ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DE LA
MAÎTRISE EN GÉNIE
CONCENTRATION RÉSEAUX DE TÉLÉCOMMUNICATIONS
M. Sc. A

PAR
Nassima FELLAG

ALGORITHME DE COURTOISIE : ORDONNANCEMENT DANS LA LIAISON MONTANTE DES RÉSEAUX LTE-ADVANCED

MONTRÉAL, LE 17 MARS 2016

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Michel Kadoch, directeur du mémoire
Département de Génie Électrique à l'École de technologie supérieure

M. Patrick Cardinal président du jury
Département de Génie Logiciel et des TI à l'École de technologie supérieure

M. Kim Khoa Nguyen, membre du jury
Département de Génie Électrique à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 17 FEVRIER 2016

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

J'adresse tout d'abord mes plus vifs remercîments à mon directeur de recherche Monsieur Michel Kadoch, pour son aide, sa patience, ses conseils et ses orientations qui m'ont été un apport considérable.

Mes sincères remerciements vont également à Monsieur Patrick Cardinal, président du jury, et Monsieur Kim Khoa Nguyen, membre du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner ce travail.

Je remercie bien évidemment mon frère Ahmed Fellag pour son encouragement et son soutien.

Mes profonds remerciements vont à toute ma famille et mes amis pour leur support continu qu'ils m'ont apportée tout au long de ma démarche.

Enfin, je remercie tous ceux qui ont contribué de près ou de loin pour l'élaboration de ce travail.

ALGORITHME DE COURTOISIE: ORDONNANCEMENT DANS LA LIAISON MONTANTE DES RESEAUX LTE-ADVANCED

Nassima FELLAG

RÉSUMÉ

L'évolution rapide du nombre d'utilisateurs et l'apparition des nouveaux services multimédia ont motivé *Third-Generation Partnership Project* (3GPP) à développer de nouvelles technologies d'accès radio. De ce fait, l'agrégation de porteuses a été introduite à partir de la version 10 de LTE (LTE-Advanced) pour faire face aux demandes croissantes en termes de débit et de bande passante. Ainsi, de garantir la QoS aux différents types de services. Cependant, cette solution demeure incomplète si elle n'est pas accompagnée d'une bonne gestion de ressources. Plusieurs approches d'ordonnancement ont été proposées dans la littérature. Or, la majorité privilégie le trafic de haute priorité. Dans cette étude, une nouvelle approche d'ordonnancement dans la liaison montante des réseaux LTE-A a été développée dans le but d'assurer l'équité de service aux différentes classes de trafics et d'augmenter le nombre d'utilisateurs satisfaits dans le réseau à travers une gestion dynamique de priorités, qui permet de privilégier les trafics de basses priorités dans des intervalles de temps bien déterminés basés sur le temps d'attente moyen de chaque classe. Les résultats de simulations obtenus montrent que les paramètres de la QoS ont été bien améliorés pour les classes de basses priorités et la QoS de la classe prioritaire n'a pas été dégradée.

Mots clés: LTE, LTE-A, courtoisie, agrégation de porteuses, ordonnancement, QoS.

ALGORITHM COURTESY: UPLINK SCHEDULING IN LTE-ADVANCED NETWORK

Nassima FELLAG

ABSTRACT

The rapid evolution of the number of users and the emergence of new multimedia services has motivated Third-Generation Partnership Project (3GPP) to develop new radio access technologies. Thus, the carrier aggregation (CA) was introduced from version 10 LTE (LTE-Advanced) to meet the increasing demands in terms of throughput and bandwidth and to ensure the QoS for different types of services. However, that solution is incomplete if it's not accompanied by a proper management of resources. Several scheduling approaches were proposed in the literature. However, the majority favors high-priority traffic. In this study, a new approach of scheduling in the uplink has been developed to ensure fairness of service to different classes of traffics and increase the number of satisfied users in the network through a dynamic management of priorities that privilege the low-priorities traffics during a specific time intervals based on the average waiting time for each class. Simulation results show that the QoS parameters were much improved for the low-priorities classes and the QoS of high-priority class was not degraded.

Keywords: LTE, LTE-A, courtesy, carrier aggregation (CA), scheduling, QoS.

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AMBR	Aggregate Maximum Bit Rate
AMC	Adaptive Modulation and Coding
APN	Access Point Name
APN-GBR	Access Point Name-Guaranteed Bit Rate
ARP	Allocation and Retention Priority
ARQ	Automatic Repeat reQuest
EPC	Evolved Packet Core
E_UTRAN	Evolved_UMTS Terrestrial Radio Access Network
FDD	Frequency Division Duplexing
FEC	Forward Error Correction
GBR	Granted Bit Rate
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat reQuest
HSS	Home Subscriber Server
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
MAC	Medium Access Control
MBR	Maximum Bit Rate
MME	Mobility Management Entity
Non-GBR	Non-Guaranteed Bit Rate
NRT	Non Real Time
OFDMA	Orthogonal Frequency-Division Multiple Access
PDCP	Packet Data Convergence Protocol

PDN	Packet Data Network
PELR	Packet error loss rate
P-GW	Packet data network Gateway
PRB	Physical Ressource Bloc
QAM	Quadrature amplitude modulation
QCI	Quality Channel Indicator
QPSK	Quadrature Phase Shift Keying
RE	Resource Element
RLC	Radio Link Control
RRM	Radio Resource Management
RT	Real Time
SAE	System Architecture Evolution
SCFDMA	Single-Carrier Frequency Division Multiple Access
S-GW	Serving Gateway
SINR	Signal-to-Interference-plus-Noise Ratio
TDD	Time division Duplexing
3GPP	Third-Generation Partnership Project
TTI	Time Transmission Interval
UE-AMBR	User Equipment-Aggregate Maximum Bit Rate
UMTS	Universal Mobile Telecommunication System
UIT/IMT-A	International Telecommunication Union/International Mobile Telecommunications-Advanced

INTRODUCTION

L'augmentation massive du volume de trafic fait appel à des débits élevés pour soutenir les applications et les services avancés, qui sont devenus une partie intégrante de notre quotidien. Pour répondre à ce besoin, l'organisme de standardisation *Third-Generation Partnership Project* (3GPP) a mené le projet *Long Term Evolution* (LTE) pour la quatrième génération des réseaux mobiles afin de fournir un débit considérablement élevé par rapport à ses prédécesseurs avec un temps de latence réduit pour l'accès aux différents services (Jyrki et Penttinen, 2012).

Les systèmes LTE sont continuellement mis à jour depuis l'introduction de la version 8 en vue de l'amélioration des performances des réseaux sans fil, la version12 est la plus récente adoptée en 2014. Par ailleurs, l'appellation de LTE a changé à LTE-*Advanced* depuis la version10 où le mot *Advanced* a été ajouté principalement pour mettre en évidence la relation entre la version 10 de LTE et l'UIT/IMT-Advanced (Cox, 2012).

LTE-Advanced est devenue la technologie la plus prometteuse avec l'émergence des nouvelles techniques d'accès radio qui introduisent l'agrégation des composantes porteuses. LTE-A fournit une bande passante et un débit plus élevés afin de confronter les exigences accrues du trafic mobile et de maximiser le nombre d'utilisateurs admis et servis. Cependant, l'expansion de la bande passante et l'accroissement du débit demeurent des solutions incomplètes si elles ne sont pas accompagnées d'une bonne gestion d'accès aux ressources où l'ordonnanceur de la liaison montante est censé d'assurer l'équité de service aux utilisateurs admis et de garantir une qualité de service acceptable pour chaque type de classe. Or, la majorité des ordonnanceurs conçus privilégie le trafic de haute priorité au détriment du trafic de basse priorité et ne définissent pas un mécanisme de gestion des priorités qui s'adapte aux différentes situations du réseau, notamment dans les situations de congestion lorsque les ressources en bande passante sont insuffisantes pour assumer la forte charge de trafic. Par conséquent, les paquets de basse priorité acceptés sont retenus dans les files jusqu'à l'achèvement du service de tous les paquets de haute priorité, ce qui engendre des

délais importants pouvant dépasser les délais tolérés et causant la perte de paquets moins prioritaires. Par contre, dans le cas où la capacité maximale des files d'attente est atteinte, les nouvelles demandes de services sont rejetées. Ainsi, la probabilité de blocage devient importante.

