maintaining the workload around the targeted values, which allows maximizing resource's utilization.

Chapter 5

A Multi-access Edge Computing Based-Architecture for Improving HTTP Adaptive Streaming

5.1 Introduction

With the rapid growth of HTTP Adaptive Streaming (HAS) multimedia video service on the Internet, improving the Quality of Service (QoS) of the video delivery will be highly requested in the Fifth Generation (5G) era. The major Mobile Network Operator (MNO) are starting the deployment on their own Content Delivery Networks (CDN), which is called Telco CDN. Telco CDN is considered as a major technical and economic response to reduce the pressure on their network resources. Since MNOs operate their own content delivery infrastructure and thus have a better control over the utilization of their resources [101] [102]. However, as discussed in Chapter 3, the performance improvement of HTTP-based adaptive streaming should be further studied to face the fast changing network condition in the current network architecture, which is a key point to make the Telco CDN be successful in the future.

The new emerging European Telecommunications Standards Institute (ETSI) standard on Multi-access Edge Computing (MEC) [103] may play an important role in this direction. MEC provides the possibility to leverage the cloud computing power by deploying IT application services at the edge of the mobile network. This will potentially facilitate content dissemination within the access network and will offer new business opportunities by integrating the MNO into the video delivery value chain.

In this chapter, we propose a novel architecture for HTTP-based adaptive streaming tailored to an MEC environment. The proposed architecture introduces an adaptation algorithm running on an MEC server so that it can relax network congestion while offering Quality of Experience (QoE) improvements for mobile users. The main idea is to dynamically control the video representations available to clients based on the current network status, thus driving the client-side video adaptation mechanism. To achieve this, we exploit the awareness of network-level information, which could be retrieved via an MEC Application Programming Interface (API) exposed by the MNO, and specific features of the Dynamic Adaptive Streaming over HTTP (DASH) standard. Our approach illustrates a use case for a mutually beneficial Content Provider (CP)-MNO collaboration over the MEC infrastructure. It is worth noting that the proposed mechanism is standards-compliant and transparently coexists with client-side bitrate adaptation algorithms. Indeed, the final quality level selection decision is still made by the clients, which corresponds to the current DASH client standard specifications.

To evaluate the performance of the mechanisms, we emulate our architecture on a virtualized testbed. Our results show that fairness among mobile users is guaranteed, and the overall perceived QoE is improved when the radio access network is congested. The proposed solution represents a win-win situation for content providers and mobile operators since it can improve end-user experience for video content/service providers while optimizing network resource utilization for mobile network operators. By actively involving the MNO in the video delivery chain, the proposed solution also allows for compensating the loss of MNO revenue after the introduction of flat rates, which had reduced the average revenue per user.

5.2 Multi-access Edge Computing: Background and Related Work

5.2.1 Multi-access Edge Computing

Multi-access Edge Computing (MEC) has emerged as a new evolution of mobile networks and it is a new ETSI standardization initiative, supported by market leaders. MEC proposes to leverage the cloud computing to deploy various applications and content caching on cloud computing-like capabilities at the edge of the networks. By bringing applications and resource-heavy tasks closer to the users, network congestion can be minimized, the latency between the users and the applications can be reduced and operators can optimize their infrastructures and differentiate their services. MEC

techniques can also contribute to scalability and QoS improvements by properly moving computing and storage features to the edge, possibly at service provisioning time, thus implementing a highly distributed, decentralized, and when possible loosely-coupled infrastructure.
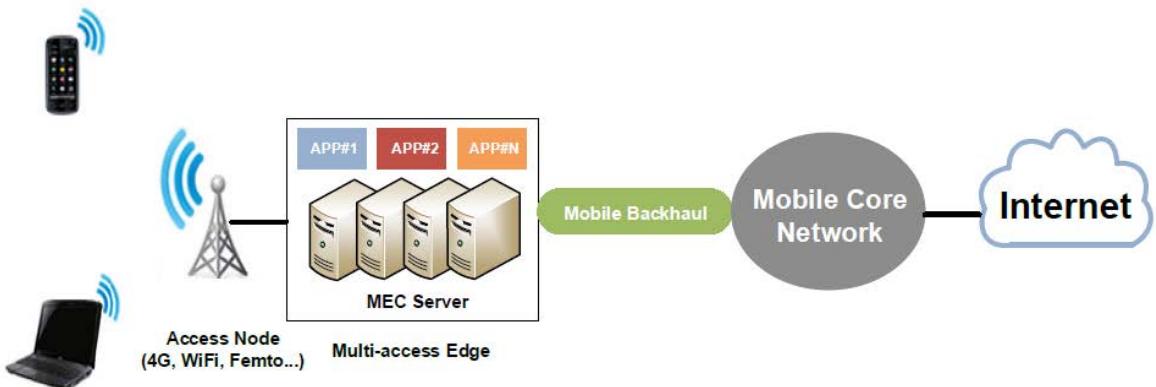
Figure 5.1 – Multi-access Edge Computing Architecture.

Figure 5.1 depicts how MEC servers are integrated into the mobile network. The key component for enabling MEC are the MEC servers, which are integrated within the operator's Radio Access Network (e.g. 3GPP, Wi-Fi or small cells). Each MEC server is capable of hosting several MEC applications. MEC applications are able to add additional information to the UE traffic and forward it to its original destination or reply directly to UE requests. These applications can be developed by authorized third parties such as content providers, which can add the flexibility to handle the traffic from/to mobile users. Moreover, operators can expose their RAN edge API to authorized third parties to provide them with radio network information in a real-time manner. Therefore, MEC paves the way for addressing the current issues of HTTP-based adaptive streaming.

MEC-based architecture is already studied in the research work to address the issues in a diversities directions. In [104], the authors leverage the MEC approach to propose an architecture called MUREN (Multi-Radio Edge Node) for managing traffic in future mobile edge networks. By using real-time radio network information, service operation and user experience are improved for the scenario of critical SMS and LR (Low Rate) waveform. In [105], the authors present an investigation on the progress of MEC and propose a platform, named WiCloud, to provide edge networking, proximate computing and data acquisition for innovative service. In [106], the authors claim the suitability of deploying EPCaaS (Evolved Packet Core (EPC) support solutions as a Service) over a uniform edge cloud infrastructure, by following the concepts of

Network Function Virtualization (NFV). They develop a state sharing mechanism across different data centers even in presence of firewall/network encapsulation. In [107], the authors propose novel local streaming service concepts and business models based on Mobile Edge Computing by utilizing a relevant theoretical framework. In [108], the authors introduce ME-VoLTE, an MEC-based video telephony system that aims to reduce power consumption of mobile devices. To this end, encoding efforts during video calls are offloaded to nearby MEC servers. In [109], the authors study energy-efficient computation offloading mechanisms for MEC in 5G heterogeneous networks. Then an optimization problem is formulated to minimize the energy consumption of the offloading system, where the energy cost of both task computing and file transmission are taken into consideration. However, in the literature above, few studies address this issue of HTTP-based adaptive streaming by leveraging MEC approach. In [110], an architecture for optimizing HTTP-based multimedia delivery in mobile networks is studied. The authors propose a network-assisted approach based on adaptive HTTP streaming with multi-layer encoding and MEC. They introduce a mobile edge-DASH adaptation function (ME-DAF) located within the Cloud Radio Access Network (C-RAN). However, their architecture still considers the Channel Quality Indicator (CQI) for the selection strategy, which may represent an important signaling overhead.

Therefore the motivation of the work our thesis is to further explore how MEC application can be applied DASH to improve the video delivery. We believe that by placing intelligent application in the MEC server, the video delivery can be optimized, which benefits both end users and operators and content providers.

5.3   Proposed architecture

5.3.1   Architecture Description

We propose an adaptive HTTP video streaming solution tailored to a MEC environment. A distinctive characteristic of our approach is that the MEC end is an active component in the video delivery chain. In particular, a special service located at the mobile edge is responsible for dynamically controlling the video representations available for delivery, exploiting (i) specific features of the MPEG-DASH standard and (ii) the availability of information on the conditions in the wireless access network, which can be made available through a MEC API. Our mechanism is standards-compliant and compatible with receiver-driven adaptive video delivery algorithms, with which it cooperates in a transparent manner. The architecture we introduce is presented in Figure 5.2.

Our use-case scenario involves a level of cooperation between the CP and the MNO, and this is implemented through the mobile edge: A CP hosts video descriptor files, potentially, in its premises, while using CDNs to host and distribute videos from the optimal locations. At the same time, the CP deploys intelligent services at the mobile edge infrastructure. These MEC services are controlled by the CP but utilize network-level information provided by the Internet Service Provider (ISP) using an MEC API to assist sophisticated mechanisms which optimize video delivery. In our case, such a MEC-level service assumes the role of a proxy/redirection node between the end-user and the video servers. At its core, our mechanism intercepts and appropriately modifies the MPEG-DASH Media Presentation Description (MPD) on the fly aiming to match current bandwidth demand with the available mobile access network resources and, consequently, to improve the user's experience.
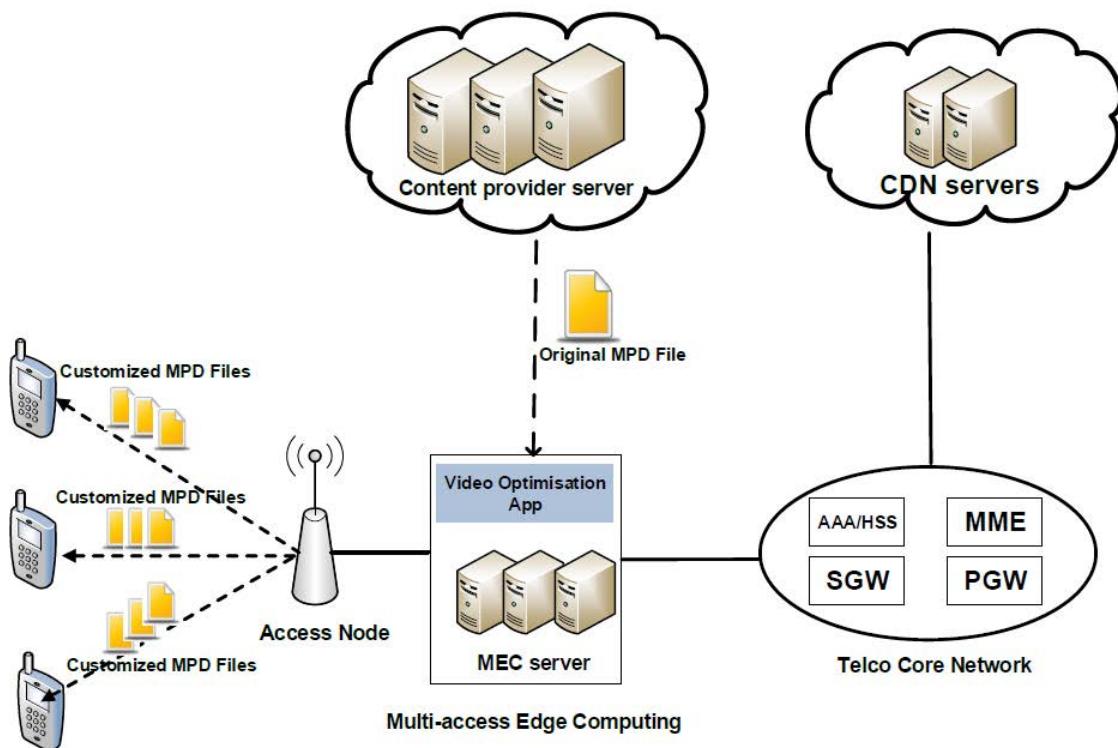
Figure 5.2 – An Architecture for Mobile Adaptive HTTP Streaming with MEC Assistance.

5.3.2   Message flow

We assume that User Equipment (UE) are already attached to the mobile network, and the MEC service is capable of periodically retrieving radio statistics and other network-level information from the network access node. Before downloading the video, each end user accesses the video catalog (potentially after a registration/login process) and selects a file to view. The video player then sends an HTTP request to retrieve the MPD file from the CP's server. This request is intercepted by the MEC service. There are two alternative modes/strategies to achieve this:

Redirect mode   In this mode (Figure 5.3), when the CP server receives the HTTP GET request from the UE for downloading an MPD file, it pushes the original MPD file to the MEC server and notifies UEs to download the MPD file from the given MEC server by sending an HTTP REDIRECT to the client. Then, following the redirect indication provided by the CP, the UE sends another HTTP GET request for the MPD file to the appropriate MEC server, which answers directly to the client by providing the file, after appropriately modifying it if necessary.

Proxy mode   In this mode (Figure 5.4), the MEC application operates as an HTTP proxy for all user traffic; if it detects an HTTP request for an MPD file directed to a content provider, it intercepts the response and modifies its content appropriately, before sending it back to the UE.

In both modes, the same adaptation logic is applied by the MEC service: Video representations are removed from/added back to the original MPD content in response to changing network conditions and demand, but also with the potential of utilizing a wealth of other information, such as user profiles and preferences.

After receiving the (transformed) MPD file, UEs begin to download video content chunk-by-chunk. Note that, in our scheme, the MPD file needs to be refreshed periodically. There are two standards-compliant options to achieve this. The first is to configure its `minimumUpdatePeriod` attribute, which instructs the video client to periodically request an up-to-date version of the file. The other option is using an in-band event (event message box mechanism, in the ISO Base Media File Format (BMFF) terminology) [111, subclause 5.10.3.3]. In this case, a special event box field is inserted in the media segment at video preparation time by the content provider, which can notify the client to refresh the MPD file at any time. In our prototype, we have opted for the first solution.
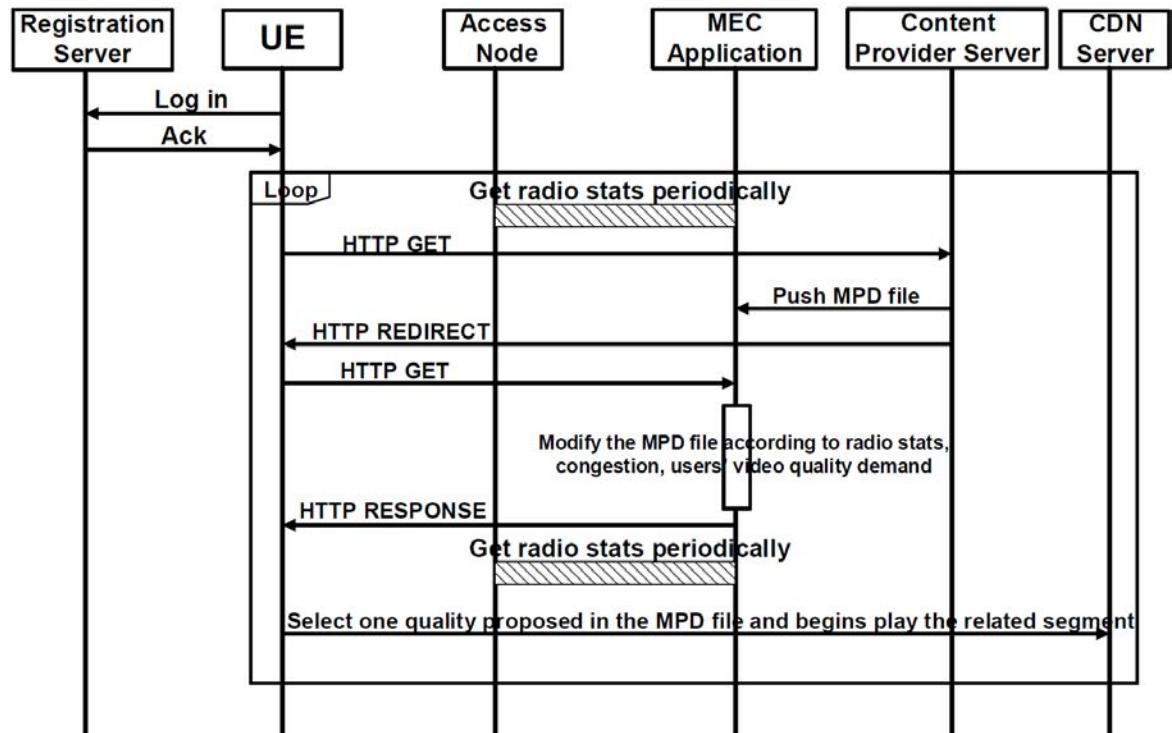
Figure 5.3 – Message Sequence (Redirect Mode).
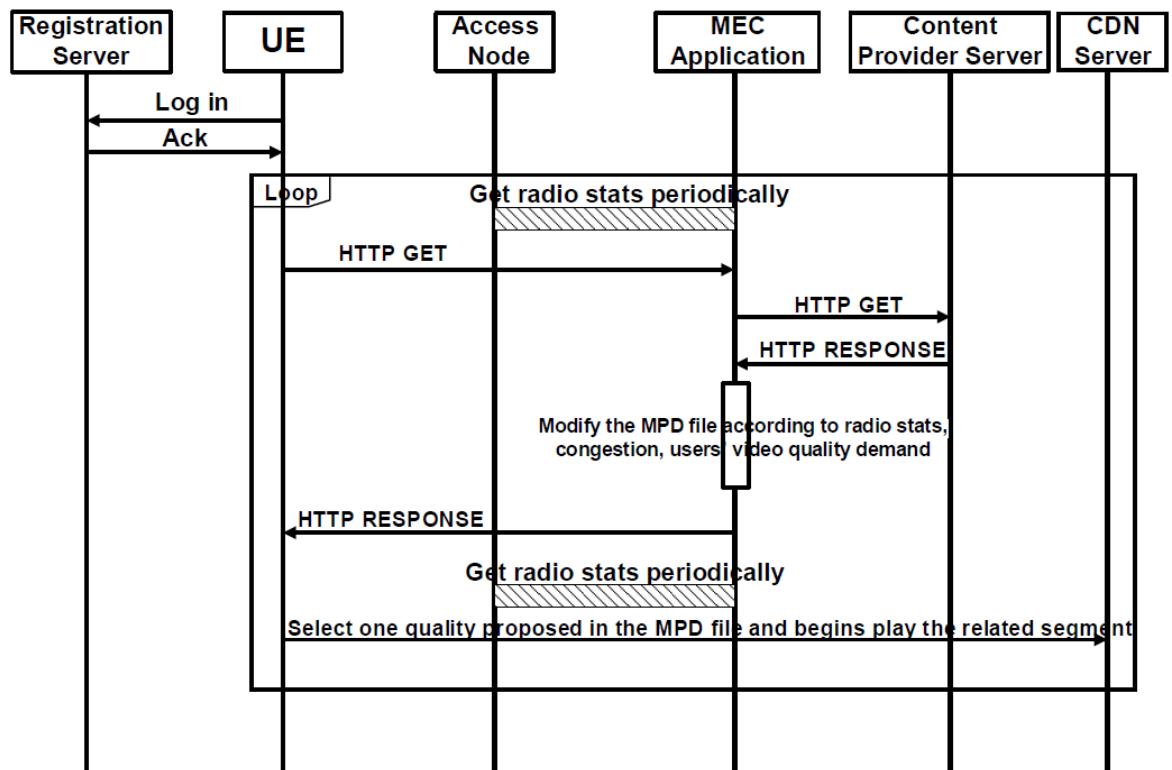
Figure 5.4 – Message Sequence (Proxy Mode).