L'algorithme de courtoisie d'optimisation de la performance dans les réseaux WIMAX fixes a été proposé (Tata et Kadoch, 2008) afin d'assurer une gestion équitable et efficace des ressources et d'optimiser le service du trafic moins prioritaire, le principe de l'algorithme consiste à transmettre les paquets de basse priorité à la place des paquets prioritaires tant que la QoS de ces derniers n'est pas affectée. L'application de l'approche a permis de réduire le temps d'attente des paquets moins prioritaires et de minimiser leur taux de perte. Bien que cette solution améliore le service des classes défavorisées. Cependant, elle n'apporte pas d'amélioration au niveau des classes prioritaires. De plus, elle ne s'intéresse pas à améliorer le débit et la probabilité de blocage dans le système. De ce fait, une nouvelle approche d'ordonnancement dans la liaison montante a été développée, à savoir l'algorithme de courtoisie d'ordonnancent dans la liaison montante des réseaux LTE-Avancé dont le but est d'augmenter le nombre de clients satisfaits et d'assurer l'équité de service aux différentes classes de trafic. Dans cette étude, l'algorithme convient aux situations de charge modérée ainsi qu'aux situations de congestion du réseau. De plus de réduire le délai moyen des classes défavorisées sans détériorer la QoS de la classe prioritaire et de minimiser le blocage et la perte de paquets dans le système et d'améliorer le débit.

La solution proposée est axée sur un schéma de base définissant trois classes de trafics relatives au trafic *handoff* et aux nouveaux trafics RT et NRT car d'une part, dans un réseau réel le trafic *handoff* est toujours favorisé par rapport au nouveau trafic et le trafic RT est prioritaire en comparant avec le trafic NRT. D'autre part, ce schéma illustre le cas de la majorité des approches antérieures qui adoptent un ordre statique de priorité. Ensuite, l'algorithme de courtoisie d'ordonnancement dans la liaison montante est appliqué, ce qui permet à chaque classe de trafic de détenir la plus haute priorité durant des intervalles de temps bien déterminés. Ainsi, l'équité de service est garantie.

Le présent document comprend quatre chapitres. Le premier chapitre résume quelques généralités de la technologie LTE/LTE-A, à savoir les caractéristiques de la technologie, l'architecture du réseau LTE/LTE-A, les paramètres de la qualité de service et les différentes entités responsables de la gestion des ressources. En outre, le chapitre présente une description des diverses politiques de gestion des files d'attente. Le deuxième chapitre est une revue de littérature des différentes solutions ayant traité le problème de l'ordonnancement dans les réseaux LTE/LTEA. Par la suite, le troisième chapitre décrit l'algorithme de courtoisie d'ordonnancement dans la liaison montante des réseaux LTE/LTE-A dont les performances sont évaluées et discutées dans le quatrième chapitre. Ce travail s'achève par une conclusion récapitulant les différentes étapes suivies pour réaliser l'objectif de cette recherche et se termine par des recommandations et des propositions de solutions pour les futurs travaux.

CHAPITRE 1

GÉNÉRALITÉS SUR LA TECHNOLOGIE LTE/LTE-A

1.1 Introduction

La technologie LTE/LTE-A présente des performances attractives en termes de flexibilité de déploiement et de richesse des services offerts. Une solution IP complète garantit un traitement approprié pour chaque type de trafic en fonction des exigences de la qualité de services et de la disponibilité de ressources dans le réseau.

Le présent chapitre est une introduction à la technologie LTE/LTE-A et aux différentes disciplines d'ordonnancement dans les files d'attente. Il fournit en premier lieu une vue globale des caractéristiques et des performances apportées dans les réseaux LTE/LTE-A en mettant l'accent sur la qualité de service et ses paramètres impliqués dans l'attribution des priorités aux flux. Ainsi, il décrit l'interface AIR avec les différentes entités et couches relatives à l'allocation des ressources radio.

La deuxième partie de ce chapitre s'intéresse aux diverses disciplines d'ordonnancement dans les files d'attente vu leur importance dans la gestion de ressources.

1.2 Généralités sur la technologie LTE/LTE-A (4G)

L'objectif du LTE (version 8) est de fournir un débit de transmission élevé, une latence faible, et un accès radio optimisé supportant multiples bandes de fréquence. En outre, son architecture a été conçue pour soutenir les flux IP avec une mobilité transparente. Les caractéristiques suivantes ont été spécifiées dans la version 8 du 3GPP. Toutefois, LTE/A a hérité de toutes les fonctionnalités du LTE version 8. (Sauter, 2011).

Modulation multi porteuses: Orthogonal Frequency Division Multiple Access (OFDMA) est la technique de modulation utilisée en voie descendante, en d'autres termes, de la station de base vers l'équipement utilisateur. Un émetteur OFDMA divise l'information en plusieurs sous-flux et transmet en parallèle les sous-flux sur différentes fréquences appelées sous-porteuses, l'avantage principal de l'OFDMA est sa robustesse contre les évanouissements dus à la propagation par trajets multiples. (Bouguen, Hardouin et Wolff, 2012).

Single-Carrier Frequency Division Multiple Access (SCFDMA): C'est une dérivée d'OFDMA, utilisée en voie montante, c'est à dire, de l'équipement utilisateur vers la station de base afin d'obtenir un facteur de crête faible.

Le facteur de crête est défini par le rapport de l'amplitude du pic du signal sur la valeur efficace du signal.

Support de Time division Duplexing (TDD) et Frequency Division Duplexing (FDD): LTE supporte les deux modes FDD et TDD pour que l'interface radio différencie entre les transmissions des équipements utilisateurs et les transmissions des stations de bases (les nœuds B). Dans le mode FDD, les stations de base transmettent sur une fréquence porteuse et les équipements utilisateurs sur une autre, les bandes de fréquences en liaisons montantes et descendantes demeurent inchangées avec des débits semblables, ce qui rend ce mode approprié pour les services temps réel (RT) tel que la voix. Tandis que dans le mode TDD, la même fréquence porteuse est maintenue. Toutefois, les moments de transmissions des mobiles et des stations de bases sont différents.

Par ailleurs, ce mode offre la possibilité au système d'exploiter les liens montants et descendants pendant une durée bien déterminée, ce qui le rend adapté au service non-temps réel (NRT) où les débits en voies descendantes sont plus importants (par exemple, la navigation web). (Ali-Yahiya, 2011).

Support de l'Adaptive Modulation and Coding (AMC): Le schéma de codage Forward Error Correction (FEC) est adopté afin de corriger les erreurs de transmission. La modulation et le codage sont basés sur les conditions du canal pour maximiser le débit dans un canal variant dans le temps. (Cox, 2012).

Largeur de bandes variables: L'interface radio Evolved UMTS Terreatrial Radio Access Network (E_UTRAN) doit pouvoir opérer dans des allocations de bandes de fréquences avec différentes tailles incluant 1,25, 1,6, 2,5, 5, 10, 15 et 20 MHz simultanément dans les liaisons montante et descendante.

Débit: Pour une largeur de bande allouée de 20 MHz, les débits maximaux sont définis par 100 Mb/s en voie descendante, soit une efficacité spectrale crête de 5bit/s/Hz et de 50 Mb/s en voie montante, soit une efficacité spectrale crête de 2,5bit/s/Hz. Il importe de rappeler que l'efficacité spectrale est définie par le rapport du débit binaire sur la bande passante.

Mobilité: L'E-UTRAN peut supporter une mobilité allant jusqu'à 500 km/h. Mais il est optimisé pour opérer avec des vitesses inférieures à 15km/h.

Retransmission de la couche liaison: Les demandes de retransmission automatique sont prises en charge au niveau de la couche liaison grâce au protocole Automatic Repeat reQuest (ARQ). Les paquets non acquittés par le récepteur sont considérés comme perdus et seront retransmis optionnellement, LTE supporte HARQ qui est un hybride efficace entre FEC et ARQ.

Support de multi-utilisateurs: LTE fournit une allocation de ressource sur les deux dimensions temporelle et fréquentielle, ce qui permet d'avoir multiples utilisateurs sur un *time slot*. Sachant qu'un time slot est défini comme l'unité la plus petite de ressource à allouer. (Ali-Yahiya, 2011).

1.2.1 Agrégation de porteuses dans LTE-Advanced

LTE-Advanced permet des débits théoriques pouvant atteindre 1Gb/s dans le lien descendant et 500 Mb/s dans le lien montant sur une bande passante maximale de 100 MHZ grâce à l'agrégation de porteuses, qui est une nouvelle technologie d'accès radio permettant d'attribuer plus d'une composante porteuse LTE (version8) au même équipement utilisateur LTE-A. L'équipement utilisateur devient associé à une seule cellule servante appelée la cellule primaire. Toutefois, il peut être associé à plusieurs cellules servantes appelées cellules secondaires en raison de la charge de trafic, des exigences de la QoS ou des considérations des politiques de déploiement. Les composantes porteuses allouées à l'équipement utilisateur sont appelées composantes porteuses secondaires, elles peuvent être contiguës ou non-contiguës et pas nécessairement de bandes passantes similaires, la porteuse de la cellule primaire est désignée composante porteuse primaire, le nombre de composantes porteuses à agréger est limité à cinq composantes. (Abu-Ali, Salahet Hassanein, 2014).

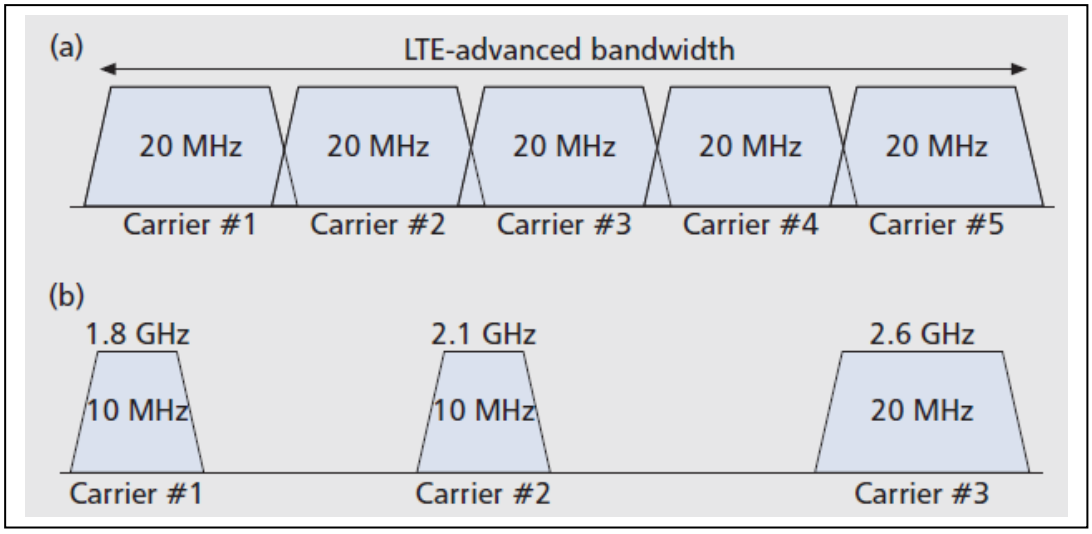


Figure 1.1 Exemple d'agrégation de porteuses
a) Agrégation de cinq composantes porteuses contiguës avec de bandes passantes similaires
b) Agrégation de trois composantes porteuses non- contiguës avec de bandes passantes distinctes
Tirée de IEEE Communications Magazine (2011)

1.2.2 Architecture LTE/LTE-Advanced

Le réseau LTE/LTE-A est fondé sur l'architecture appelée *Evolved Packet Switched* connue aussi sous le nom de SAE/LTE ou *System Architecture Evolution*/LTE. La figure1.2 présente l'architecture générale qui est composée de deux éléments principaux, à savoir le réseau d'accès *Evolved UTRAN* (E-UTRAN) également désigné LTE et le réseau cœur *Evolved Packet Core* (EPC).

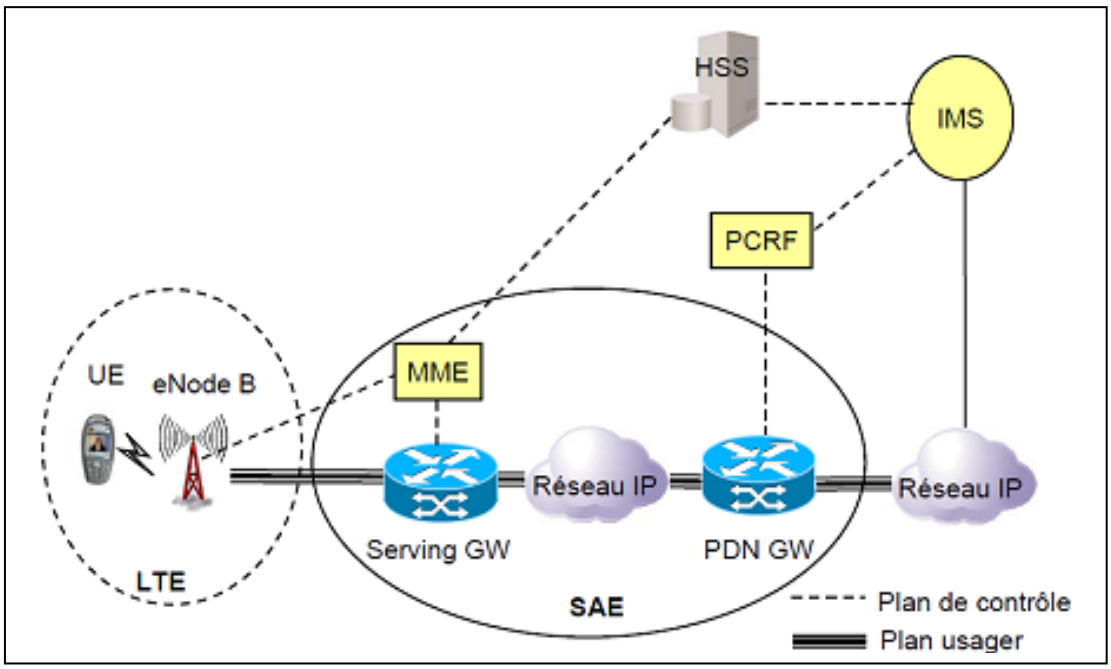


Figure 1.2 Architecture EPC
Tirée de http://fr.slideshare.net/Opethienne/lte
Consulté (décembre 2015)

1.2.2.1 Evolved UMTS Terrestrial Radio Access Network (E-UTRAN)

Comprend un seul nœud, le nœud B qui est une station de base responsable des échanges des transmissions radio avec l'équipement utilisateur. L'envoi des transmissions s'effectue sur la liaison descendante et la réception sur la liaison montante. D'autre part, le nœud B fournit les fonctionnalités requises pour la gestion de ressources radio incluant le contrôle d'admission,

le contrôle radio de porteurs, l'allocation de ressources et la gestion des signalisations. En outre, il s'occupe du chiffrement et de la compression d' d'entête IP**.** (Penttinen, 2012).

Comme montrée par la figure 1.3**,** l'interface S1 relie l'EPC avec le nœud B, alors que les différentes stations de bases s'interconnectent entre elles via l'interface X2. La station de base peut être connectée à plusieurs MME (*mobility management entity*) et S-GWS (*serving gateway*). Tandis que l'équipement utilisateur ne peut se connecter qu'à un seul nœud B à la fois. (Hallahan et Peha, 2010).

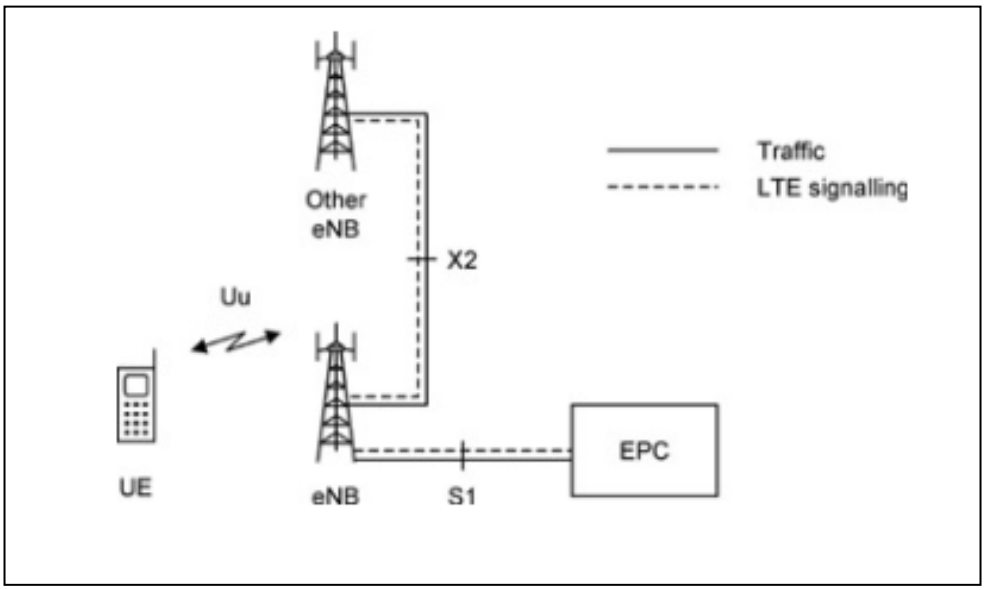


Figure 1.3 Architecture de l'E-UTRAN
Tirée de Jyrki et Penttinen (2012)

1.2.2.2 Evolved Packet Core (EPC)

a) Serveur d'abonnés Home Subscriber Server (HSS)

Une base de données centrale qui contient des informations sur tous les abonnés de l'opérateur du réseau. Le HSS est l'un des éléments rares du LTE/LTE-A qui a été reporté de l'UMTS et GSM.

b) Packet data network Gateway (P-GW)

Constitue le point de contact de l'EPC avec le monde extérieur via l'interface SGi. Chaque P-GW échange des données avec un ou plusieurs dispositifs externes ou des réseaux de données à commutation de paquets tels que les serveurs de l'opérateur du réseau ou internet. Chaque réseau est identifié par le nom du point d'accès APN. Un opérateur de réseau utilise généralement différents APN, par exemple un APN pour ses propres serveurs et un autre pour l'internet.