5.4   Network-aware MEC-assisted video rate control

An MPD update event triggers a mechanism at the MEC end which controls the available representations for each requested video, exploiting awareness of the current network conditions. As noted, this MEC (server)-side scheme is used in combination with traditional receiver-driven bitrate adaptation mechanisms.

We assume that there are n mobile DASH clients in the same cell requesting video content, and each client has the same service priority. To control the set of available video representations, our adaptation mechanism needs to have access to the following information, which is assumed to be available via an MEC API exposed by the network operator:

- The number of UEs (n) currently active and accessing the video service in the cell.

- The wireless access network bandwidth (BW) available/allocated to the video streaming service.

- The network congestion state.

Each time a user requests the MPD file for a video item, our mechanism decides which (if any) representations should be removed from it. This decision is mainly guided by the available network capacity. By simple intuition, since all users have equal priorities, a reasonable choice would be to disallow representations whose bitrate is more than $\frac{BW}{n}$. At the same time, however, for practical reasons, there can be users who, due to poor signal conditions, consume video at a rate significantly lower than $\frac{BW}{n}$, leaving unused access network capacity which could be exploited by other stations. To address this issue, our scheme allows for representations with higher bitrate than $\frac{BW}{n}$, but gradually removing them if network congestion persists.

Let R denote the set of available representations for a specific video, with $r_{max} \in R$ being the one with the highest bitrate in the original MPD file, and r_{max}^* the one with the highest bitrate in the transformed MPD file. Furthermore, the service operator defines a bandwidth margin B_m which is interpreted as follows: All representations whose bitrate is more than $B_m + \frac{BW}{n}$ are removed from the served MPD file. This margin is successively decreased if the network keeps being congested. The network state is expressed by the binary variable $S_{congested}$ (true in the case of congestion).

Our MEC-side adaptation mechanism is summarized in Alogrithm 1. We begin by setting $S_{congestion} = $ False and $B_m = M$, where M is the maximum configured

bandwidth margin. Each time a request for an MPD file is received, the MEC-side application retrieves the current number n of active UEs in its managed cell, as well as the current values for BW and the network state (`isNetworkCongested()` function) using the MEC API. For each successive time the network is found to be congested, the bandwidth margin is reduced by a step δ and the `getRepresentation()` function is called to map the per user $B_m + \frac{BW}{n}$ bandwidth to the maximum bitrate representation to be included in the transformed MPD file; all representations with bitrate higher than that of r^*_{max} are removed. On the contrary, if the network is not found to be congested, we set $S_{congestion} = $ False, $B_m = M$, and $r^*_{max} = r_{max}$.

We consider two strategies with respect to congestion reporting. The *reactive* strategy (default) returns the network congestion state *after* detecting a change in the available network capacity and conditions. On the other hand, the *proactive* strategy implies the existence of a prediction mechanism which identifies congestion in advance. Although the details of such a mechanism are outside the scope of this work, in Section 5.5.4 we assume that an accurate congestion prediction module is in place and evaluate the performance improvements it can bring about compared to the reactive strategy.

Algorithm 1 MEC-side adaptation algorithm

Initialization: Set $B_m = M$; Set $S_{congested} = False$
Iteration:
1: Get n; Get BW; $S_{congested} = $ isNetworkCongested();
2: **if** $S_{congested} = True$ **then**
3:     $B_m = max(0, B_m - \delta)$;
4:     $r^*_{max} = $ getRepresentation$(B_m + \frac{BW}{n})$;
5: **else**
6:     $B_m = M$;
7:     $r^*_{max} = r_{max}$;
8: **end if**
9: $f' = f$; /* f': transformed MPD file */
10: **for all** $r \in R | $bitrate$(r) > $bitrate$(r^*_{max})$ **do**
11:     removeRepresentation(r, f');
12: **end for**
13: return f';

5.5   Performance evaluation

To demonstrate the advantages of our MEC-assisted approach in terms of user experience, we develop an emulated testbed where carry out a set of measurements, which we present in this section.

5.5.1   Experimental methodology and setup

We have set up a virtualized testbed to emulate a mobile edge environment where our architecture is deployed. Our setup is illustrated in Figure 5.5. In our testbed, a Virtual Machine (VM) where the nginx [112] HTTP server runs represents the origin video server managed by a content provider. A separate VM stores the MPD files which are served by an Apache [113] HTTP server and retrieved by clients. The MEC application is hosted in a third VM and is aware of the current number of clients in the network cell it serves, as well as the available bandwidth for the shared link. Its role is to capture client requests for MPD files and, given its network awareness and the current client demand, to execute the adaptation algorithm, retrieving the requested MPD file from the source and appropriately transforming it, before serving it to clients.

Video player instances,[1] each representing a single client, are hosted in a separate VM which is connected to the MEC VM via a virtual link. Using the `netem` Linux utility, we shape the bandwidth of the virtual connection to emulate congestion conditions on the shared wireless access link. The above VMs are hosted on a single workstation and are managed by the kvm [114] hypervisor.

Each client accesses a 9-minute video sequence. In particular, using ffmpeg, we created 5 different H.264/Advanced Video Coding (AVC) representations of the Blender Foundation's "Big Buck Bunny" [115] open-source movie (with approximate bitrates of 128, 512, 1024, 2048, 4096 Kbps) with a 1280×720 resolution, and prepared them for DASH delivery using GPAC's MP4Box utility [116]. Each video segment has a duration of two seconds. Importantly, we use the `minimumUpdatePeriod` MPD directive to instruct each client to request an updated version of the manifest file every 5 s.

---

[1]We use the MP4Client video player shipping with the open-source GPAC framework [116], which we have appropriately modified to record video interruption statistics. We have also modified the ffmpeg H.264 decoder library to record picture quality data. We utilize these data for QoE assessment purposes (see Section 5.5.3).
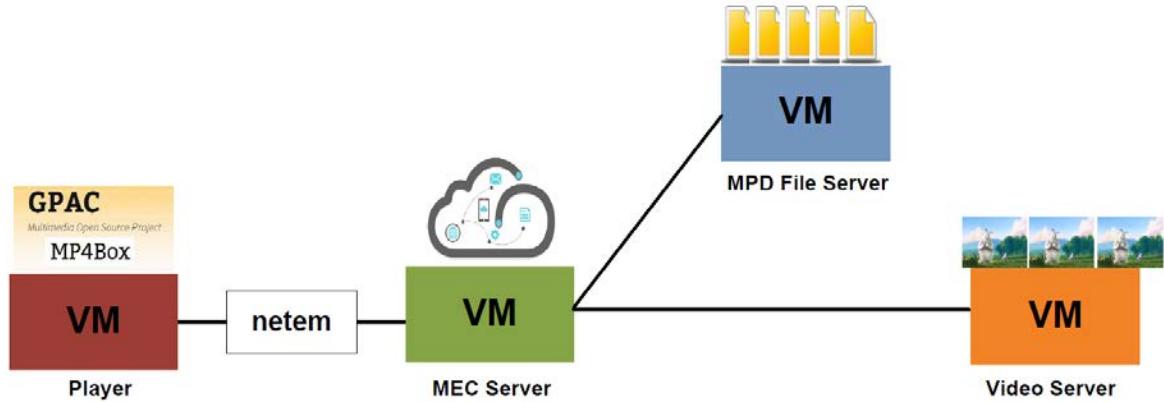
Figure 5.5 – Testbed Description

5.5.2   Candidate strategies

We compare three adaptation strategies: (i) A purely user-driven bitrate adaptation strategy, where each client uses the measured chunk download times to decide whether to switch to a more appropriate video representation (bitrate) to match the available bandwidth. This is the default mechanism applied by MP4Client, the video player we use in our tests. (ii) Our proposed MEC-assisted reactive (default) strategy (see Section 5.4), in which the network congestion state is returned by the MEC API as soon as the mobile edge proxy detects that the bandwidth demand or network conditions have changed. (iii) Our proposed MEC-assisted proactive strategy, which assumes that there is an accurate mechanism for predicting congestion and the available bandwidth. Note that in cases (ii) and (iii), the default bitrate adaptation scheme is still used at the client side to select among the available video representations present in the MPD file after the modifications carried out by the MEC proxy.

5.5.3   Performance metrics

We evaluate our system in terms of the average achieved QoE for all users at any time instant. QoE is defined in [117] by ITU-T and it is a purely subjective measure from the users' perspectives to describe how satisfied they are with the service. The QoE can be influenced by many factors such as network, client and terminal according to different services. Since HAS uses a reliable transport protocol such as TCP, the playout interruptions may occur due to the retransmissions after packet losses. Moreover, the quality of video, which is decided by encoding parameters (i.e. Quantization Parameter (QP)), can vary with bandwidth fluctuations. Therefore, the playout interruption and video bitrate change are the main factors which significantly degrade users' QoE in

HAS. The metric that expresses QoE is the Mean Opinion Score (MOS), i.e., the expected rating that a panel of users would give to the quality of the transmitted video in the 1-5 (poor-excellent) scale.

QoE is subjective, and we estimate it using objective, measurable video-service-level parameters (interruption statistics, encoding parameters), which we translate to MOS estimates using the Pseudo-Subjective Quality Assessment (PSQA) approach [118] to monitor the QoE by the users, automatically and accurately in real time. The PSQA methodology involves training a Random Neural Network (RNN) using data from subjective tests, where a set of parameters affecting quality is monitored and the ratings of users are recorded. The trained RNN classifier can then be applied to calculate the expected MOS for specific values of the input parameters.

In this work, we use the PSQA tool developed by Singh et al. [58] for H.264/AVC-encoded HTTP video quality estimation. This tool produces a single QoE estimate for a 16-second video window. During video playout, we maintain an exponentially-weighted moving average of the instantaneous MOS according to the following equation:

$$MOS_{i+1} = 0.2MOS_i + 0.8s_{i+1},$$

where MOS_i is the moving average calculated up to window i and s_{i+1} is the MOS sample calculated for the $(i+1)$-th window.

The input parameters for the QoE estimation tool are the following:

- The number of interruptions in the 16-second window.

- The average and the maximum interruption duration.

- The average value of the Quantization Parameter (QP) across all picture macroblocks in the measurement window. QP represents the transformation coefficients when encoding the video and it controls the compression level. The higher the QP value is, the more video information is lost and the lower the video bitrate and quality become. As per the H.264/AVC standard [119], QP values range from 0 to 51.

These parameters need to be appropriately normalized (see [58] for details).

According to various studies [120, 121], switching among video representations also affects QoE. However, recent results [122, 123] indicate that the effects of frequent bitrate switching are less pronounced compared to other factors. For reasons of completeness, and since our PSQA model does not account for this factor, we also present bitrate switching statistics for each of the adaptation mechanisms we compare.

Finally, we evaluate our system in terms of fairness. One desirable property is to have all users (with the same service priority and with comparable signal conditions, as we assume in our tests) enjoy approximately the same video quality levels. Our fairness metric is Jain's Fairness Index (JFI) [124], which we measure at regular time intervals during our experiments. In our context, JFI is given by the following formula:

$$J_t = \frac{(\sum_{i=1}^n MOS_{i,t})^2}{n \sum_{i=1}^n MOS_{i,t}^2},\tag{5.1}$$

where $MOS_{i,t}$ is the QoE value calculated for the $i-$th user at time t and n is the total number of users simultaneously accessing the video service in a cell. J_t is bounded between 0 and 1; higher values indicate higher fairness.

5.5.4 Results

Each of our experiments starts with a period of 15 seconds during which 10 users join the video service following a Poisson process. We emulate the network load in two distinct periods: (i) From $t = 120$ s to $t = 240$ s, when the available bandwidth is limited to 10 Mbps, and (ii) from $t = 300$ s to $t = 420$ s, when it is limited to 1.8 Mbps. We repeat the experiment introducing load at the exact same time instances for the three candidate strategies (see Section 5.5.2). In the case of the proactive MEC-assisted strategy, we assume that the mobile edge has predicted the decrease in the available bandwidth 5 s in advance.

Figure 5.6 presents the evolution of the average QoE achieved across all users. Each point represents the average of 5 iterations of the same experiment. The shaded regions represent periods when the load was introduced (available bandwidth amounting to 10 Mbps and 1.8 Mbps respectively). Our scheme brings tangible performance benefits, especially with the proactive MEC-assisted strategy. In this case, QoE improvements reach up to approximately 1 point in the 1-5 MOS scale.

Our approach also improves on fairness, as Figure 5.7 suggests. In the experiment, each point represents the average JFI value across 5 iterations of the same experiment. The shaded regions represent periods when the load was introduced (available bandwidth amounting to 10 Mbps and 1.8 Mbps respectively). During the high-load period, the proactive MEC-assisted strategy brings an up to 10% increase in terms of JFI.