Chaque équipement utilisateur est affecté à une passerelle PDN par défaut dès qu'il est allumé pour qu'il soit toujours connecté à un réseau par défaut. Par ailleurs, il peut être affecté à d'autres passerelles PDN dans le cas où il souhaite se connecter à d'autres réseaux supplémentaires. Chaque passerelle PDN demeure la même pendant toute la durée de la connexion.

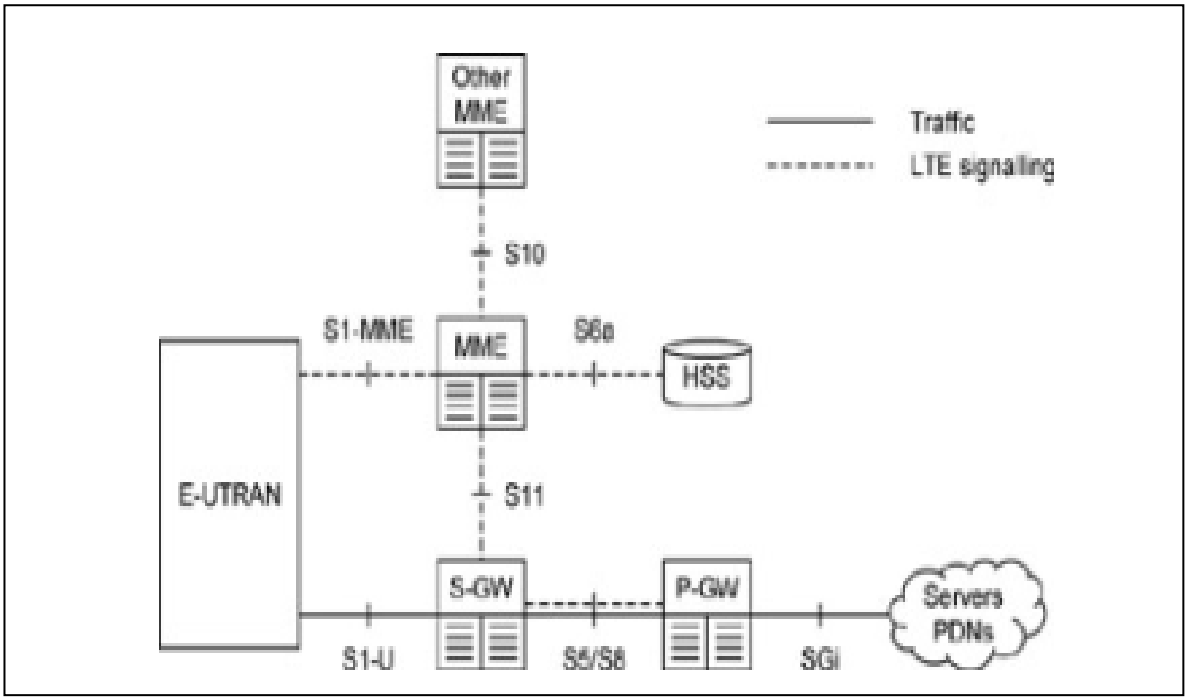


Figure 1.4 Composants principaux de l'EPC
Tirée de Jyrki et Penttinen (2012)

c) Serving Gateway (S–GW)

Il agit comme un routeur et transmet les données entre le nœud B et la passerelle PDN. Un réseau typique peut contenir plusieurs S-GW dont chacune s'occupe des mobiles dans une zone géographique bien déterminée. Chaque mobile est affecté à une seule passerelle de service. Cependant, la passerelle ne demeure pas la même durant la connexion, elle peut changer quand le mobile change de zone.

d) Mobility Management Entity (MME)

Contrôle les opérations de haut niveau du mobile en lui envoyant des messages de signalisation concernant la sécurité et la gestion des flux de données qui ne sont pas impliquées dans les communications radio. Un réseau typique peut contenir plusieurs MME dont chacun prend en charge une zone géographique spécifique. (Penttinen, 2012).

1.2.3 Qualité de service

La qualité de service est un composant primordial pour la livraison satisfaisante des services aux utilisateurs. Dans les réseaux LTE/LTE-A, la notion de canaux logiques EPS est introduite afin de définir plusieurs classes de QoS supportée**s** par l'E-UTRAN et l'EPC.
Les données véhiculées dans un même canal logique EPS subissent le même traitement et requirent la même qualité de service. Néanmoins, ce traitement diffère d'un canal à un autre par les priorités accordées aux paramètres d'EPC.

1.2.3.1 Porteur EPS

Le porteur est un canal logique établit entre l'équipement utilisateur et la passerelle P-GW, il permet de véhiculer tout le trafic de l'équipement utilisateur à travers le même canal physique ou le canal radio. Ainsi, plusieurs porteurs doivent être créés afin de distinguer entre les méthodes de traitement des différentes transmissions au sein du même canal telles que la politique d'ordonnancement, la politique de gestion de files d'attente, etc.…

Deux types de porteurs peuvent coexister dans un réseau LTE: porteur par défaut et porteur dédié.

a) Porteur par défaut

Chaque équipement utilisateur dispose d'au moins un porteur par défaut créé lors du premier attachement au réseau et demeure disponible durant toute la période de Connexion.

b) Porteur dédié

Établit après le porteur par défaut et utilisé lorsque les exigences de la QoS pour un trafic particulier sont différentes des dispositions de la QoS fournies par le porteur par défaut, en outre, tous les flux nécessitant la même QoS sont transmis via le même porteur dédié. De ce fait, le porteur constitue l'unité fondamentale pour discuter les mécanismes de la QoS disponibles dans le réseau LTE/LTE-A.

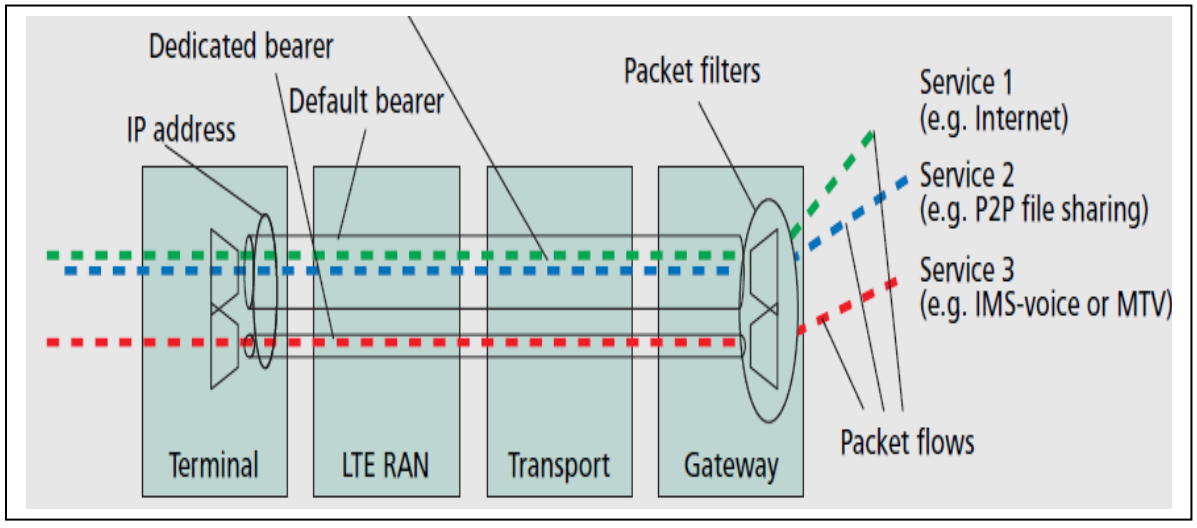


Figure 1.5 Type de porteurs
Tirée de P. Kandolcar (2013)

Une deuxième classification des porteurs peut être élaborée en fonction des garanties de la QoS attribuées au porteur lorsqu'il est établi. Un porteur GBR (*Garanted Bit Rate*) est un porteur associé à un débit binaire garanti qui est un débit de données moyen à long terme que

l'équipement utilisateur s'attend à recevoir. Les Porteurs GBR sont adaptés aux services temps réel tels que la voix. Quant au porteur non GBR, aucun débit binaire minimum n'est assuré par le réseau, ce qui le rend approprié pour les services non temps réel tel que la navigation web.

1.2.3.2 Paramètres de la qualité de service pour le porteur

Pour répondre aux exigences de la qualité de service, le porteur EPS doit définir des paramètres déterminant les traitements préférentiels qu'il puisse recevoir. Chaque porteur GBR ou non GBR est associé à ces deux paramètres**,** *Allocation and Retention Priority* (ARP) et *QoS Class Identifier* (QCI).

a) Allocation and Retention Priority (ARP)

Lorsque les ressources dans le réseau sont limitées, une demande d'établissement ou de modification de porteur peut être rejetée, notamment lorsqu'un porteur GBR est sollicité et la capacité radio est limitée. De ce fait, le paramètre ARP est introduit pour faciliter la prise de décision, il comporte trois composantes, une valeur scalaire unique contenant les informations sur le niveau de priorité d'un porteur et deux valeurs de drapeau distinctes se référant à la capacité de préemption et la vulnérabilité de préemption respectivement.

Les niveaux de priorités d'ARP sont utilisés pour assurer que les requêtes émanant d'un porteur de haute priorité sont privilégiées par rapport aux autres demandes provenant des porteurs moins prioritaires, ce qui offre la possibilité au réseau de choisir le porteur de faible priorité à préempter et par conséquent, libérer les ressources nécessaires. Le choix du porteur est basé sur la valeur de la capacité de préemption qui définit si un porteur donné est dans la mesure d'être préempté. D'autre part, la valeur de la capacité de vulnérabilité définie si un porteur est susceptible d'être préempté y compris le porteur de plus haute priorité ARP. Il convient de noter qu'après la mise en place du porteur, la valeur d'ARP n'a aucun effet sur le traitement de la transmission des paquets. (Ekström, 2009).

b) Quality Class Identifier (QCI)

Une fois les porteurs sont établis à l'aide des mécanismes de contrôle d'accès fournis par l'ARP, le nœud B doit encore savoir comment traiter les paquets de chaque porteur et allouer les ressources nécessaires. Ainsi, il est indispensable de définir un second paramètre pour accomplir la tâche de gestion de ressources. La technologie LTE/LTE-A définit une valeur scalaire désignée QCI afin de déterminer un groupe de paramètres de la QoS permettant de traiter les paquets à acheminer de chaque porteur. Ce traitement est effectué grâce à l'ordonnancement qui permet d'allouer les ressources radio à chaque porteur. Les paramètres principaux de la QoS sont les suivants:

Débit: Trois catégories de débit sont définies.

- *Guaranteed Bit Rate* (GBR): Les ressources réseaux allouées dans le cas du débit GBR demeurent inchangées car c'est un flux de données de service garanti;
- *Maximum Bit Rate* (MBR): utilisé pour limiter le GBR;
- *Aggregate Maximum Bit Rate* (AMBR): Paramètre défini pour le flux non-GBR, AMBR peut être APN-GBR ou UE-AMBR. APN-GBR désigne le débit maximum atteint par tous les porteurs non-GBR et les connexions PDN d'un PDN donné, alors que UE-AMBR. se réfère au débit binaire maximum autorisé pour tous les porteurs non-GBR agrégés d'un utilisateur.

Délai: Spécifié par le budget de retard des paquets, neuf catégories de délai sont définies citées dans le tableau.1.1.

Perte de paquets: le taux d'erreur de perte de paquets, comme le délai, neuf catégories sont définies dans le tableau.1.1.

Priorité: Spécifiée par l'ARP pour indiquer à la fois les priorités de planification et de rétention des flux de données. (Abu-Ali, Salah et Hassanein, 2014).

Le tableau1.1 résume les valeurs de QCIs déjà normalisées avec le niveau de la priorité, le délai des paquets, le PELR (Packet error loss rate) et des exemples de services mappés QCI. (Hallahan, et Peha, 2010).

Tableau 1.1 Valeurs normalisées de QCI avec leurs caractéristiques de QoS
Tiré de Hallahan et Peha (2010)

Resource Type	QCI	Priority	Packet Delay Budget[12]	Packet Error Loss Rate[13]	Example Services
GBR	1	2	100 ms	10^-2	Conversational Voice
	2	4	150 ms	10^-3	Conversational Video (Live Streaming)
	3	3	50 ms	10^-3	Real Time Gaming
	4	5	300 ms	10^-6	Non-Conversational Video (Buffered Streaming)
Non-GBR	5	1	100 ms	10^-6	IMS Signalling
	6	6	300 ms	10^-6	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing progressive video, etc.)
	7	7	100 ms	10^-3	Voice, Video (Live Streaming), Interactive Gaming
	8	8	300 ms	10^-6	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing progressive video, etc.)
	9	9	300 ms	10^-6	QCI typically used for the default bearer of a UE/PDN

1.2.4 Gestion de ressources radio

Les protocoles spécifiques de la liaison radio incluant les protocoles des couches *Radio Link Control* (RLC), *Medium Access Control* (MAC) et *Packet Data Convergence Protocol* (PDCP) sont situés au niveau du nœud B. En outre, le nœud B comprend également le module *Radio Resource management* (RRM) dont la principale fonction est de contrôler l'utilisation de ressources radio dans le système avec la prise en considération que les exigences des porteurs en QoS soient satisfaites à moindre coût possible.

Les algorithmes de contrôle d'admission et d'allocation dynamique de ressources radio sont exécutés par le nœud B au niveau du RRM.

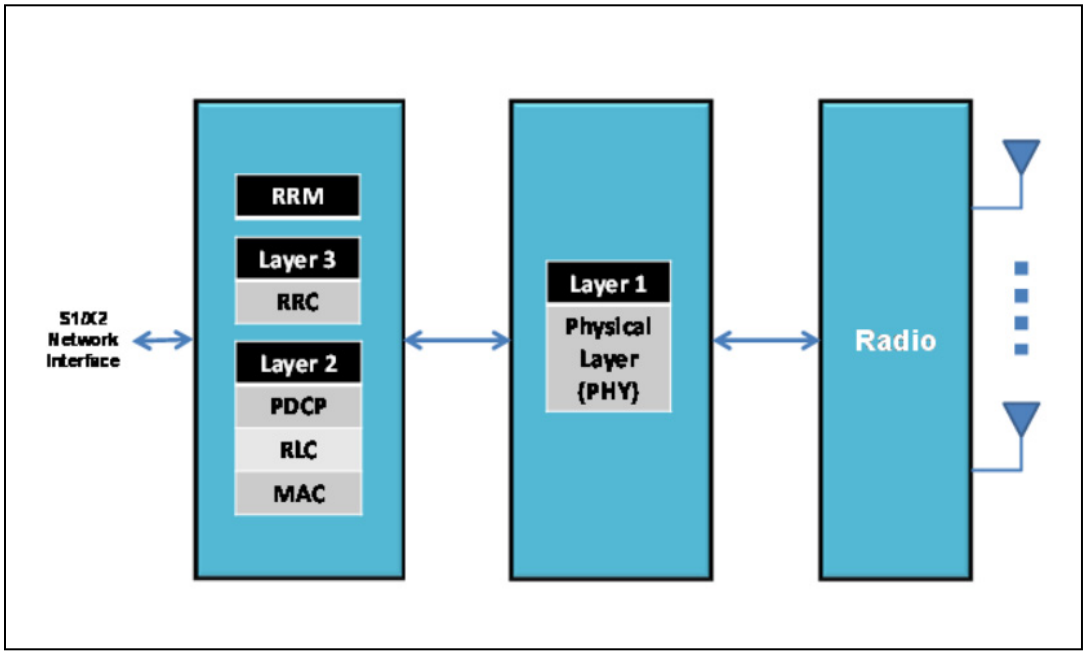


Figure 1.6 Architecture en couches du nœud B
Tirée de http://saitechnology.com/index.php/saisystem-solutions/sai-lte-system-solutions/sai-lte-enodeb%20.10
Consulté (août 2014)

1.2.4.1 Contrôle d'admission

Une sous fonction du RRM qui admet ou rejette les nouvelles demandes d'établissement des porteurs selon la disponibilité des ressources radio, ce qui permet de garantir la QoS appropriée aux sessions déjà en cours. Cette procédure est réalisée par le nœud B pour chacune de ses cellules, un porteur GBR n'est accepté que s'il y'a suffisamment de ressources inoccupées dans les liaisons montantes et descendantes simultanément.

1.2.4.2 Allocation de ressources radio

Le contrôle d'admission retourne l'état actuel d'allocation des ressources, l'étape suivante consiste à allouer les ressources libres de l'interface d'accès RAN aux différents utilisateurs, en appliquant les algorithmes d'ordonnancement disponibles au niveau du RRM. L'allocation de ressources se déroule au niveau de la couche MAC. Cependant, il est important de présenter le rôle de la couche physique dans cette procédure. (Nasir, 2011).

a) Couche physique

Appelée également la couche 1, son rôle est de fournir des services de transport sur l'interface air à la couche MAC. Les caractéristiques les plus importantes de la couche physique sont les techniques de modulation déjà citées, à savoir la modulation multi porteuses, l'OFDM en lien descendant et sa dérivée SC-FDMA en lien montant. Toutefois, elle réalise également les fonctions suivantes pour la transmission des données:

1) **La modulation:** Associe les bits à des symboles qui sont transmis sous forme d'onde électromagnétique.