Our improvements are mainly due to the fact that our algorithm reduces the frequency and the duration of playout interruptions (Table 5.1): When applied proactively, our MEC-assisted strategy leads to approximately 54% shorter and four times less frequent interruptions compared to the default client-driven adaptation mechanism.
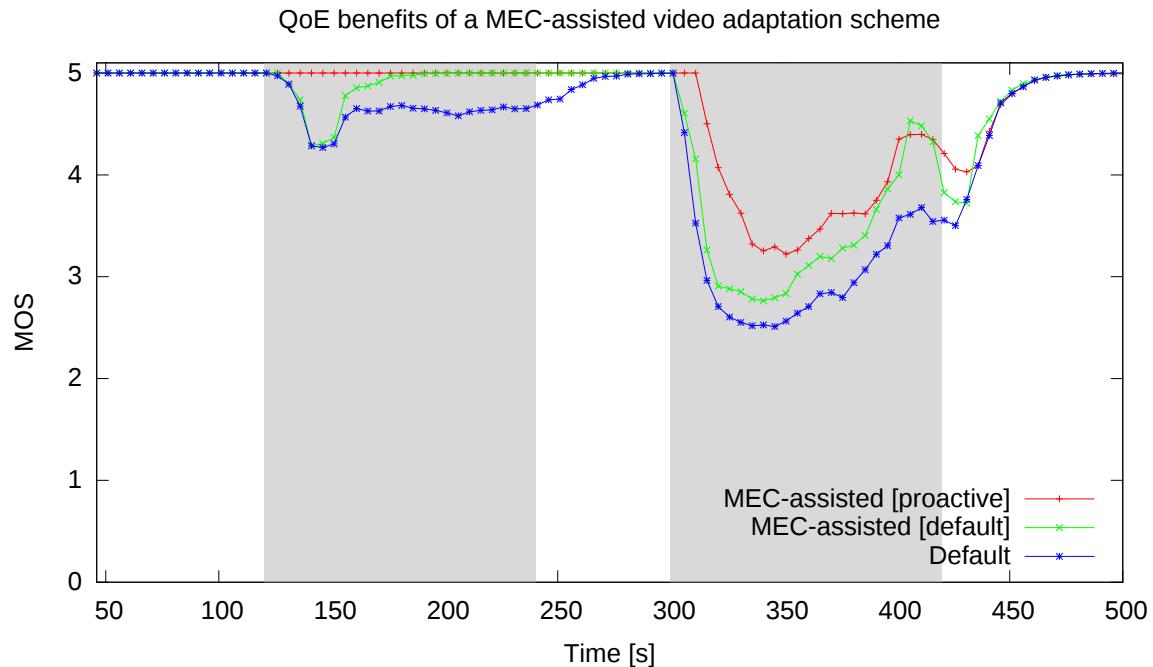
Figure 5.6 – MOS across all 10 Users.
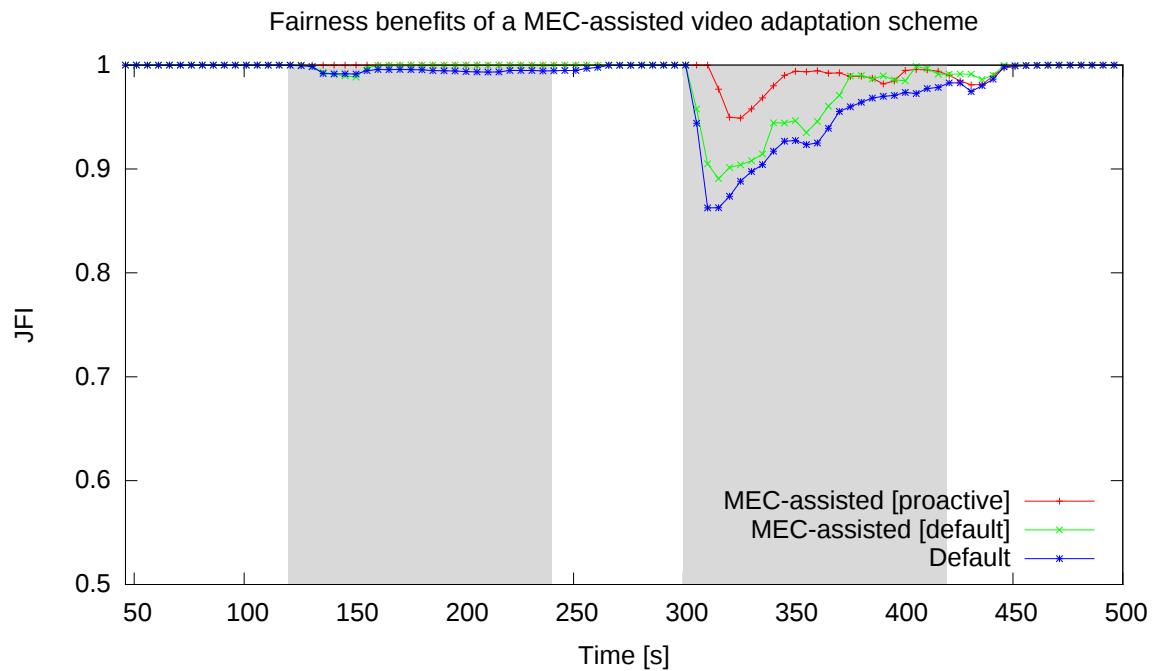
Figure 5.7 – Fairness Achieved for Each Adaptation Strategy in Terms of Jain's Fairness Index.

It should be noted that our mechanism also results in fewer bitrate switches, as shown in Table 5.1. In the measurement, each reported value is the mean across all 10 clients and averaged over five experiment iterations where each client watched the full 9-minute video. This is also attributable to the fact that in view of congestion, after our mechanism takes effect, there are fewer available video representations to select from.

5.6 Summary

In this chapter, we focused on enhancing the performance of Dynamic Adaptive Streaming over HTTP (DASH) in a mobile network environment. To address this challenge, we introduced a novel architecture based on Multi-access Edge Computing (MEC). The proposed adaptation mechanism, running as a MEC service, could modify DASH video manifest files in real time, responding to network congestion and dynamic demand, thus driving clients towards selecting more appropriate quality/bitrate video representations. Experiments with our scheme on a virtualized testbed we have developed demonstrated its Quality of Experience (QoE) benefits compared to traditional, purely client-driven, bitrate adaptation approaches; in the face of congestion, our scheme notably improved both on the achieved Mean Opinion Score (MOS) and on fairness. The strength of this approach is threefold. First, fairness is achieved without explicit communication among clients and thus no significant overhead is introduced into the network. Second, the system of coordination proxies is transparent to the clients, that is, the clients do not need to be aware of its presence. Third, the HAS principle is maintained, as the in-network components only provide the clients with new information and suggestions, while the rate adaptation decision remains the sole responsibility of the clients themselves.

Table 5.1 – Interruption and bitrate switching statistics.

	Interruption duration (s)	Interruptions/m	Switches/m
MEC-assisted (proactive)	0.58	1.04	6.27
MEC-assisted (default)	1.10	1.88	10.90
Default	1.26	4.01	22.45

Chapter 6

A Multi-access Edge Computing-Assisted Streaming Architecture for Wireless Heterogeneous Networks

6.1 Introduction

With the proliferation of advanced mobile devices such as tablets and smartphones, video streaming especially HTTP-based adaptive streaming is gaining popularity and becoming one of the major Internet services for mobile consumers. Since contemporary mobile devices are usually equipped with multiple radio interfaces (e.g. 3G/LTE, Wi-Fi, etc.), exploiting such diverse wireless access networks to improve the video streaming services is promising. In case of congestion on one wireless network, a user should be able to switch to another network smoothly, without any QoE degradation. Similarly, from the operators' perspective, it is essential to guarantee the QoE of mobile users and maximize network performance by assisting the users to make the network selection decision for the provided video streaming service. However, optimizing the video quality selection strategy for video streaming over multiple wireless networks, considering the video service's requirements, the wireless channel profiles and the costs of the different links remains an open issue. Therefore, there is clearly a need for an application which can make both video quality selection and network selection in an optimal manner.

In Chapter 5, we have proposed a Multi-access Edge Computing (MEC)-based architecture for improved HAS video delivery in the context of one access network. In this chapter, we extend it to propose a MEC-assisted solution to maximize QoE and fairness of mobile users by combining video quality and network selection in a multi-access heterogeneous network. We have investigated in Chapter 3 the state of art which focused on improving DASH performance. Existing approaches can be classified as either client-based, proxy-based or SDN-based schemes. The proposed solution belongs to the proxy-based approach, as we consider using a mobile edge cloud for improving DASH performance. However, compared to most existing work, which considers only a single access network [135], the proposed solution assumes a multi-access heterogeneous network environment. By modifying the MPD files, in real time, based on the network information collected by MEC service, our approach can achieve an improved overall QoE and a refined fairness for the clients in case of network congestion. One of the main advantages of our algorithm is its compatibility with the actual Dynamic Adaptive Streaming over HTTP (DASH) standard, as DASH client remains unaware of the proposed procedure. Indeed, the final quality level selection decision is still made by the clients, which corresponds to the current DASH client standard specifications. Leveraging the existence of multi-homed devices has the advantages of better resiliency to changes in network conditions. It also allows going further in access network optimization with the objective to deliver a more stable video experience. Therefore, it is promising to exploit such diverse wireless access networks to improve the quality of mobile video streaming services.

6.2 Problem formulation

6.2.1 QoE estimation

Our goal is to maximize the sum of the Quality of Experience (QoE) of all clients. QoE is a subjective measure and the metric often used to quantify it is the Mean Opinion Score (MOS), i.e., the expected rating that a panel of users would give to the quality of the transmitted video in the 1-5 (poor-excellent) scale. In the problem formulation, we utilize the tool developed in [132] to translate the video bitrate received by the clients into MOS values.

6.2.2 Model description

We assume a network setting where n users, each having m available networks, access a video service where h distinct video representations are available. The binary variable $x_{i,j,k}$ represents whether user i accesses video representation k over network j, where $i \in [1,n]$, $j \in [1,m]$, $k \in [1,h]$, and b_k and E_k denote the video bitrate of representation k and its translated MOS value, respectively.

To achieve this objective, the MEC application, which is instantiated on the virtualisation infrastructure of the mobile edge host [133], should find the maximum permitted video quality, which should be viewed by each client and the network over which it can be downloaded. Our model assumptions are summarized below:

- At any time, each user downloads a single video representation.

- At any time, each user uses a single radio interface to access the video service[1].

- Each access network has a specific total capacity C_j, $j \in [1,m]$.

- Each individual radio link between the user's terminal i and the network j has a specific capacity l_i^j, which is decided by the modulation, codec, etc.

We model the problem of user-video representation-network assignment as a Binary Integer Linear Programming:

$$\text{Maximize} \quad \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{h} x_{i,j,k} E_k \tag{6.1}$$

$$\text{subject to} \quad \sum_{i=1}^{n} \sum_{k=1}^{h} x_{i,j,k} b_k \leq C_j, \forall j \in [1,m] \tag{6.2}$$

$$x_{i,j,k} b_k \leq l_i^j, \forall i \in [1,n], \forall j \in [1,m], \forall k \in [1,h] \tag{6.3}$$

$$\sum_{k=1}^{h} \sum_{j=1}^{m} x_{i,j,k} \leq 1, \forall i \in [1,n] \tag{6.4}$$

$$x_{i,j,k} \in \{0,1\}, \forall i \in [1,n], \forall j \in [1,m], \forall k \in [1,h]. \tag{6.5}$$

This model has the objective of maximizing the overall QoE perceived by all the clients. The set of constraints in (6.2) guarantees that the capacity of network j is not exceeded by the assignment, since, for each network, the sum of allocated bitrates to all users is constrained by the capacity of the network (C_j). Constraints (6.3) ensure

---

[1]Multi-path through multiple access networks is not considered in this paper.

that a video representation is not assigned to user i over network j if its bitrate cannot be accommodated by the capacity of the user's radio link. A guarantee that a user is assigned *at most* one video representation is given by (6.4). The latter set of constraints also guarantee that each user uses *at most* one network for the video service (since he receives at most one video representation).

Deriving exact optimal solutions for large problem instances is prohibitively expensive computationally. We therefore propose a low-complexity heuristic algorithm in the following section. We show it to be effective by comparing its output with the optimal solution derived using the IBM ILOG CPLEX Optimizer [134] in Section 6.5.1.

6.3 Algorithm design

Due to the complexity of the problem, we develop a heuristic algorithm. Its goal is to derive a solution that maximizes network wide proportional fairness while attempting to maximize the QoE across all the clients (we want to avoid that some clients have the very high video quality, whereas other clients have very low video qualities). The proposed solution takes into account the number of active users' equipment (UE), the video quality, and the available bandwidth in each network. The main idea is to fill the network bandwidth by increasing the video quality gradually in a water-filling like manner.

The notations used in the proposed algorithm are:

- $\mathcal{U} = \{1, 2, \cdots, n\}$: The set of identifiers of the mobile users intending to view a video.

- $\mathcal{N} = \{1, 2, \cdots, m\}$: The set of identifiers of the access networks the mobile users can connect to.

- q_k, where $k \in [1, h]$: The video quality which can be rendered at the mobile users' level. We denote q_1 as the lowest quality and q_h as the highest quality. The function $b(q_k)$ denotes the required bitrate for the video quality k.

- $\mathcal{S}_j$, where $j \in [1, m]$: This variable, which contains a set of tuples (i, q_k), indicates the list of users associated with network j and their corresponding quality. The function $TotalBR(\mathcal{S}_j)$ and $RemainBR(\mathcal{S}_j)$ denote the overall allocated bitrate to all the users in network j and the remaining bitrate, respectively.

- $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_m\}$: This represents the global solution containing all the networks and their corresponding users and video quality, which will be the final output of the proposed heuristic algorithm.

- B_j: Current available bandwidth in network j.

- p_j: Current number of assigned users in network j.

The algorithm consists of the three following steps:

Step 1: The main idea behind this step (*Network Selection*), is to associate temporarily the users with an access network and the lowest video quality q_1 (see algorithm 2). First, the algorithm finds the set $\mathcal{N}_c$ of networks in $\mathcal{N}$ having enough bandwidth to transmit the lowest video quality. Then, it decides which network in $\mathcal{N}_c$ is the most suitable for each client. In order to place efficiently the users among the access networks while keeping the same level of enhancement, the users are placed successively in the network with the best fair share (see lines 12 and 14). After each placement, the network with not enough resources is automatically removed from the list $\mathcal{N}_c$, as shown in lines 17 and 18. Note that the algorithm ends when all users are assigned to an access network or when $\mathcal{N}_c$ is empty, which means that there are not enough resources to serve all users. In this last case, the *Block*($[i..n]$) function allows barring, for a certain period of time, all users from i to n.

Step 2: The objective of this step (*Local optimization of resources*), is to increase fairly the video quality of the users in each network (see algorithm 3). It is based on the relaxation principle, in which the user with the worst quality is selected, using the function GetWorse(), and gradually assigned a better quality, when possible. Indeed, a better quality is assigned to a user only when it can handle it (see line 7).

The algorithm ends when there is no user to improve. This may happen when $\mathcal{S}'_j = \{\phi\}$, or when the worst user has the highest quality q_h, or when there are no available resources to allocate. In this last case, the remaining amount of resources, given by function RemainBR(), is too small for a quality improvement.

Step 3: After having assigned the different users to a predefined access network with the best possible quality, this step "*Global optimization of resources)*" optimizes further the quality and network assignment (see algorithm4). The main idea here is to first select the user with the lowest video quality using the function FindUserMin(), which returns a user Id equal to 0 when there is no user's quality to improve. In this case the optimal solution $\mathcal{S}$ is returned. Otherwise, the network with the maximal remaining bandwidth is selected as a potential destination, using the function MaxNetRemainBR(). If there are enough resources to host the selected user, it is assigned to this network. Otherwise,

---

Algorithm 2 Network Selection (Step 1)

Init: $\mathcal{S}_j = \{\phi\}$, $\mathcal{N}_c = \{\phi\}$, $p_j = 0$, **for all** $j \in \mathcal{N}$;

 1: **for all** $j \in \mathcal{N}$ **do**
 2:     **if** $B_j \geq b(q_1)$ **then**
 3:        $\mathcal{N}_c = \mathcal{N}_c \cup \{j\}$;
 4:     **end if**
 5: **end for**
 6: **for all** $i \in \mathcal{U}$ **do**
 7:     **if** $\mathcal{N}_c = \{\phi\}$ **then**
 8:        Block($[i..n]$);
 9:        **return** $\mathcal{S}$;
10:     **else**
11:        **for all** $j \in \mathcal{N}_c$ **do**
12:          $Avg_j \leftarrow \frac{B_j}{p_j + 1}$;
13:        **end for**
14:        $net = \underset{j \in \mathcal{N}_c}{\arg\max} \;\; Avg_j$;
15:        $\mathcal{S}_{net} = \mathcal{S}_{net} \cup \{(i, q_1)\}$;
16:        $p_{net}++$;
17:        **if** $B_{net} < (p_{net} + 1) \times b(q_1)$ **then**
18:          $\mathcal{N}_c = \mathcal{N}_c - \{net\}$;
19:        **end if**
20:     **end if**
21: **end for**
22: **return** $\mathcal{S}$;

---

the algorithm ends by returning the optimal solution $\mathcal{S}$. Note that this process repeats until no increment on video quality is possible, which guarantees the convergence of the proposed greedy algorithm.