2) **Le codage du canal:** Protège l'information contre les erreurs de transmission.

3) **Les traitements spatiaux (dits MIMO):** Permettent la transmission des symboles codés de plusieurs antennes.

Symbole, slot, bloc radio et trame: Un symbole désigne l'unité la plus petite d'information à transmettre sur une sous porteuse, il comprend plusieurs bits dont le nombre dépend du schéma de modulation adopté, dans le cas où les conditions radio sont excellentes, la modulation 64-QAM est utilisée pour transférer 6 bits par symbole ($2^6 = 64$), Dans des conditions moins idéales, 16-QAM ou QPSK est utilisée pour transférer 4 ou 2 bits par symbole.

Un symbole est également considéré comme un élément de ressources RE, six ou sept symboles consécutifs sont regroupés sur 12 sous porteuses afin de constituer un bloc de ressources PRB d'une durée de 0.5 ms, soit un slot, le nombre de symboles dans un PRB dépend du type du préfixe cyclique utilisé. (Sauter, 2011).

La figure 1.7 représente le cas de sept symboles sur 12 sous porteuses, deux slots forment une sous trame de durée de 1 ms, une sous trame représente en LTE l'intervalle d'ordonnancement ou le TTI, en d'autres termes, à chaque 1 ms, de nouvelles ressources sont attribuées par le nœud B aux différents utilisateurs, le nombre de blocs de ressources parallèles dans chaque sous-trame dépend de la bande passante du système, si par exemple pour une fréquence porteuse de 10-MHz, 600 sous porteuses sont disponibles, il en résulte, un total de 50 blocs de ressources à allouer à un ou à plusieurs utilisateurs par sous trame, dix sous trames rassemblées constituent une trame LTE de durée de 10 ms. Les ressources radio en lien descendant et montant se présentent sous la forme d'une grille temps-fréquence.

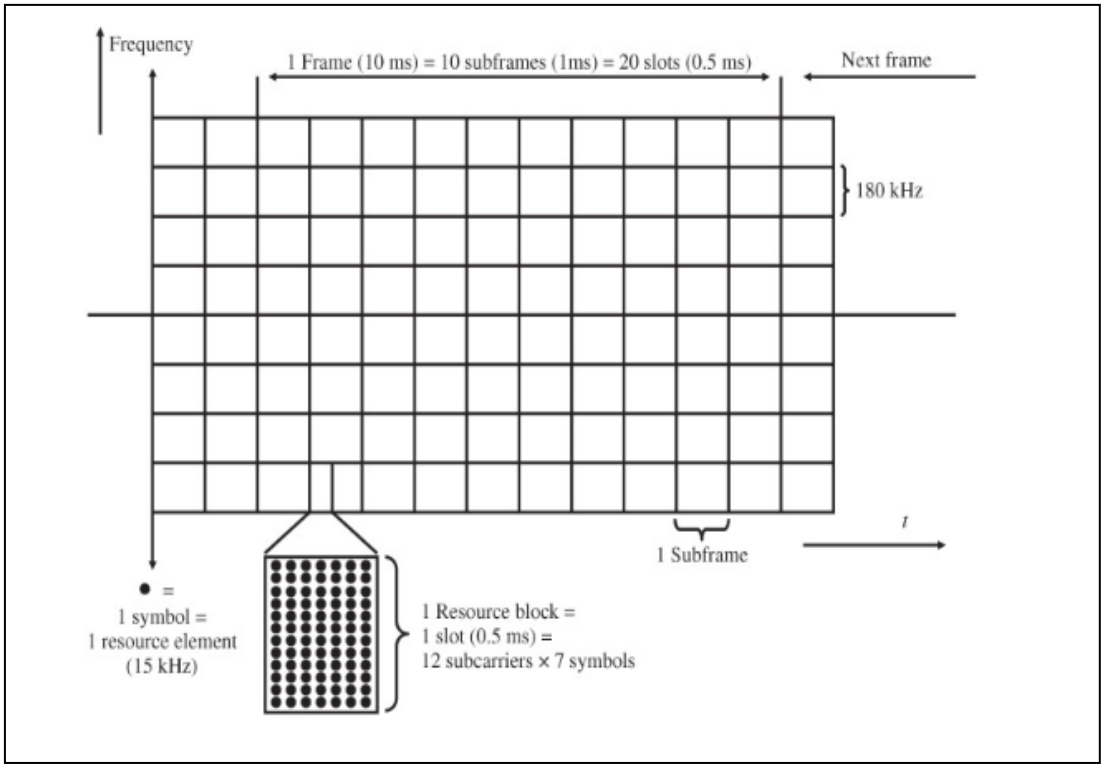


Figure 1.7 Grille de ressources LTE
Tirée de Cox (2012)

Insertion du Préfixe cyclique

Afin d'éviter l'interférence entre symboles un émetteur OFDMA insère au début de chaque symbole une copie de la dernière partie du même symbole. Le préfixe cyclique constitue aussi une période de garde afin d'empêcher la transmission des informations utiles. Deux types de préfixe cycliques peuvent exister, à savoir préfixe normal de durée de 4,7 µs et préfixe étendu dont la durée est de16,6 µs.(Cox, 2012).

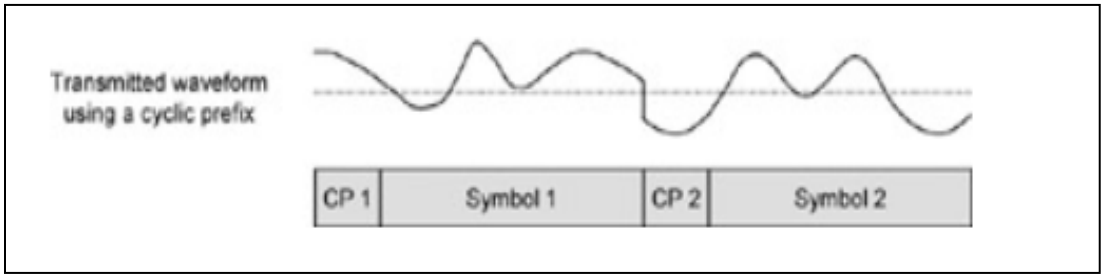


Figure 1.8 Opération d'insertion du préfixe cyclique
Tirée de Cox (2012)

b) Couche Meduim Access Control (MAC)

La couche MAC est responsable du multiplexage et démultiplexage des données entre la couche physique et la couche RLC**,** elle prend en charge la correction d'erreurs par retransmission grâce à l'implémentation de l'hybride ARQ. D'autre part, elle s'occupe des mécanismes d'accès aléatoires de la liaison montante avec le maintien de la synchronisation, la priorisation des flux sur le lien montant et enfin, l'allocation de ressources radio en s'appuyant sur les mesures effectuées par la couche physique.

1) Ordonnanceur MAC: Comme déjà mentionné, l'allocation de ressources en LTE s'effectue simultanément dans les dimensions temporelle et fréquentielle par blocs de PRB, les PRB attribués à un équipement utilisateur peuvent ne pas être adjacents. L'entité responsable de l'exécution des algorithmes d'allocation de ressources est l'ordonnanceur MAC, en fonction de sa mise en œuvre, l'ordonnanceur MAC peut allouer des ressources en se basant sur: les exigences de la qualité de service, les conditions instantanées du canal,

l'équité, etc. En outre, il veille à ce que le processus HARQ soit exécuté en temps opportun sachant que dans le réseau LTE, la retransmission des paquets est effectuée après 8 ms de la réception d'un message d'acquittement négatif. (Bianchi et Al, 2013).

2) Gestion de ressources avec agrégation de porteuses dans LTE-A: La structure de la gestion de ressources avec agrégation de porteuses dans LTE-A est décrite dans la figure.1.9.