6.4   Architecture description

In the proposed architecture illustrated in Figure 6.1, the users can connect to the mobile core network through the different available access networks. The architecture is composed of the following entities:

User Equipment (UE) The UE is equipped with multiple radio interfaces (i.e. multi-homed terminal). When it receives an MPD file, the UE downloads video chunks from the URLs indicated in the file. Since the video streaming is over HTTP, session continuity is guaranteed when the UE switches across access networks.

Algorithm 3 Local optimization of resources (Step 2)

Init: $\mathcal{S}'_j = \mathcal{S}_j$;

 1: **while** $\mathcal{S}'_j \neq \{\phi\}$ **do**
 2:     $(i, q_k) = \text{GetWorse}(\mathcal{S}'_j)$;
 3:     **if** $q_k \geq q_h$ **then**
 4:         **return** $\mathcal{S}_j$;
 5:     **end if**
 6:     **if** $(\text{RemainBR}(\mathcal{S}_j) > b(q_{k+1}) - b(q_k))$ **then**
 7:         **if** $l_i^j \geq b(q_{k+1})$ **then**
 8:             $\text{Assign}(j, i, q_{k+1})$;
 9:         **else**
10:             $\mathcal{S}'_j = \mathcal{S}'_j - \{(i, q_k)\}$;
11:         **end if**
12:     **else**
13:         **return** $\mathcal{S}_j$;
14:     **end if**
15: **end while**
16: **return** $\mathcal{S}_j$;
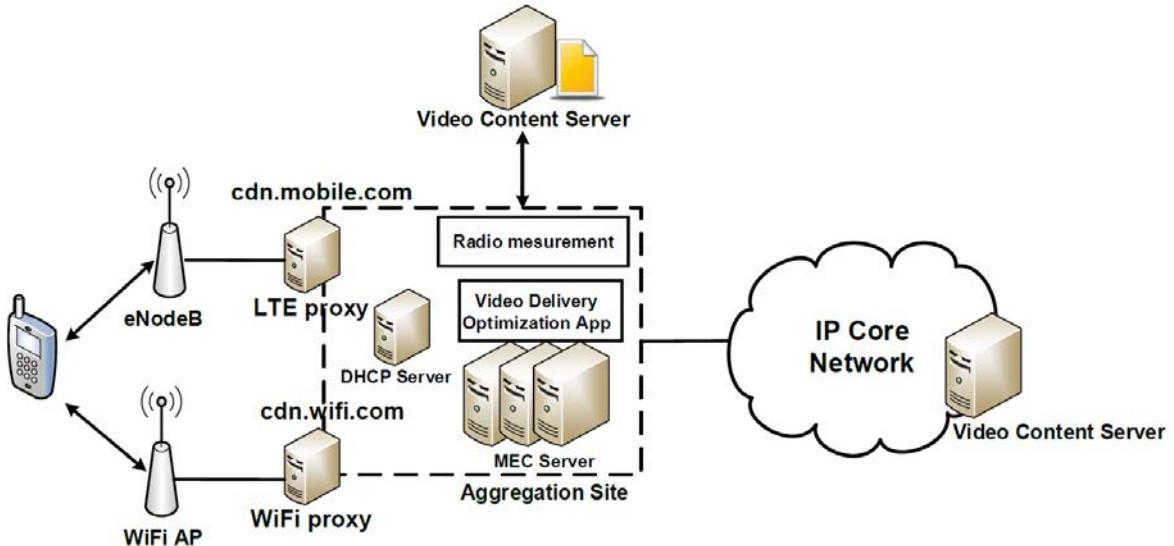
Figure 6.1 – An Architecture for Mobile Adaptive HTTP Streaming with MEC in Heterogeneous Network

Access Point (AP) The APs consist of different possible access technologies such as an eNodeB, a Wi-Fi AP, a Femtocell AP, etc. All these APs are managed by the network operator and they are connected to a common mobile backhaul network.

MEC server The MEC server, which is managed by the network operator, hosts various MEC applications.

Algorithm 4 Global optimization of resources (Step 3)

1: **while** true **do**
2:     $(net_s, i, q_k) = \text{FindUserMin}(\mathcal{S})$;
3:     **if** $i = 0$ **then**
4:         **return** $\mathcal{S}$;
5:     **end if**
6:     $net_d = \text{MaxNetRemainBR}(\mathcal{S} - \{net_s\})$;
7:     **if** $(\text{RemainBR}(net_d) < b(q_{k+1}))$ **then**
8:         **return** $\mathcal{S}$;
9:     **else**
10:         $\text{Move}(i, q_{k+1}, net_s, net_d)$;
11:         Go back to **Step 2** for net_s;
12:     **end if**
13: **end while**

LTE and Wi-Fi proxies The proxies are in charge of intercepting and relaying the UE requests and the MPD server responses. They are identified by different domain names (e.g. cdn.lte.com and cdn.wifi.com), which are included in the representation field of the MPD file to indicate the path for transmitting video chunks.

Multi-homed DHCP server The multi-homed DHCP server serves multiple radio access networks. It configures different gateways towards the multiple available access networks, to direct chunk requests via the appropriate network interface.

MEC server-based video delivery optimization application This application is located in the MEC server. It runs an optimization algorithm, which can run periodically or be triggered when network congestion or QoE degradation are detected. The application periodically measures the network status (link capacities, the network load, the video encoded quality), using, among others, the Radio Network Information Service (RNIS) [133] MEC API. These metrics are then used to optimize the path and the video quality selection schemes with the objective to guarantee an optimized video delivery. The calculated result is then applied by transforming the original MPD file, which is sent by the MPD server located at the service provider.

MPD server The MPD server, which is managed by the content or the service provider, stores the original MPD files with the video encoding levels and the default chunk addresses on the video content server(s). It responds to the request of the MEC application for sending the original MPD file.

Video Content Server Video content server is managed by the content provider. It stores the original video files.

6.4.1 Message flow
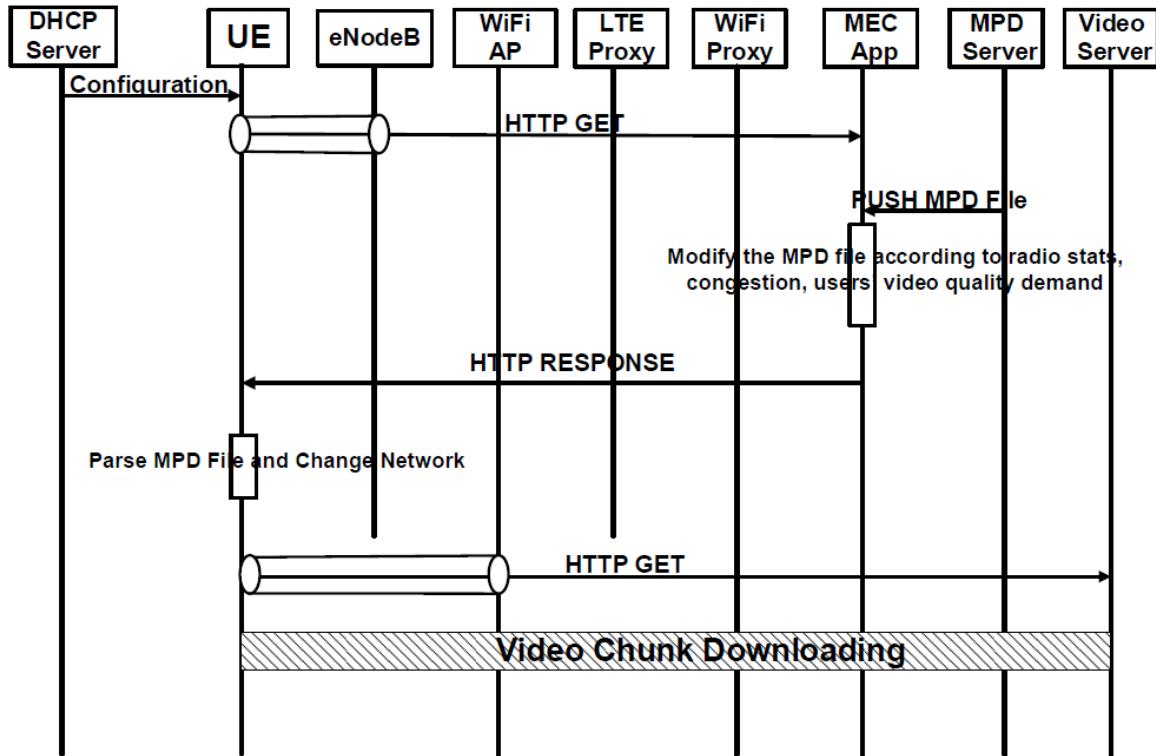
Figure 6.2 – Message flow for the video quality and network selection procedure

Figure 6.2 describes how UEs interact with the proxies and the MEC application to select the network and the video quality according to the received MPD file. In this scenario, the UE is already connected to the LTE network. To access the video streaming service, it first sends a request for the MPD file to the LTE proxy using HTTP, which the LTE proxy relays to the MPD server. The original MPD file is intercepted and modified by the MEC video delivery optimization application. The modified MPD, which contains the optimal path and video quality levels is sent back to the client via the LTE network. In the illustrated example, the optimal path is via the Wi-Fi link since the Wi-Fi proxy address is included in the URL written in the MPD file. Then the UE sends the request to the Wi-Fi proxy for downloading the video chunks and the Wi-Fi proxy relays each request to the video content server.

6.4.2 MPD File Updates

To apply our optimization solution, we need a mechanism so that the clients can receive updated versions of the MPD file. We achieve this in an MPEG DASH compliant,

receiver-driven and transparent way by exploiting the `minimumUpdatePeriod` attribute of the standard. When this attribute is present in the MPD file, the client is instructed to periodically request an up-to-date version of the file from the video server.

An MPD file example which is compatible with the DASH standard is showed as follows:

```xml
1 <?xml version="1.0"?>
2 <MPD profiles="urn:mpeg:dash:schema:mpd:2011" minBufferTime="PT1.500S"
      type="dynamic" timeShiftBufferDepth="PT0H0M0.000s"
      minimumUpdatePeriod="PT0H0M10.000S" maxSegmentDuration="PT0H0M2.000
      S" maxSegmentDuration="PT0H0M2.000S" profiles="
      urn:mpeg:dash:profile:isoff-live:2011">
3   <Period id="GENID_DEF" start="PT0H0M0.000S">
4     <AdaptationSet segmentAlignment = "true" bitstreamSwitching="true"
        maxWidth="1280" maxHeight="720" maxFrameRate="24" par="16:9"
        lang="und">
5       <SegmentTemplate initialization="http://@path/bbb_init.mp4"/>
6       <Representation id="1" mimeType="video/mp4" width="1280" height="
          720" frameRate="24" sar="1:1" startWithSAP="1" bandwidth="
          117301">
7         <SegmentTemplate timescale="12288" media="bbb_128$Number$.m4s"
            startNumber="1" duration="24576"/>
8       </Representation>
9       <Representation id="2" mimeType="video/mp4" width="1280" height="
          720" frameRate="24" sar="1:1" startWithSAP="1" bandwidth="
          238080">
10        <SegmentTemplate timescale="12288" media="http://@path/bbb_256$
            Number$.m4s" startNumber="1" duration="24576"/>
11      </Representation>
12    <Representation id="3" mimeType="video/mp4" width="1280" height="720
        " frameRate="24" sar="1:1" startWithSAP="1" bandwidth="487065">
13        <SegmentTemplate timescale="12288" media="http://@path/bbb_512$
            Number$.m4s" startNumber="1" duration="24576"/>
14      </Representation>
15    </AdaptationSet>
16  </Period>
17 </MPD>
```

We assume that three distinct video representations are available in the MPD file example. For each representation, the corresponding required transmission bitrate is indicated in the `bandwidth` field. This bitrate corresponds to the b_k variable in our system model. The `media` field indicates the address where the video chunks are stored, and the `path` attribute indicates the base URL where the content is stored. In our

approach, this is a domain name which is resolved by the DNS server as described in Figure 6.1. Thus, by appropriately modifying this path, the mobile users are directed to select the appropriate network interface to download video chunks.

By modifying periodically `bandwidth`, `path`, and `media` fields in real-time according to the proposed algorithm, the MEC application can control the maximum permitted video quality, which can be downloaded by mobile users, and their corresponding network interface. Therefore, the mobile users can always get an updated MPD file and select the optimal video quality and network.

6.5   Performance evaluation

6.5.1   Numerical analysis

We implement the model presented in Section 6.2.2 using the IBM ILOG CPLEX Optimizer. Our proposed heuristic algorithm is implemented in Python. For evaluating the scalability of this two approaches, we measure the execution time of the CPLEX QoE-aware solution and our heuristic algorithm. We vary the number of users from 10 to 100, and the video is encoded into 4 bitrates: 128, 256, 512, 1024 Kbps. Each user has two available networks, n_1 and n_2, with bandwidth b_1 and b_2, respectively. We generate 1000 different network configurations randomly (i.e., the combination of b_1 and b_2). The individual link capacity of these networks is fixed to a constant, which can support up to 1 Mbps. Each measurement is repeated 10 times. Then, we calculate the average execution time over the 1000 network configurations and the 10 repetitions to obtain the results presented in Figure 6.3. It can be noticed that the execution time of CPLEX QoE-aware solution increases exponentially with the number of clients, which indicates that it is not scalable. Our proposed heuristic, though, scales linearly.

We further compare the two algorithms in terms of our model's QoE-aware objective function (sum of QoE). The bandwidth b_1 of network n_1 varies from $100 Kbps$ to $50 Mbps$ and the bandwidth b_2 of network n_2 varies from $50 Kbps$ to $25 Mbps$. We define optimality as

$$\text{optimality}(\%) = \frac{QoE_{\text{heuristic}}}{QoE_{\text{CPLEX}}} \times 100 \tag{6.6}$$

Our results are shown in Figure 6.4, for 10, 50, and 100 clients. We can see from Figure 6.4a that from the beginning of the experiment, when the network is heavily loaded, optimality is equal to 100%, which indicates that the two algorithms derive equivalent solutions, reaching the same overall QoE. With the increase of network
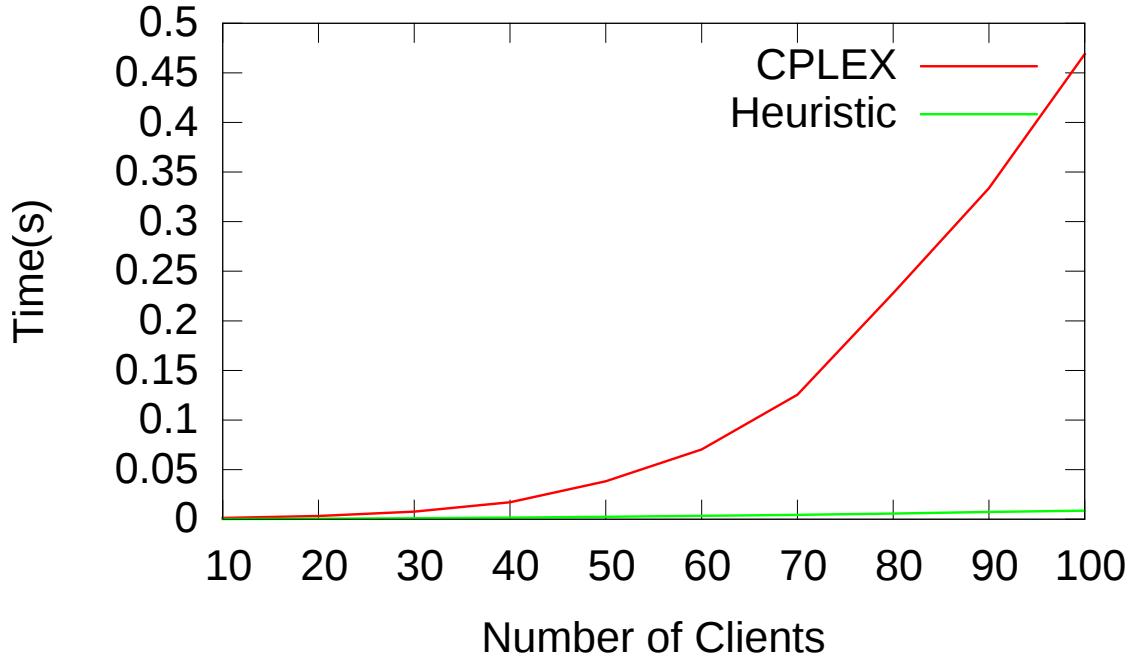
Figure 6.3 – Execution time with different number of users.

bandwidth, the optimality starts to degrade. The worst case across our experiments for the heuristic solution is approximately at 80% of the optimal (with 10 clients), which is a positive result. As network bandwidth continues to increase, the two algorithms allow obtaining the same overall QoE (the optimality metric converges to 100%) when the network is lightly or not loaded. Figure 6.4b shows the CDF of the optimality metric. In nearly 90% of all scenarios, our heuristic scores more than 91% of the optimal one. Based on our experiments, our heuristic algorithm approximates well the optimal solution while having an execution time which allows it to be applied to real MEC settings.

6.5.2 Testbed experiments

The evaluation is carried out with DASH clients accessing to two different networks. We have set up a local-area testbed of 9 Linux-based machines. Our setup is shown in Figure 6.5. One machine acts as the video server running the nginx [112] HTTP server, where video content is stored. For simplicity, the MPD server and the MEC application are hosted within this server.

The other 8 machines (DASH clients) and equipped with two Ethernet interfaces *eth*1 and *eth*2 to emulate multi-homed UEs with LTE and WiFi interfaces. Clients run
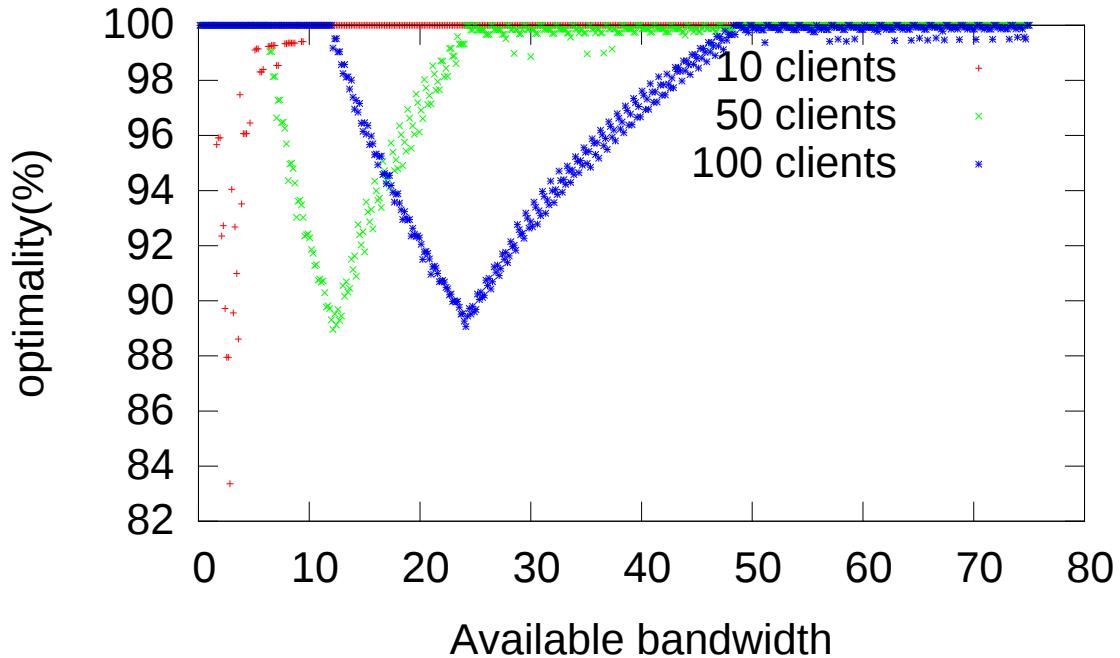
(a) Optimality with different number of users and bandwidths
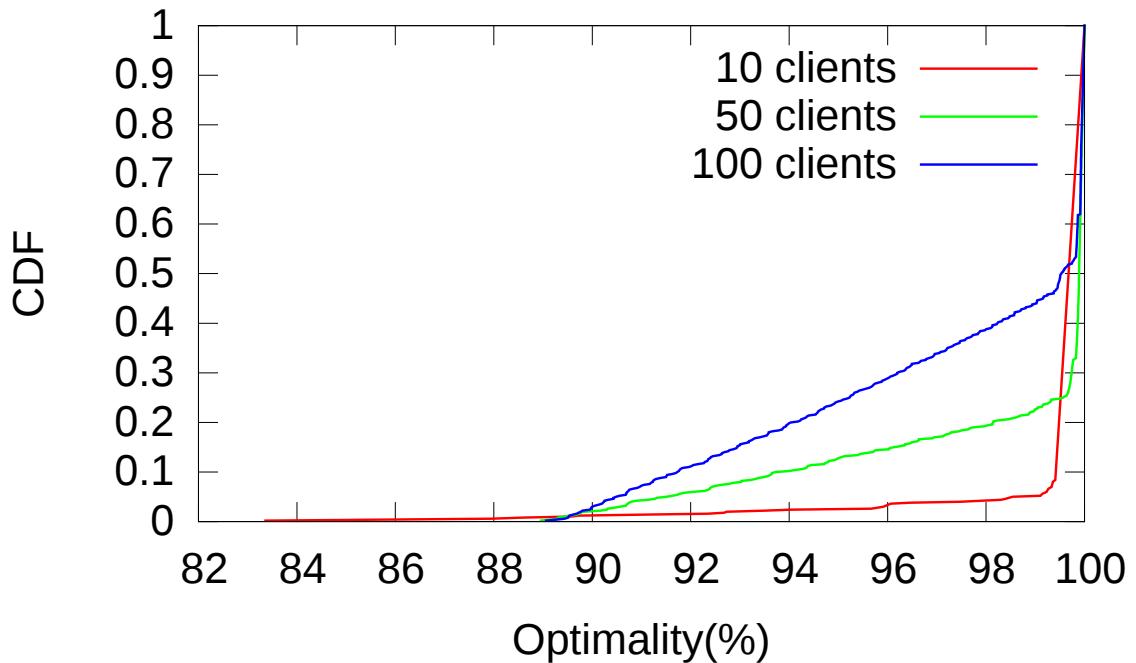
(b) CDF of optimality with different number of users

Figure 6.4 – Optimality Measurement
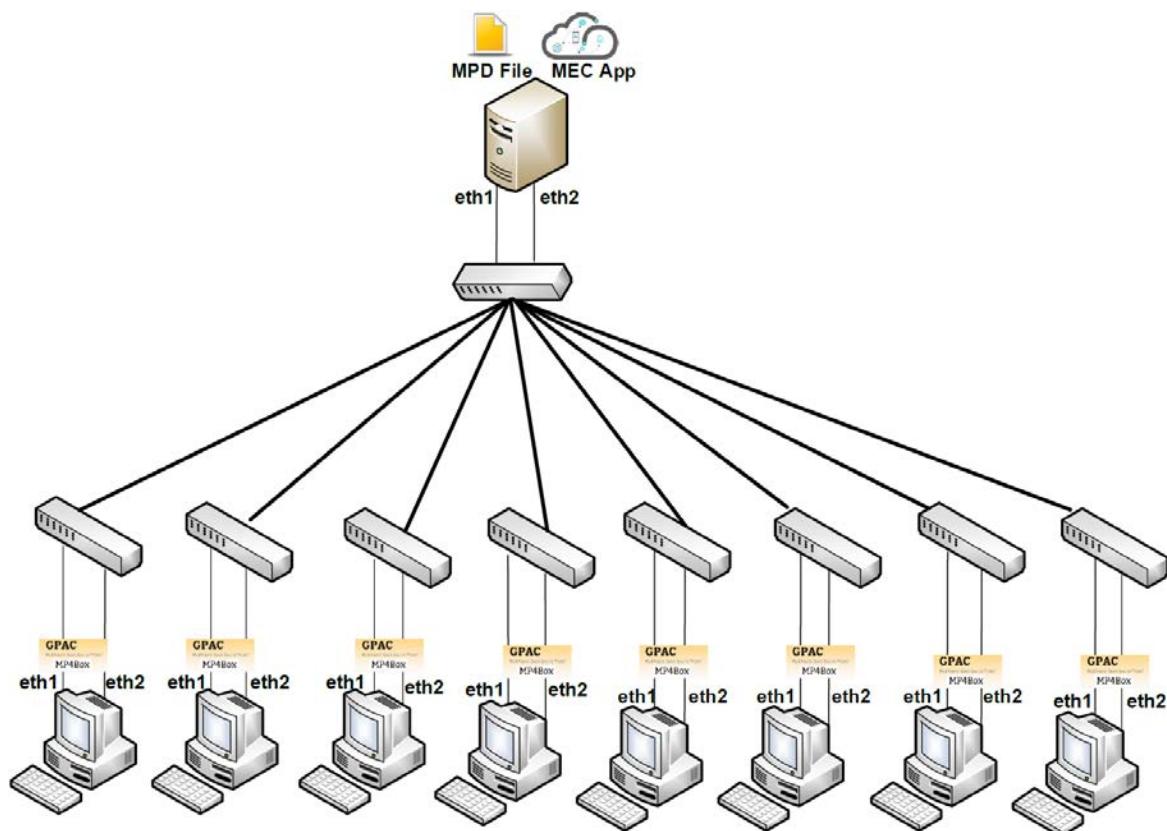
Figure 6.5 – Testbed implementation.

the MP4Client video player [116], which we have extended to record the Quantization Parameter (QP) and the video interruption statistics that are used for QoE assessment.

The MEC video delivery optimization application is aware of the current number of clients in the system, as well as the available bandwidth for each network. It intercepts client requests for MPD files and executes the adaptation algorithm according to its network awareness and the current client demand. It retrieves the requested MPD files from the source and appropriately transforms them to steer clients towards the best network and best video quality.

Each client accesses a 5 minute video sequence. In particular, using ffmpeg, we created different H.264/AVC representations of the Blender Foundation's "Big Buck Bunny" [115] open-source movie (with approximate bitrates of 128, 256, 512, 1024, 2048, 4096 Kbps) with a 1280×720 resolution, and prepared them for DASH delivery using GPAC's MP4Box utility [116]. Each video segment has a duration of 2 s. We use the `minimumUpdatePeriod` MPD attribute to instruct each client to request an updated version of the manifest file every 10 s.

6.5.3   Performance metrics

We evaluate our system in terms of the average achieved QoE for all users at any time instant. We estimate QoE using objective, measurable video-service-level parameters (interruption statistics, encoding parameters), which we translate to MOS estimates. In DASH, two important parameters that affect QoE are video encoding quality (determined by the Quantization Parameter (QP) used for encoding) and playout interruptions (i.e., buffering delays due to poor bandwidth). Several QoE estimation model are proposed considering these parameters. In our previous work [135], we used the PSQA tool developed by Singh et al. [58] for H.264/AVC-encoded HTTP video quality estimation. The PSQA methodology involves training a Random Neural Network (RNN) using data from subjective tests, where a set of parameters affecting quality is monitored and the ratings of users are recorded. The trained RNN classifier can then be applied to calculate the expected MOS for specific values of the input parameters. This approach allows having an accurate estimation of the QoE when the video quality is high. However, this model is not accurate with low video quality (i.e., high QP value). On the other hand, the tool introduced in [132] (denoted as *Vipeer*) is a purely QP-based QoE estimation model without taking playout interruptions into account.

To compare the performance of the different strategies considered in our experiments, we combine these two QoE estimation approaches under the assumption that

impairments due to playout interruptions and QP have an addictive effect on QoE. The additivity assumption is typical in the context of VoIP (see the E-model [136]), but it has also been proposed for video services [137]. We apply it in this work as it simplifies the comparison of the discussed strategies, but, as Hoßfeld et al. show [138], multi-factor QoE models and the assumptions therein require further study.

In this combined model, we calculate the two impairments independently. Since the maximum value of MOS is 5 (excellent perceived quality), the QP impairment is given by (6.7) using the VIPEER model and assuming no interruptions. The playout interruption impairment is given by (6.8), using the PSQA model assuming the perfect QP. The variable w denotes the measurement result from our experiment. The final MOS given in (6.9), is used as our QoE estimate.

$$I_{\mathrm{QP}} = 5 - MOS_{\mathrm{Vipeer}}(w) \tag{6.7}$$

$$I_{\mathrm{interruption}} = 5 - MOS_{\mathrm{PSQA}}(w) \tag{6.8}$$

$$MOS_{\mathrm{Final}} = 5 - I_{\mathrm{QP}} - I_{\mathrm{interruption}} \tag{6.9}$$

6.5.4 Candidate strategies

In our testbed experiments, since the number of clients is small, the running time for solving the optimization problem using CPLEX is only about several ms. Therefore, we can compare the optimal solution and the proposed heuristic to accurately evaluate the effectiveness of these solutions. We compare five adaptation strategies: (i) Wi-Fi first strategy. In this strategy, we do not control the MPD file adaptation. Clients decide on the video quality selection, and always select the Wi-Fi network when it is available. This is one of typical user behavior, since the monetary cost is always the clients' concern and Wi-Fi is often low cost. (ii) Random access strategy. In this strategy, the original MPD file is not modified by MEC application and the clients make the video quality adaptation decision. During playing the video, all the clients connect to one of the two networks randomly without any preference. This can happen in real network since they often switch between Wi-Fi and mobile network to get a better service. (iii) Our proposed heuristic strategy. In this strategy, the algorithm is running in the MEC server and modifies the MPD file periodically according to the result calculated by the algorithm. (iv) CPLEX Bitrate-optimal strategy. In this strategy, we modify the objective function of the model in Section 6.2.2 to maximize

the sum of the video bitrates received by all users. (v) CPLEX QoE-optimal strategy. In this strategy, the algorithm maximizes the overall QoE using IBM ILOG CPLEX Optimizer following the constraints presented in the formulated model.

6.5.5 Results

The experiment starts with an initialization period of 15 s during which 8 users join the video service following a Poisson process. Table 6.1 shows the constraints of aggregated bandwidth in the LTE and Wi-Fi networks. For simplicity, we assume that the capacity of each individual radio link (i.e. LTE and WiFi) of users is constant at 1 Mbps. We emulate the network load from $t = 90s$ to $t = 210s$, when the available bandwidth for LTE and Wi-Fi is limited to 6.4Mbps and 3.2Mbps, receptively. We repeat the experiment with the same load at the exact same time instances for the five candidate strategies (see Section 6.5.4). Each experiment is repeated 5 times.

Figure 6.6 presents the evolution of the overall QoE of 8 users with the five candidate strategies. Each point represents the average of 5 iterations of the same experiment. Users have the worst MOS when they apply the Wi-Fi-First strategy in case of heavy network loads. This is because no matter whether the Wi-Fi network is congested or not, all the clients will connect to the it network when it is available. Therefore, they will compete for the limited bandwidth, which will result in video quality degradation and interruptions. The random strategy has better performance since they switch between the two networks frequently, however, the MOS is still not optimal. Although the CPLEX Bitrate-optimal mechanism maximizes the overall transmission bitrates and the bandwidth utilization of the two networks, QoE is not optimized since this strategy does not consider fairness of users. The overall QoE with our heuristic reaches approximately the optimal QoE for all users, which validates the numerical evaluation in section 6.5.1.