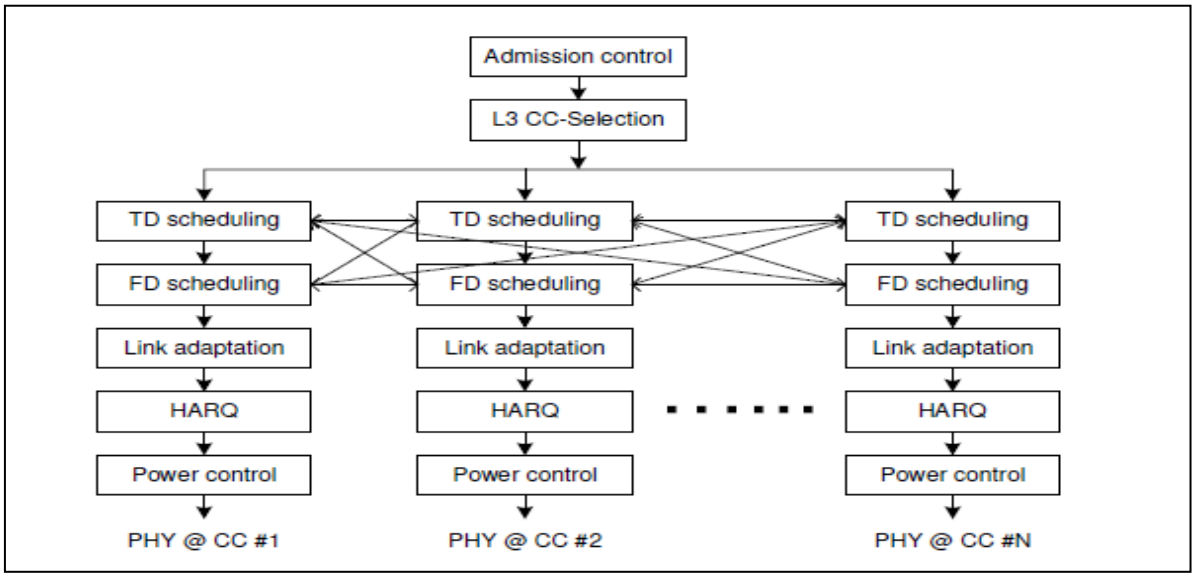


Figure 1.9 Structure de gestion de ressources avec CA dans LTE-A
Tirée de Wang, Rosa et Pedersen (2010)

Différents blocs de RRM opèrent indépendamment sur chaque composante porteuse afin de maintenir la compatibilité entre LTE-A et LTE et de permettre aux utilisateurs des deux versions de coexister. La gestion de ressources se déroule en deux étapes principales. Premièrement, le nœud B réalise un contrôle d'admission pour pouvoir affecter les composantes porteuses et planifier les utilisateurs sur toutes les composantes porteuses.

Ensuite, l'ordonnancement des paquets dans le domaine temporel et fréquentiel est effectué afin d'attribuer les PRB aux utilisateurs en considérant les exigences de la QoS, l'efficacité du système et l'équité de service entre les utilisateurs. (Wang, Claudio et Pedersen, 2010).

1.3 Disciplines d'ordonnancement dans les files d'attente

Plusieurs problèmes rencontrés dans les réseaux liés à l'allocation d'un nombre limité de ressources partagées entre les utilisateurs concurrents, les applications ou les classes de services. La gestion des files d'attente permet de gérer l'accès aux ressources de la bande passante par la sélection du prochain paquet à transmettre. Différentes disciplines d'ordonnancement dans les files d'attente sont définies où chacune essaye de trouver la meilleure balance entre la complexité, le contrôle et l'équité. (Chuck, 2001).

1.3.1 File d'attente First-In, First-Out (FIFO)

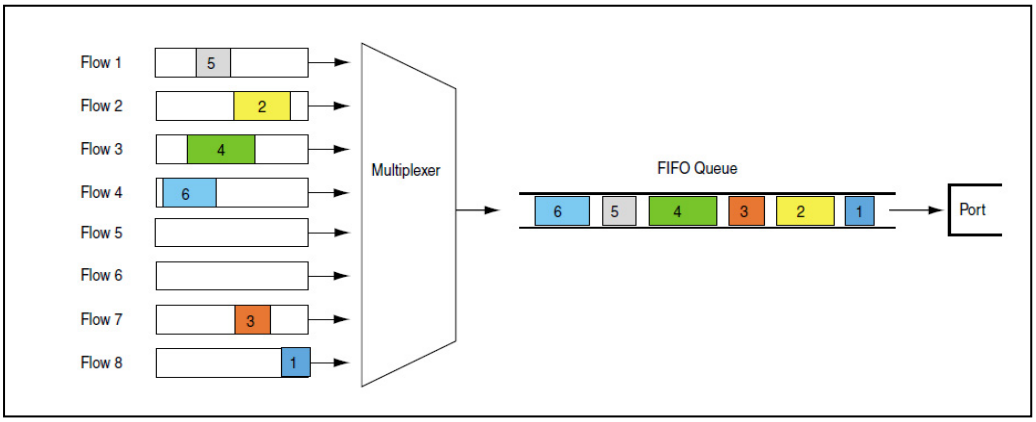


Figure 1.10 File d'attente First-In, First-Out
Tirée de http://users.jyu.fi/~timoh/kurssit/verkot/scheduling.pdf
Consulté (septembre 2015)

La File d'attente *First-in, First-out* (FIFO) est la discipline la plus basique de la gestion des files d'attente, tous les paquets sont placés dans une seule file et sont traités équitablement, les paquets sont servis suivant l'ordre premier arrivé, premier servi d'où FIFO est également connue sous le nom *First-Come, First-Served* ou (FCFS).

1.3.2 File d'attente Priority Queuing (PQ)

Priority Queuing est la base de plusieurs algorithmes permettant de fournir des méthodes simples pour soutenir la différentiation de service. Dans *Priority Queueing*, les paquets sont classifiés par le système et placés dans différentes files d'attentes, les paquets qui se trouvent au niveau de la tête de la file d'attente sont servis uniquement si toutes les files d'attentes de plus haute priorité sont vides. Au sein de chaque file d'attente, le service des paquets est planifié suivant la discipline FIFO.

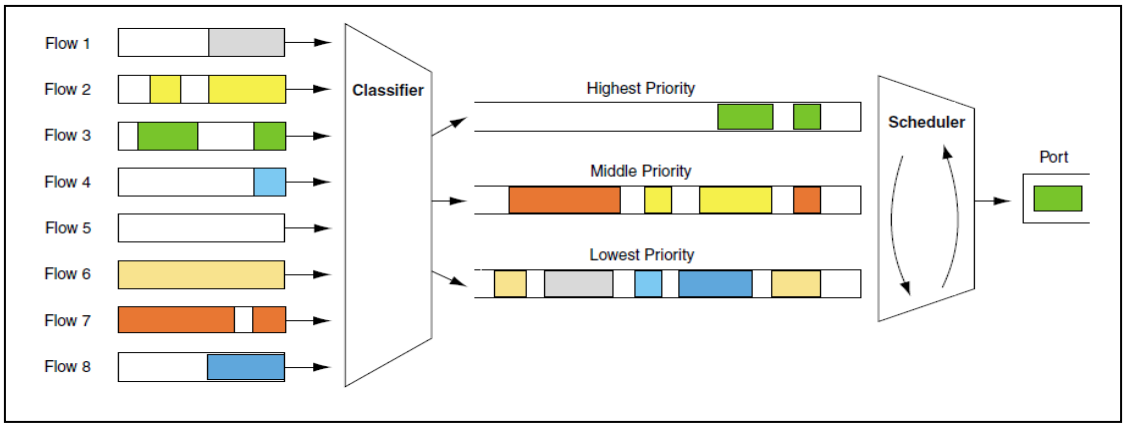


Figure 1.11 File d'attente Priority Queuing
Tirée de http://users.jyu.fi/~timoh/kurssit/verkot/scheduling.pdf
Consulté (septembre 2015)

1.3.3 File d'attente Weighted Fair Queuing (WFQ)

Elle est désignée pour soutenir les flux avec différentes exigences en termes de bande passante en attribuant des poids à chaque file d'attente afin de déterminer le pourcentage de la bande passante alloué à chaque file. Deux types de WFQ existent: *Flow based* WFQ et *Class Based* WFQ (CBWFQ). Dans CBWFQ, les flux sont groupés en des classes basées sur différents critères tels que le type de protocole et la tolérance au délai, les poids sont attribués aux classes de trafic au lieu à des flux indépendants, Les paquets à servir sont choisis en fonction du poids de la file d'attente. (Alsawaai, 2010).