We further compare these strategies in terms of fairness, and Jain's Fairness Index (JFI) [124] in particular. In this context, JFI is given by the following formula:

$$J_t = \frac{(\sum_{i=1}^{n} MOS_{i,t})^2}{n \sum_{i=1}^{n} MOS_{i,t}^2}, \tag{6.10}$$

where $MOS_{i,t}$ is the QoE value calculated for the $i-$th user at time t and n is the total number of users simultaneously accessing the video service in a cell. J_t is bounded between 0 and 1; higher values indicate higher fairness.

Figure 6.7 shows that the QoE-optimal and heuristic strategies have achieved higher *JFI*, since the algorithms will remove some high-bitrate representations to avoid competitive behavior of users in case of network congestion. The JFI in Wi-Fi first strategy decreased approximately 20% compared to the previous strategies when the network is under heavy loads, which is due to the frequency and the duration of playout interruptions and bitrate switching.

Table 6.1 – Bandwidth Configuration

	0-90 (s)	90-210 (s)	210-300 (s)
LTE	100 Mbps	6.4 Mbps	100 Mbps
WiFi	50 Mbps	3.2 Mbps	50 Mbps

Figure 6.6 – QoE benefits of a MEC-assisted video adaptation scheme.

6.6 Summary

In this chapter, we focused on enhancing the performance of dynamic adaptive streaming over HTTP in the context of wireless heterogeneous network environments. Since network and video quality selection are key factors that impact user experience, we proposed a MEC-based architecture and provided a binary integer linear programming
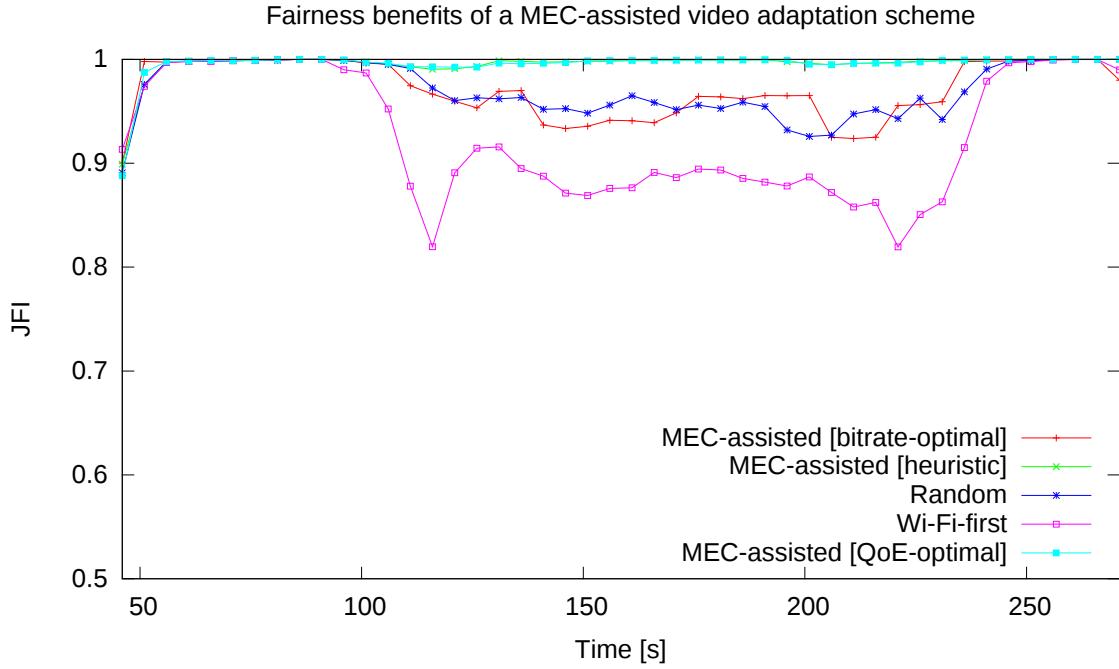
Figure 6.7 – Fairness benefits of a MEC-assisted video adaptation scheme

formulation to the problem of network and video quality selection. To use the proposed solution for realistic use cases, we introduced a heuristic to solve it efficiently. The proposed solution can be built into a MEC service that mediates mobile video content delivery, without requiring any modification at the client or the content provider ends. This service uses transparent and standards-compliant techniques for QoE-optimized video streaming, responding to network congestion and dynamic demand. Our testbed experiments demonstrated that our proposed algorithm can achieve close-to-optimal performance in terms of overall QoE and fairness.

Chapter 7

Conclusion and Future Work Directions

7.1   Summary

Continuous evolution of mobile Telecommunication network technologies from 1G to 4G will continue with the expected standardisation and deployment of 5G technologies in a very near future to face the tremendous growth in mobile data traffic. Current network infrastructure is characterized by its heterogeneity, where various types of access network technologies such as LTE and Wi-Fi coexist. Since these access technologies have different characteristics such as network capacity and coverage, the selection of the appropriate access network becomes a critical issue. On the other hand, video service is dominating the mobile network trafic, which creates a huge amount of data over the current heterogeneous network resulting in frequent network congestion occuring. Since mobile devices are alredy equipped with different network interfaces, which allows them to connect multiple networks, it is very challenging for network operators to avoid network congestion while guaranteeing the QoE of mobile video users. Therefore, in this thesis, we addressed two critical issues in heterogeneous network environments: network selection and performance improvement of DASH. We gave an analysis of the state of the art for two aspects and identified their limitations and proposed novel approaches. In this chapter, we summarize our major contributions and discuss future research directions.

7.1.1 Major Contribution Synopsis

Control Theory-based Network Selection Scheme in Heterogeneous Network Environment

The first part of the contributions focused on the network selection issue in heterogeneous network environments. Firstly we presented the problem analysis by implementing a simulator with ns-3 to explore the impact on the network performance using oversimplified network selection strategy. With the simulation results, we noticed that each network had an ideal workload to be reached for the optimized performance. Then we defined the method to solve the network selection problem: determining the ideal workload for each network and designing an admission controller to make the number of sessions converge to this workload. In the first contribution, we started by describing the network selection model for LTE and Wi-Fi interworking with one traffic class. Then we analyzed the controllability and observability of the model and proved that it was possible to design an LQR controller to control the system by regulating the parameters of the system model. We validated the controller analytically with Matlab and evaluated the performance of the controller with different configurations. In the second contribution of this part, we extended the model presented in the first contribution to describe the network selection scenario in a multi-access network with multi-class traffic. With the same principle, the LQR controller was designed and validated by Matlab simulation. Then we discussed the integration of proposed controller within a 3GPP standardized network using ANDSF module and we introduced the admission probability which was sent to mobile users when they made network selection decision. Finally we implemented the real scenario with admission probability contralto for the network selection and the simulation results revealed that the network workload could be always kept at the ideal level in case of network congestion.

Performance Improvement of HTTP Adaptive Streaming in Heterogeneous Network

The second part of the contributions in this thesis addressed the performance issue of DASH, one of the standard implantation of HAS in the heterogeneous network. We proposed to apply the emerging standard MEC enhance the performance of DASH. Firstly, we gave a brief introduction about MEC and its current application. We noticed that applying MEC for implementing DASH were not studied enough in the current state of the art. Therefore, in the first contribution, we introduced a innovate DASH architecture based on the MEC for the mobile users who have only one access

network. The idea was to develop an intelligent adaptation algorithm, running in the MEC server, to update the MPD file in real time to drive clients towards selecting more appropriate video representations before delivering it to mobile users according to the collected network information. We described the functionality of each node in this proposed architecture and discussed two modes for implementing it. Then we detailed the adaptation algorithm and deployed our scheme in a virtualized testbed. The experiment results showed that fairness among mobile users is guaranteed and the overall perceived QoE is improved compared to the purely client-driven, bitrate adaptation approaches under congested network situation. The second contribution of this work was to extend the MEC-based architecture to optimize the performance of DASH for the users with multi-homed devices in a multi-access heterogeneous network environment. In this scenario, mobile users can download the video chunk with one of their activated network interfaces thus the problem became to select the appropriate video quality to play with and best network to download. We formulated the problem as an optimization problem specifying the objective and constraints with the aim to maximize the total QoE of all the clients. IBM CPLEX tool was a good candidate to solve this problem. However, we measured the scalability of this tool and noticed the concern when the number of clients increased. Thus, a heuristic algorithm was proposed to approximate the optimal solution with low complexity. Then we presented the new involved network entities in the MEC-based architecture and the call flow for the implementation details. Then we make a performance comparison between the two approaches and the results showed that the proposed heuristic algorithm saved much more computation time than the CPLEX solution when the number of clients was increasing. The heuristic solution performed less optimal, however, the optimality loss was not more than 20% compared to the CPLEX solution, which as acceptable. Finally, we conducted experiments to measure the DASH performance of five strategies including the CPLEX and heuristic solution on an emulated testbed. The results revealed that the proposed heuristic algorithm performed significant improvement on the overall MOS and fairness by making the video quality and network selection when the network was in a congestion situation.

7.2   Future Work Directions

In our future work, we identify some interesting research directions regarding to our contributions in this thesis.

7.2.1 Extension of the Control Theoretic Model for Network Selection Problem

The proposed model can be extended to support the following scenarios:

- In our proposed control theory based network selection model, we consider a desired number of sessions to derive the targeted network occupancy. However, since each session can have a different radio link quality and thus consume different network resources, a better objective should be to take into account other parameters such as the throughput to complete our model.

- Our model assists in making network selection decision before mobile users start their session. After selecting a network, they will steer to this network until the end of the session. However, when a burst of packets is injected into the network chosen by certain mobile users, they will intend to handoff to other networks. Therefore, the handoff behavior should be considered in our control theory based model and the system characteristics such as controllability and observability should be evaluated to design a new kind of controller.

- Our model introduces the blocking probability as a cost to keep the network workload at the desired occupancy. For a scenario where a prioritized traffic class exists, the blocking probability could be canceled for high priority traffic class. Therefore, this model should be modified and the impact on the controllability of the system should be studied.

7.2.2 SDN and NFV in MEC-based Architecture for Streaming Service

We will explore the new networking paradigms SDN and NFV to enhance our proposed MEC-based architecture for more agility and flexibility. Since the SDN controller can provide a global view of available resources and network topology information to the network application such as MEC application via the northbound API, it will be very promising to combine the SDN and NFV with the MEC architecture to deliver the high performance streaming service over the future mobile network. Moreover, We will leverage the NFV technology to implement the architecture in a fully virtualized environment to evaluate the scalability of our solution in a large scale test.

7.2.3   Extending the proposed optimization Model

The optimization model developed in Chapter 6 focuses on maximizing the received bitrate of all the clients by following the network bandwidth constraints. In the reality, other attributes should be considered to extend this model. For example, mobile users may concern more about their budget on the data consumption. In this case, the network pricing model should be studied and objective function and constraints in the optimization model should be adapted to support this scenario. Another important characteristic is the battery level of mobile devices and different energy consumption using different network interfaces. For example, when the battery level is low, mobiles users often switch off the 4G interface even if the 4G network can deliver better video quality. Some mobile users may pay for the premium profile to have a guaranteed video service, therefore, users' classifications and traffic prioritization should be added in the extended model.

7.2.4   Future implementation in a real MEC environment

For evaluating our proposed optimization model, we implemented a testbed to emulate the MEC environment and the radio link. The next step is to deploy the solution in a real MEC infrastructure to measure its efficiency. The practical aspects such as how to perform radio information collection, available network bandwidth measurement and prediction should be further studied. As MEC is built on a virtualized environment, the computing resources such as CPU and memory utilization to perform the optimization process is an interesting perspective to investigate.

7.2.5   Video streaming performance improvement using control theory

Since we have established a control theory-based model to solve the network selection problem in heterogeneous network environments, applying this model to the MEC-based architecture to improve the DASH performance in a multi-access network will be a very interesting research direction. Thus, the model should be adapted to support video quality selection. The controller can be implemented in the MEC server as an application. Moreover, MEC server can get more network information such as throughput, delay, network occupancy which can be the feedback of the controller, the model can be enriched to support more variables. The comparison with other DASH

performance improvement approaches will be studied to figure out the advantages of the proposal.

List of publications

- International Conferences:

  1. Houssem-Eddine Chihoub, Shadi Ibrahim, Yue Li, Gabriel Antoniu, María Pérez, Luc Bougé, **'Exploring Energy-Consistency Trade-offs in Cassandra Cloud Storage System,"** in 27th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2015), Florianopolis Santa Catarina, October 2015.

  2. Yue Li, Yassine Hadjadj-Aoul, Philippe Bertin, Gerardo Rubino, **"Control theory based interface selection mechanism in Fixed-Mobile Converged network,"** in 16th IEEE International Conference on High Performance Switching and Routing (IEEE HPSR 2015), Budapest, Hungary, July 2015.

  3. Yue Li, Yassine Hadjadj-Aoul, Philippe Bertin, Gerardo Rubino, **"A control theoretic strategy for intelligent interface selection in heterogeneous network environments,"** in 13th Annual IEEE Consumer Communications & Networking Conference (IEEE CCNC 2016), Las Vegas, USA, January 2016.

  4. Yue Li, Pantelies A. Frangoudis, Yassine Hadjadj-Aoul and Philippe Bertin, **"A Mobile-Edge-Computing-Based Architecture for Improved Adaptive HTTP Video Delivery,"** in 2nd IEEE Conference on Standards for Communications & Networking (IEEE CSCN 2016), Berlin, Germany, Oct 2016.

  5. Yue Li, Pantelies A. Frangoudis, Yassine Hadjadj-Aoul and Philippe Bertin, **"A Mobile Edge Computing-assisted Streaming Architecture for Wireless Heterogeneous Networks,"** (accepted) in 22nd IEEE Symposium on Computers & Communication (IEEE ISCC 2017), Heraklion, Crete, Greece, July 2017.

- Patents

    4. Yue Li, Yassine Hadjadj-Aoul and Philippe Bertin, **" Optimisation des flux DASH dans les réseaux mobiles 4G et au-delà,"**, France, submitted in April, 2016

Bibliography

[1] Cisco. *Cisco Visual Networking Index: Forecast and Methodology, 2014-2019*, White Paper, May 2015.

[2] C .E .Shannon. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, USA, 1948

[3] A. Damnjanovic and J. Montojo and Y. Wei and T. Ji and T. Luo and M. Vajapeyam and T. Yoo and O. Song and D. Malladi. A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications*, 18(3):10-21, June 2011.

[4] E. Gustafsson and A. Jonsson. Always best connected. *IEEE Wireless Communications*, 10(1):49-55, Feb 2003.

[5] COMBO. *COMBO (COnvergence of fixed and Mobile BrOadband access/aggregation networks), FP7 ICT Integrated Project.* [Online]. Available: http://www.ict-combo.eu/.

[6] Y. Li and Y. Hadjadj-Aoul and P. Bertin and G. Rubino. Control theory based interface selection mechanism in Fixed-Mobile Converged network. In *IEEE International Conference on High Performance Switching and Routing (HPSR 2015)*, pages 1-6, July 2015.

[7] Y. Li and Y. Hadjadj-Aoul and P. Bertin and G. Rubino. A control theoretic strategy for intelligent interface selection in heterogeneous network environments. In *IEEE Annual Consumer Communications Networking Conference (CCNC 2016)*, pages 1006-1011, Jan 2016.

[8] Y. Li and P. A. Frangoudis and Y. Hadjadj-Aoul and P. Bertin. A Mobile Edge Computing-based architecture for improved adaptive HTTP video delivery. In *IEEE Conference on Standards for Communications and Networking (CSCN 2016)*, pages 1-6, Oct 2016.

[9] Y. Li and P. A. Frangoudis and Y. Hadjadj-Aoul and P. Bertin. A Mobile Edge Computing-assisted Streaming Architecture for Wireless Heterogeneous Networks (accepted). In *IEEE Symposium on Computers and Communication (ISCC 2017)*, pages 1-6, July 2017.

[10] R. Trestian and O. Ormond and G. M. Muntean. Game Theory-Based Network Selection: Solutions and Challenges. *IEEE Communications Surveys Tutorials*, 14(4):1212-1231, 2012.

[11] P. Fishburn. *Utility theory for decision making.* Wiley New York, 1970.

[12] L. Liang and H. Wang and P. Zhang. Net Utility-Based Network Selection Scheme in CDMA Cellular/WLAN Integrated Networks. In *IEEE Wireless Communications and Networking Conference (WCNC 2007)*, pages 3313-3317, March 2007.

[13] Yisheng Lai and Kok Keong Chai and Yue Chen. A utility-based intelligent network selection for 3G and WLAN heterogeneous networks. In *IET International Conference on Wireless Communications and Applications (ICWCA 2012)*, pages 1-6, April 2012.

[14] L. Wang and G. S. Kuo. Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks - A Tutorial. *IEEE Communications Surveys Tutorials*, 15(1):271-292, 2013.

[15] Ormond, O. and Murphy, J. and Muntean, G.. Utility-based Intelligent Network Selection in Beyond 3G Systems. In *IEEE International Conference on Communications (ICC 2006)*, pages 1831-1836, Jun 2006.

[16] Q. T. Nguyen-Vuong and Y. Ghamri-Doudane and N. Agoulmine. On utility models for access network selection in wireless heterogeneous networks. In *IEEE Network Operations and Management Symposium (NOMS 2008)*, pages 144-151, April 2008.

[17] J. McNair and Fang Zhu. Vertical handoffs in fourth-generation multi-network environments. *IEEE Wireless Communications*, 15(1):8-15, June 2004.

[18] E. Stevens-Navarro and V. W. S. Wong. Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks. In *IEEE 63rd Vehicular Technology Conference (VTC2006-Spring)*, pages 947-951, May 2006.

[19] F. Bari and V. C. M. Leung. Use of non-monotonic utility in multi-attribute network selection. In *Wireless Telecommunications Symposium (WTS 2007)*, pages 1-8, April 2007.

[20] Qingyang Song and A. Jamalipour. Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques. *IEEE Wireless Communications*, 12(3):42-48, June 2005.

[21] F. Bari and V. C. M. Leung. Automated network selection in a heterogeneous wireless network environment. *IEEE Network*, 21(1):34-40, Jan 2007.

[22] F. Bari and V. C. M. Leung. Multi-Attribute Network Selection by Iterative TOPSIS for Heterogeneous Wireless Access. In *IEEE Consumer Communications and Networking Conference (CCNC 2007)*, pages 808-812, Jan 2007.

[23] Chamodrakas, Ioannis and Martakos, Drakoulis. Utility-based Fuzzy TOPSIS Method for Energy Efficient Network Selection in Heterogeneous Wireless Networks. *Applied Soft Computing*, 11(4):3734-3743, June 2011.

[24] Chung-Ju Chang and Tsung-Li Tsai and Yung-Han Chen. Utility and Game-Theory Based Network Selection Scheme in Heterogeneous Wireless Networks. In *IEEE Wireless Communications and Networking Conference (WCNC 2009)*, pages 1-5, April 2009.

[25] N. D. Tripathi. *Generic adaptive handover algorithms using fuzzy logic and neural networks*. PhD dissertation, Virginia Polytechnic Inst. and State Univ., Blacksburg, 1997.

[26] J. Hou and D. C. O'Brien. Vertical handover-decision-making algorithm using fuzzy logic for the integrated Radio-and-OW system. *IEEE Transactions on Wireless Communications*, 5(1):176-185, Jan 2006.

[27] Meriem Kassar and Brigitte Kervella and Guy Pujolle. An overview of vertical handover decision strategies in heterogeneous wireless networks. *Computer Communications*. 31(10):2607-2620, June 2008.

[28] Wenhui Zhang. Handover decision using fuzzy MADM in heterogeneous networks. In *IEEE Wireless Communications and Networking Conference (WCNC 2004)*, pages 653-658, March 2004.

[29] Alkhwlani, M.M. and Ayesh, A.. Access network selection using combined fuzzy control and MCDM in heterogeneous networks. In *International Conference on Computer Engineering & Systems (ICCES 2007)*, pages 108-113, Nov 2007.

[30] P. M. L. Chan and Y. F. Hu and R. E. Sheriff. Implementation of fuzzy multiple objective decision making algorithm in a heterogeneous mobile environment. In *IEEE Wireless Communications and Networking Conference (WCNC 2002)*, pages 332-336, Mar 2002.

[31] Malanchini, I. and Cesana, M. and Gatti, N.. Network Selection and Resource Allocation Games for Wireless Access Networks. *IEEE Transaction On Mobile Computing*, 12(12):2427-2440, Dec 2013.

[32] E. H. Watanabe and D. S. Menasche and E. de Souza e Silva and R. M. M. Leao. Modeling Resource Sharing Dynamics of VoIP Users over a WLAN Using a Game-Theoretic Approach. In *IEEE International Conference on Computer Communications (INFOCOM 2008)*, pages 915-923, April 2008.

[33] S. Vassaki and A. D. Panagopoulos and P. Constantinou. Bandwidth allocation in wireless access networks: Bankruptcy game vs cooperative game. In *International Conference on Ultra Modern Telecommunications Workshops (ICUMT 2009)*, pages 1-4, Oct 2009.

[34] Chatterjee, Mainak and Lin, Haitao and Das, Sajal K.. Non-cooperative Games for Service Differentiation in CDMA Systems. *Journal Mobile Networks and Applications*, 10(6):939-946, Dec 2005.

[35] Josephina Antoniou and Vicky Papadopoulou and Vasos Vassiliou and Andreas Pitsillides. Cooperative user-network interactions in next generation communication networks. *Journal Computer Networks: The International Journal of Computer and Telecommunications Networking*, 54(13):2239-2255, Sep 2010.

[36] H. Pervaiz and J. Bigham. Game Theoretical Formulation of Network Selection in Competing Wireless Networks: An Analytic Hierarchy Process Model. In *International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2009)*, pages 292-297, Sep 2009.

[37] D. Niyato and E. Hossain. A Cooperative Game Framework for Bandwidth Allocation in 4G Heterogeneous Wireless Networks. In *IEEE International Conference on Communications (ICC 2006)*, pages 4357-4362, June 2006.

[38] E. A. Feinburg and A. Shwartz. *Handbook of Markov decision processes: methods and applications*, Kluwer Academic Publishers, 2002.

[39] E. Stevens-Navarro and Y. Lin and V. W. S. Wong. An MDP-Based Vertical Hand-off Decision Algorithm for Heterogeneous Wireless Networks. *IEEE Transactions on Vehicular Technology*, 57(2):1243-1254, March 2008.

[40] X. Gelabert and J. Pérez-Romero and O. Sallent and R. Agustí. A Markovian Approach to Radio Access Technology Selection in Heterogeneous Multiaccess/Multiservice Wireless Networks. *IEEE Transactions on Mobile Computing*, 7(10):1257-1270, Oct 2008.

[41] 3GPP. Access to the 3GPP Evolved Packet Core(EPC) via non-3GPP access networks (Release 11); Stage 3. TS 24.302, Dec 2014.

[42] 3GPP. Access Network Discovery and Selection Function(ANDSF) Management Object (MO) (Release11). TS 24.312, Dec 2014.

[43] T. Ahmed and D. Barankanira and S. Antoine and Xiaofeng Huang and H. Duvocelle. Inter-system mobility in evolved packet system (EPS): Connecting non-3GPP accesses. In *International Conference on Intelligence in Next Generation Networks (ICIN 2010)*, pages 1-6, Oct 2010.

[44] D. Xenakis and N. Passas and L. Merakos and C. Verikoukis. ARCHON: An ANDSF-assisted energy-efficient vertical handover decision algorithm for the heterogeneous IEEE 802.11/LTE-advanced network. In *IEEE International Conference on Communications (ICC 2014)*, pages 3166-3171, June 2014.

[45] D. S. Kim and Y. Noishiki and Y. Kitatsuji and H. Yokota. Efficient ANDSF-assisted Wi-Fi control for mobile data offloading. In *International Wireless Communications and Mobile Computing Conference (IWCMC 2013)*, pages 343-348, Oct 2013.

[46] Y. M. Kwon and J. S. Kim and J. Gu and M. Y. Chung. Andsf-based congestion control procedure in heterogeneous networks. In *International Conference on Information Networking (ICOIN 2013)*, pages 547-550, Jan 2013.

[47] J. Orimolade and N. Ventura and O. Falowo. Inter-system mobility in evolved packet system (EPS): Connecting non-3GPP accesses. In *Mediterranean Electrotechnical Conference (MELECON 2016)*, pages 1-6, April 2016.

[48] OPENET. *Openet's Network Selection Intelligence solution.* [Online]. Available: http://www.openet.com/areas-expertise/network-optimization/wi-fi-offload/.

[49] Aptilo. *Aptilo's Wi-Fi offload solution.* [Online]. Available: https://www.aptilo.com/mobile-data-offloading/.

[50] Aruba. *Carrier-class public Wi-Fi for 3G/4G offload.* [Online]. Available: www.arubanetworks.com/assets/so/SB_Offload.pdf/.

[51] IEEE 802.11 TG. IEEE Standard for Information Technology-Telecommunications and information exchange between systems-Local and Metropolitan networks-specific requirements-Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 9: Interworking with External Networks. Feb, 2011.

[52] Microsoft. *Smooth streaming: The official microsoft IIS site.* [Online]. Available: https://www.iis.net/downloads/microsoft/smooth-streaming/

[53] R. Pantos and W. May. *HTTP Live Streaming.* [Online]. Available: https://tools.ietf.org/html/draft-pantos-http-live-streaming-07/

[54] Adobe. *HTTP dynamic streaming: Flexible delivery of on-demand and live video streaming.* [Online]. Available: http://www.adobe.com/products/hds-dynamic-streaming.html/

[55] ISO/IEC DIS 23009-1.2 Dynamic adaptive streaming over HTTP (DASH), April 2012.

[56] ETSI 3GPP. 3GPP TS 26.247; Transparent end-to-end packet-switched streaming service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH), Dec 2011.

[57] I. Sodagar. The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *IEEE MultiMedia*, 18(4):62-67, April 2011.

[58] K. D. Singh and Y. Hadjadj-Aoul and G. Rubino. Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC. In *IEEE Consumer Communications and Networking Conference (CCNC 2012)*, pages 127-131, Jan 2012.

[59] Akhshabi, Saamer and Begen, Ali C. and Dovrolis, Constantine. An Experimental Evaluation of Rate-adaptation Algorithms in Adaptive Streaming over HTTP. In *ACM Multimedia Systems Conference (MMSys 2011)*, pages 157-168, Feb 2011.

[60] Huang, Te-Yuan and Handigol, Nikhil and Heller, Brandon and McKeown, Nick and Johari, Ramesh. Confused, Timid, and Unstable: Picking a Video Streaming Rate is Hard. In *ACM Conference on Internet Measurement Conference (IMC 2012)*, pages 225-238, Nov 2012.

[61] Akhshabi, Saamer and Anantakrishnan, Lakshmi and Begen, Ali C. and Dovrolis, Constantine. What Happens when HTTP Adaptive Streaming Players Compete for Bandwidth? In *International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2012)*, pages 9-14, Jun 2012.

[62] Z. Li and X. Zhu and J. Gahm and R. Pan and H. Hu and A. C. Begen and D. Oran. Probe and Adapt: Rate Adaptation for HTTP Video Streaming At Scale. *IEEE Journal on Selected Areas in Communications*, 32(4):719-733, April 2014.

[63] Müller, Christopher and Lederer, Stefan and Timmerer, Christian. An Evaluation of Dynamic Adaptive Streaming over HTTP in Vehicular Environments. In *ACM Workshop on Mobile Video (Movid 2012)*, pages 37-42, Feb 2012.

[64] Riiser, Haakon and Bergsaker, Håkon S. and Vigmostad, Paul and Halvorsen, Pål and Griwodz, Carsten. A Comparison of Quality Scheduling in Commercial Adaptive HTTP Streaming Solutions on a 3G Network. In *ACM Workshop on Mobile Video (Movid 2012)*, pages 25-30, Feb 2012.

[65] I. Ben Mustafa and M. Uddin and T. Nadeem. Understanding the intermittent traffic pattern of HTTP video streaming over wireless networks. In *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2016)*, pages 1-8, May 2016.

[66] Adhikari, Vijay Kumar and Guo, Yang and Hao, Fang and Varvello, Matteo and Hilt, Volker and Steiner, Moritz and Zhang, Zhi-Li. Unreeling netflix: Understanding and improving multi-CDN movie delivery. In *IEEE International Conference on Computer Communications (INFOCOM 2012)*, pages 1620-1628, March 2012.

[67] Liu, Chenghao and Bouazizi, Imed and Gabbouj, Moncef. Rate Adaptation for Adaptive HTTP Streaming. In *ACM Multimedia Systems Conference (MMSys 2011)*, pages 169-174, Feb 2011.

[68] X. Zhu and Z. Li and R. Pan and J. Gahm and H. Hu. Fixing multi-client oscillations in HTTP-based adaptive streaming: A control theoretic approach. In *IEEE International Workshop on Multimedia Signal Processing (MMSP 2013)*, pages 230-235, Sept 2013.

[69] Jarnikov, Dmitri and Özçelebi, Tanir. Client intelligence for adaptive streaming solutions. In *IEEE International Conference on Multimedia & Expo (ICME 2010)*, pages 1499-1504, July 2010.

[70] V. Ramamurthi and O. Oyman. Video-QoE aware radio resource allocation for HTTP adaptive streaming. In *IEEE International Conference on Communications (ICC 2014)*, pages 1076-1081, June 2014.

[71] Yin, Xiaoqi and Jindal, Abhishek and Sekar, Vyas and Sinopoli, Bruno. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *ACM Conference on Special Interest Group on Data Communication (SIGCOMM 2015)*, pages 325-338, Aug 2015.

[72] Tian, Guibin and Liu, Yong. Towards Agile and Smooth Video Adaptation in Dynamic HTTP Streaming. In *International Conference on Emerging Networking Experiments and Technologies (CoNEXT 2012)*, pages 109-120, Dec 2012.

[73] C. Zhou and X. Zhang and L. Huo and Z. Guo. A control-theoretic approach to rate adaptation for dynamic HTTP streaming. In *Visual Communications and Image Processing (VCIP 2012)*, pages 1-6, Nov 2012.

[74] S. Petrangeli and M. Claeys and S. Latré and J. Famaey and F. De Turck. A multi-agent Q-Learning-based framework for achieving fairness in HTTP Adaptive Streaming. In *IEEE Network Operations and Management Symposium (NOMS 2014)*, pages 1-9, May 2014.

[75] N. Bouten and S. Latré and J. Famaey and W. Van Leekwijck and F. De Turck. In-Network Quality Optimization for Adaptive Video Streaming Services. *IEEE Transactions on Multimedia*, 16(8):2281-2293, Dec 2014.

[76] N. Bouten, J. Famaey, S. Latré, R. Huysegems, B. Vleeschauwer, W. Leekwijck, and F. De Turck. QoE optimization through in-network quality adaptation for HTTP Adaptive Streaming. In *International Conference on Network and Service Management (CNSM 2012)*, pages 336-342, Oct 2012.

[77] Wei Pu and Zixuan Zou and Chang Wen Chen. Video adaptation proxy for wireless Dynamic Adaptive Streaming over HTTP. In *International Packet Video Workshop (PV 2012)*, pages 65-70, May 2012.

[78] V. Krishnamoorthi and N. Carlsson and D. Eager and A. Mahanti and N. Shahmehri. Helping Hand or Hidden Hurdle: Proxy-Assisted HTTP-Based Adaptive Streaming Performance. In *IEEE International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2013)*, pages 182-191, Aug 2013.

[79] *Squid Proxies*. [Online]. Available: http://www.squidproxies.com/.

[80] K. J. Ma and R. Bartos. HTTP Live Streaming Bandwidth Management Using Intelligent Segment Selection. In *IEEE Global Telecommunications Conference (GLOBECOM 2011)*, pages 1-5, Dec 2011.

[81] Mok, Ricky K. P. and Luo, Xiapu and Chan, Edmond W. W. and Chang, Rocky K. C.. QDASH: A QoE-aware DASH System. In *ACM Multimedia Systems Conference (MMSys 2012)*, pages 11-22, Feb 2012.

[82] Essaili, Ali El and Schroeder, Damien and Staehle, Dirk and Shehada, Mohammed and Kellerer, Wolfgang and Steinbach, Eckehard G.. Quality-of-experience driven adaptive HTTP media delivery. In *IEEE International Conference on Communications (ICC 2013)*, pages 2480-2485, June 2013.

[83] Petrangeli, Stefano and Famaey, Jeroen and Claeys, Maxim and Latré, Steven and De Turck, Filip. QoE-Driven Rate Adaptation Heuristic for Fair Adaptive Video Streaming. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(2):1-24, March 2016.

[84] I. Rubin and S. Colonnese and F. Cuomo and F. Calanca and T. Melodia. Mobile HTTP-based streaming using flexible LTE base station control. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2015)*, pages 1-9, June 2015.

[85] H. B. Chang and I. Rubin and S. Colonnese and F. Cuomo and O. Hadar. Joint Adaptive Rate and Scheduling for Video Streaming in Multi-Cell Cellular Wireless Networks. In *IEEE Global Communications Conference (GLOBECOM 2015)*, pages 1-6, Dec 2015.

[86] De Cicco, Luca and Mascolo, Saverio and Palmisano, Vittorio. Feedback Control for Adaptive Live Video Streaming. In *ACM Multimedia Systems Conference (MMSys 2011)*, pages 145-156, Feb 2011.

[87] H. Kim and N. Feamster. Improving network management with software defined networking. *IEEE Communications Magazine*, 51(2):114-119, Feb 2013.

[88] McKeown, Nick and Anderson, Tom and Balakrishnan, Hari and Parulkar, Guru and Peterson, Larry and Rexford, Jennifer and Shenker, Scott and Turner, Jonathan. OpenFlow: Enabling Innovation in Campus Networks. *ACM SIGCOMM Computer Communication Review*, 38(2):69-74, April 2008.

[89] Georgopoulos, Panagiotis and Elkhatib, Yehia and Broadbent, Matthew and Mu, Mu and Race, Nicholas. Towards Network-wide QoE Fairness Using Openflow-assisted Adaptive Video Streaming. In *ACM SIGCOMM Workshop on Future Human-centric Multimedia Networking (FhMN 2013)*, pages 15-20, Aug 2013.

[90] J. J. Quinlan and A. H. Zahran and K. K. Ramakrishnan and C. J. Sreenan. Delivery of adaptive bit rate video: balancing fairness, efficiency and quality. In *IEEE International Workshop on Local and Metropolitan Area Networks (LANMAN 2015)*, pages 1-6, April 2015.

[91] Cetinkaya, C. and Ozveren, Y. and Sayit, M.. An SDN-assisted system design for improving performance of SVC-DASH. In *Federated Conference on Computer Science and Information Systems (FedCSIS 2015)*, pages 819-826, Sept 2015.

[92] Chin-Feng Lai and Ren-Hung Hwang and Han-Chieh Chao and Hassan, M. and Alamri, A.. A buffer-aware HTTP live streaming approach for SDN-enabled 5G wireless networks. *IEEE Network*, 29(1):49-55, Jan 2015.

[93] S. Laga and T. Van Cleemput and F. Van Raemdonck and F. Vanhoutte and N. Bouten and M. Claeys and F. De Turck. Optimizing scalable video delivery through OpenFlow layer-based routing. In *IEEE Network Operations and Management Symposium (NOMS 2014)*, pages 1-4, May 2014.

[94] *Network simulator 3*. [Online]. Available: https://www.nsnam.org/.

[95] Keshav, Srinivasan. A Control-theoretic Approach to Flow Control. *ACM SIGCOMM Computer Communication Review*, 25(1):3-16, August 1991.

[96] C. V. Hollot and V. Misra and D. Towsley and Wei-Bo Gong. A control theoretic analysis of RED. In *IEEE The Conference on Computer Communications (INFOCOM 2001)*, pages 1510-1519, April 2001.

[97] Tarek Abdelzaher, Yixin Diao, Joseph L. Hellerstein, Chenyang Lu, and Xiaoyun Zhu. *Introduction to control theory and its application to computing systems*, Springer-Verlag, pages 185-216, 2008

[98] F.L. Lewis, D. Vrabie, and V.L. Syrmos. *Optimal Control*, 3rd edition, John Wiley, 2013.

[99] M.C. de Oliveira and J. Bernussou and J.C. Geromel. A new discrete-time robust stability condition. *Systems & Control Letters*, 37(4):261-265, August 1999.

[100] Zhou, Kemin and Doyle, John C. and Glover, Keith. *Robust and Optimal Control*, Prentice-Hall, Inc, Upper Saddle River, NJ, USA, 1996.

[101] Tuncer, Daphné and Charalambides, Marinos and Landa, Raul and Pavlou, George. More control over network resources: An ISP caching perspective. In *International Conference on Network and Service Management (CNSM 2013)*, pages 26-33, Oct 2013.

[102] M. Claeys and D. Tuncer and J. Famaey and M. Charalambides and S. Latré and G. Pavlou and F. De Turck. Proactive multi-tenant cache management for virtualized ISP networks. In *International Conference on Network and Service Management (CNSM 2014)*, pages 82-90, Nov 2014.

[103] M. Patel et al.. Mobile-Edge Computing Introductory Technical WhitePaper, Sep 2015.

[104] Mathieu Bouet and Vania Conan and Hicham Khalife and Kevin Phemius and Jawad Seddar. MUREN: delivering edge services in joint SDN-SDR multi-radio nodes. *CoRR*, June 2016.

[105] H. Li and G. Shou and Y. Hu and Z. Guo. Mobile Edge Computing: Progress and Challenges. In *IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud 2016)*, pages 83-84, March 2016.

[106] E. Cau and M. Corici and P. Bellavista and L. Foschini and G. Carella and A. Edmonds and T. M. Bohnert. Efficient Exploitation of Mobile Edge Computing for Virtualized 5G in EPC Architectures. In *IEEE International Conference on Mobile*

Cloud Computing, Services, and Engineering (MobileCloud 2016), pages 100-109, March 2016.

[107] O. Mäkinen. Streaming at the Edge: Local Service Concepts Utilizing Mobile Edge Computing. In *International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2016)*, pages 1-6, Aug 2016.

[108] M. T. Beck and S. Feld and A. Fichtner and C. Linnhoff-Popien and T. Schimper. In *Intelligence in Next Generation Networks (ICIN 2015)*, pages 38-44, Feb 2015.

[109] K. Zhang and Y. Mao and S. Leng and Q. Zhao and L. Li and X. Peng and L. Pan and S. Maharjan and Y. Zhang. Energy-efficient Offloading for Mobile Edge Computing in 5G Heterogeneous Networks. *IEEE Access*, 4:5896-5907, August 2016.

[110] Fajardo, J.O. and Taboada, I. and Liberal, F.. Improving content delivery efficiency through multi-layer mobile edge adaptation. *IEEE Network*, 29(6):40-46, Nov 2015.

[111] ISO/IEC 23009-1:2014. Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, May 2014.

[112] *nginx.* [Online]. Available: https://www.nginx.com/.

[113] *The Apache HTTP server project.* [Online]. Available: https://httpd.apache.org/.

[114] A. Kivity and Y. Kamay and D. Laor and U. Lublin and A. Liguori. kvm: the Linux virtual machine monitor. In *Ottawa Linux Symposium (OLS 2007)*, June 2007.

[115] *Big Buck Bunny.* [Online]. Available: https://peach.blender.org/.

[116] *GPAC multimedia open source project.* [Online]. Available: https://gpac.wp.mines-telecom.fr/.

[117] ITU-T SG12. Definition of Quality of Experience. COM12 - LS 62 - E, TD 109rev2 (PLEN/12), Geneva, Jan 2007.

[118] Rubino, Gerardo. Quantifying the Quality of Audio and Video Transmissions over the Internet: The PSQA Approach. *Communication Networks & Computer Systems*, Imperial College Press, Jan 2005.

[119] ITU-T. Series H: Audiovisual and Multimedia Systems – Infrastructure of audiovisual services – Coding of moving video – Advanced video coding for generic audiovisual services, Feb 2016.

[120] Ni, Pengpeng and Eg, Ragnhild and Eichhorn, Alexander and Griwodz, Carsten and Halvorsen, Pål. Flicker Effects in Adaptive Video Streaming to Handheld Devices. In *ACM international conference on Multimedia (MM 2011)*, pages 463-472, Dec 2011.

[121] Rodríguez, Demóstenes Z. and Wang, Zhou and Rosa, Renata L. and Bressan, Graça. The impact of video-quality-level switching on user quality of experience in dynamic adaptive streaming over HTTP. *EURASIP Journal on Wireless Communications and Networking*, 29(1):1-15, Jan 2015.

[122] Hyunwoo Nam and Kyung-Hwa Kim and Henning Schulzrinne. QoE matters more than QoS: Why people stop watching cat videos. In *IEEE International Conference on Computer Communications (INFOCOM 2016)*, pages 1-9, April 2016.

[123] Tobias Hoßfeld and Michael Seufert and Christian Sieber and Thomas Zinner and Phuoc Tran-Gia. Identifying QoE optimal adaptation of HTTP adaptive streaming based on subjective studies. *Journal Computer Networks*, 81(C):320-332, April 2015.

[124] Jain, R. and Chiu, D. and Hawe, W.. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Technical Report TR-301, DEC Research, Sep 1984

[125] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure. TCP Extensions for Multipath Operation with Multiple Addresses. RFC 6824, 2013.

[126] T. Ojanperä and J. Vehkaperä. Network-assisted multipath DASH using the distributed decision engine. In *International Conference on Computing, Networking and Communications (ICNC 2016)*, pages 1-6, Feb 2016.

[127] M. Xing and S. Xiang and L. Cai. A Real-Time Adaptive Algorithm for Video Streaming over Multiple Wireless Access Networks. *IEEE Journal on Selected Areas in Communications*, 32(4):795-805, April 2014.

[128] P. Houzé and E. Mory and G. Texier and G. Simon. Applicative-layer multipath for low-latency adaptive live streaming. In *IEEE International Conference on Communications (ICC 2016)*, pages 1-7, May 2016.

[129] Han, Bo and Qian, Feng and Ji, Lusheng and Gopalakrishnan, Vijay. MP-DASH: Adaptive Video Streaming Over Preference-Aware Multipath. In *International Conference on Emerging Networking Experiments and Technologies (CoNEXT 2016)*, pages 129-143, Dec 2016.

[130] Corbillon, Xavier and Aparicio-Pardo, Ramon and Kuhn, Nicolas and Texier, Géraldine, Gwendal Simon. Cross-layer Scheduler for Video Streaming over MPTCP. In *ACM Multimedia Systems Conference (MMSys 2016)*, pages 1-12, May 2016.

[131] Moon, Seonghoon and Yoo, Juwan and Kim, Songkuk. Adaptive interface selection over cloud-based split-layer video streaming via multi-wireless networks. *Future Generation Computer Systems*, 56(C):664-674, March 2016.

[132] T.Groléat, M.K.Sbai, S.Vaton, Y.Hadjadj-Aoul, K. Singh, S.Moteau. *Advances on monitoring and integration to the demonstrator*. [Online]. Available: https://recherche.telecom-bretagne.eu/vipeer/publications/.

[133] European Telecommunications Standards Institute (ETSI). Mobile Edge Computing (MEC); Framework and Reference Architecture, March 2016.

[134] S. Popinet, *IBM: IBM ILOG CPLEX Optimization Studio V12.4*. [Online]. Available: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

[135] Corbillon, Xavier and Aparicio-Pardo, Ramon and Kuhn, Nicolas and Texier, Géraldine, Gwendal Simon. Cross-layer Scheduler for Video Streaming over MPTCP. In *ACM Multimedia Systems Conference (MMSys 2016)*, pages 1-12, May 2016

[136] ITU-T Recommendation G.107. The E-model: A computational model for use in transmission planning, April 2009.

[137] ITU-T Recommendation P.1201. Parametric non-intrusive assessment of audiovisual media streaming quality. Amendment 2: New Appendix III - Use of ITU-T P.1201 for non-adaptive, progressive download type media streaming, Dec 2013.

[138] Tobias Hoßfeld and Lea Skorin-Kapov and Poul E. Heegaard and Martín Varela and Kuan-Ta Chen. On Additive and Multiplicative QoS-QoE Models for Multiple

QoS Parameters. In *5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)*, pages 44-48, Aug 2016.

Acronyms

1G First Generation. i

4G Fourth Generation. i

5G Fifth Generation. 31, 55

ABC Always Best Connection. 1, 9, 33

AHP Analytic Hierarchy Process. 16

ANDSF Access Network Discovery and Selection Function. i, 5, 18, 34, 48

API Application Programming Interface. 55

AVC Advanced Video Coding. 64

BMFF Base Media File Format. 60

BS Base Station. 1

C-RAN Cloud Radio Access Network. 57

CDMA Code Dvision Multiple Access. 16

CDN Content Delivery Networks. 23, 55

CP Content Provider. 55

CQI Channel Quality Indicator. 29, 57

DASH Dynamic Adaptive Streaming over HTTP. i, 2, 5, 24, 28, 55, 69, 71

DT LQR Discrete-Time Linear Quadratic Regulator. 40

EPC Evolved Packet Core. 18, 57

ETSI European Telecommunications Standards Institute. 55

FL Fuzzy Logic. 13

FMC Fixed–Mobile Convergence. 3

GRA Gray Relational Analysis. 11

GSM Global System for Mobile Communications. 9

HAS HTTP Adaptive Streaming. 6, 23, 32, 55

HD High Definition. 2

HSPA High Speed Packet Access. 1

HTTP Hypertext Transfer Protocol. 23

IETF Internet Engineering Task Force. 23

ISP Internet Service Provider. 58

JFI Jain's Fairness Index. 66

LQR Linear–Quadratic Regulator. i, 4, 34

LTE Long Term Evolution. 9, 34

MADM Multiple Attribute Decision Making. 11, 20

MDP Markov Decision Process. 17, 22, 28

MEC Multi-access Edge Computing. i, iv, vi, viii, 2, 4, 5, 7, 32, 55, 56, 69, 71

MEW Multiplicative Exponential Weighting. 11

MNO Mobile Network Operator. 5, 55

MO Management Objects. 48

MOS Mean Opinion Score. i, 5, 15, 65, 69

MPC Model Predictive Control. 28

MPD Media Presentation Description. 24, 58

MPEG Moving Picture Experts Group. 24

NAT Network Address Translation. 23

NERF Network Event Reporting Function. 20

NFV Network Function Virtualization. 57

NG-POP Next Generation Point of Presence. 3

NO Network Operator. 33

ns-3 Network Simulator 3. i, 5, 6, 34

OSMF Open Source Media Framework. 29

PD Proportional Derivative. 28

PIC Proportional Integral Controller. 28

PLMN Public Land Mobile Network. 19

PSQA Pseudo-Subjective Quality Assessment. 66

QoE Quality of Experience. i, 1, 23, 33, 55, 69

QoS Quality of Service. 1, 20, 40, 55

QP Quantization Parameter. 65, 66

RAGA Re-buffering Aware Gradient Algorithm. 28

RAT Radio Access Technologies. 1, 9

RNN Random Neural Network. 66

RTP Real-Time Transport Protocol. 23, 32

SAW Simple Addicive Weighting. 11

SDN Software Defined Networking. 30, 31

SVC Scalable Video Coding. 31

TCP Transmission Control Protocol. 28

TOPSIS Technique for Order Preference by Similarity to an Ideal Solution. 11

UE User Equipment. 18, 59

UMTS Universal Mobile Telecommunications System. 9

URI Uniform Resource Identifier. 24

VHO Vertical Handover. 17

VM Virtual Machine. 64

VNI Visual Networking Index. 1

VoD Video on Demand. 29

WiMAX Worldwide Interoperability for Microwave Access. 9

WLAN Wireless Local Area Network. 9

XML eXtensible Markup Language. 24