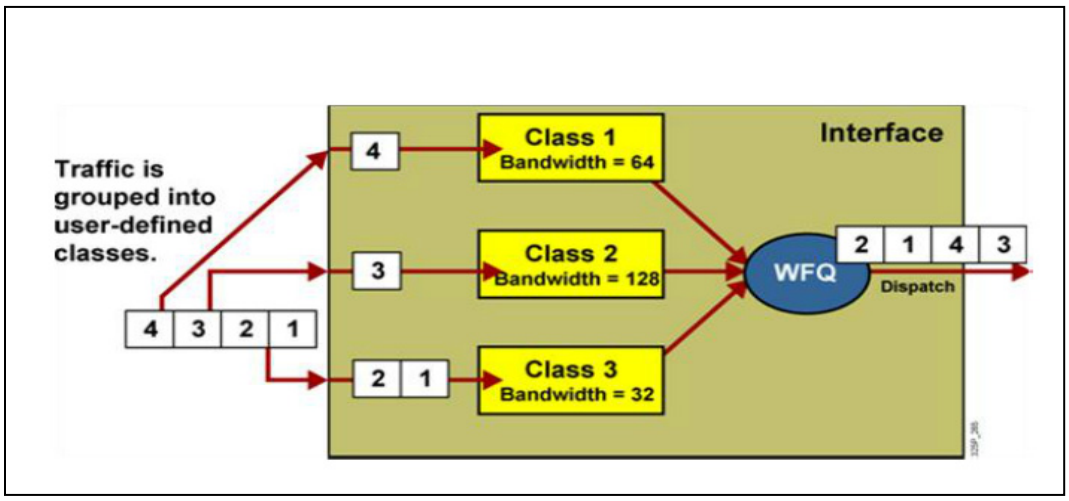


Figure 1.12 File d'attente Class Based WFQ (CBWFQ)
Tirée de http://www.myshared.ru/slide/882969/
Consulté (décembre 2015)

1.4 Conclusion

Ce chapitre commence par une présentation des caractéristiques principales de la technologie LTE/LTE-A en mettant l'accent sur l'agrégation de porteuses qui est une nouvelle technologie d'accès radio adoptée dans les systèmes LTE-A. Ensuite, le chapitre décrit les deux parties constituant l'architecture du réseau, à savoir l'E-UTRAN et l'EPC**.** Par ailleurs, le concept de la qualité de service a été présenté à travers la mise en place des porteurs pour chaque type de flux en tenant compte des paramètres ARP et QCI qui permettent d'attribuer différentes priorités aux porteurs lors de leurs établissements et aux flux durant le traitement. D'autre part, la technologie LTE définit une sous-fonction RAC au niveau du nœud B afin de contrôler l'admission d'un nouveau porteur. L'admission d'un nouveau porteur nécessite l'allocation au préalable de ressources pour garantir la QoS appropriée à la classe de trafic véhiculé dans ce porteur. Le mécanisme d'allocation de ressources radio est également discuté en présentant d'abord la technique de modulation qui constitue une fonctionnalité primordiale de la couche physique du nœud B. Ensuite, l'ordonnanceur des paquets MAC dont la fonction est d'exécuter les algorithmes d'allocation de ressources est décrit. Le chapitre présente également dans la seconde partie les différentes politiques d'ordonnancement dans les files d'attente, notamment FIFO, PQ, FQ et CBWFQ qui seront utilisées dans la conception de la nouvelle approche proposée.

CHAPITRE 2

REVUE DE LA LITTÉRATURE

2.1 Introduction

Les systèmes LTE/LTE-A sont conçus pour servir diverses classes de trafic à travers les réseaux à commutation de paquets basés IP. En raison des exigences incompatibles en termes de QoS pour chaque classe de trafic, les systèmes LTE/LTE-A sont dotés de mécanismes d'ordonnancement permettant de soutenir la différentiation de service lors de l'attribution des ressources blocks. Comme le standard 3GPP n'exige pas l'adoption d'une approche particulière, la conception d'ordonnanceurs est laissée ouverte aux chercheurs et concepteurs.

Ce chapitre porte en premier lieu sur une étude de quelques travaux de recherche ayant abordé la gestion de ressources dans les réseaux LTE/LTE-A. L'étude présente une classification d'ordonnanceurs dans le lien montant et s'intéresse à la catégorie d'ordonnanceurs basés QoS en raison de l'importance des paramètres de délai et de débit dans l'optimisation de la gestion de ressources. Ensuite, quelques algorithmes d'ordonnancement dans le lien descendant sont exposés dans le but de faire une analyse complète des différents aspects adoptés dans l'ordonnancement. En second lieu, l'algorithme de courtoisie d'optimisation de ressources dans le lien montant dans les réseaux WIMAX fixes est présenté. L'algorithme définit une politique de gestion des priorités permettant d'améliorer le service de trafic de basse priorité sans affecter la QoS de trafic de haute priorité. En fin, une évaluation critique des solutions existantes est réalisée en vue d'une conception d'un mécanisme d'ordonnancement robuste.

2.2 Classification des algorithmes d'ordonnancement dans la liaison montante

L'ordonnancement dans la liaison ascendante est complexe pour plusieurs raisons, Premièrement, l'équipement utilisateur a une source d'énergie limitée. Deuxièmement, il est difficile de prévoir le nombre de ressources radio dont l'équipement utilisateur a besoin afin

d'échanger les données avec la station de base. Suivant la fonction objective considérée et les classes de trafic véhiculées dans les canaux radio, les auteurs de (Bendaoud, Abdennebi et Didi, 2014) définissent quatre familles d'ordonnanceurs.

2.2.1 Ordonnanceurs legacy

Cette famille comprend le célèbre algorithme classique, l'algorithme Round Robin (RR), qui consiste à diviser les PRB disponibles en groupes de N_{PRB} puis à répartir les groupes formés sur les différents UE disponibles (N_{UE}).

$$\frac{N_{PRB}}{N_{UE}} \tag{2.1}$$

2.2.2 Ordonnanceurs Best Effort

Leur objectif est de maximiser l'utilisation des ressources radio et d'assurer l'équité entre les différentes classes de trafic en se basant sur la métrique *Propotionnal Fair*. Le mot *best effort* signifie que les algorithmes qui figurent dans cette classe sont des algorithmes gloutons visant d'optimiser la gestion des ressources.

2.2.3 Ordonnanceurs d'optimisation de puissance

Le but principal de cette classe d'algorithmes est d'étendre la durée d'activité de l'équipement utilisateur en minimisant la puissance du signal transmis. Les ordonnanceurs appartenant à cette famille appliquent généralement certains traitements de la QoS de sorte qu'ils réduisent la puissance d'émission afin de maintenir les exigences minimales de la QoS.

2.2.4 Ordonnanceurs basés QoS

La catégorie d'ordonnanceurs basés QoS prend en considération les paramètres de délai, de débit maximum et de nombre d'utilisateurs desservis pour offrir la QoS requise aux

utilisateurs. L'ordonnancement basé QoS a fait l'objet de plusieurs études. Les algorithmes présentés dans la section suivante donnent un aperçu des différentes solutions proposées dans la littérature.

2.3 Algorithmes d'ordonnancement basés QoS

L'étude effectuée dans (Ben Ali, Zarai et Kamoun, 2010) est une combinaison d'un procédé adaptatif de contrôle d'admission et d'un algorithme d'ordonnancement basé QoS afin de réduire la probabilité de blocage des appels *handoff*. Le trafic reçu est séparé en fonction de la tolérance au délai dans trois files d'attente, file d'attente NRT, file d'attente RT et file d'attente RT-TLR contenant les paquets temps réel dont la tolérance au délai est nulle.

De plus, l'algorithme définit dans l'ordre croissant un système de priorité de six classes de services: NC-NRTI, HC-NRTI, NC-TLRI, HC-TLRI, NC-INTLRI, HCINTLR où NC et HC désignent respectivement les nouveaux appels et les appels *handoff*. Le système de priorité permet de déterminer l'ordre préférentiel de traitement des classes.

D'autre part, les appels sont retenus dans les files d'attente lorsque la cellule est surchargée pour être servis selon la tolérance au délai d'attente, c'est– à– dire, l'appel avec une faible tolérance est traité en premier**.** Dans le cas où deux requêtes arrivent simultanément dans deux différentes files, leurs délais d'attente sont comparés avec le délai maximum toléré pour chaque classe. Afin de privilégier les appels *handoff* dans les situations de congestion, l'algorithme de réservation de ressources basé QoS est exécuté par le nœud B. Il consiste à dégrader le nombre de PRB alloué aux appels de basse priorité qui exigent une large bande passante. Le nombre de PRB résultant après la dégradation ne doit pas être inférieur à PRBmin. Si le nombre de PRB obtenu pour servir l'appel *handoff* est insuffisant, l'appel *handoff* est bloqué. Par ailleurs, pour traiter les nouveaux appels retenus dans les files d'attente quand la cellule est surchargée, le nombre de PRB attribué aux appels NRT est dégradé pour atteindre le nombre de PRB nécessaire pour accepter un nouvel appel. Le nouvel appel est rejeté si le nombre de PRB est au-dessous de nombre de PRB requis.

L'algorithme proposé a été comparé avec l'algorithme de référence, les résultats ont démontré une diminution de la probabilité de blocage pour les appels *handoff* lorsque le nombre des utilisateurs dans le réseau dépasse un certain seuil. De plus, l'utilisation de ressources a augmenté suite à l'application de l'algorithme de réservation de ressources.

Une version avancée de l'algorithme d'ordonnancement *Proportional Fair* est développée dans (Dhameliya, Bhoomarker et Zafar, 2014). Son but est d'améliorer le débit dans le lien montant des utilisateurs LTE-A qui se trouvent aux bords de la cellule et qui ont un SINR faible. L'algorithme se déroule en trois étapes

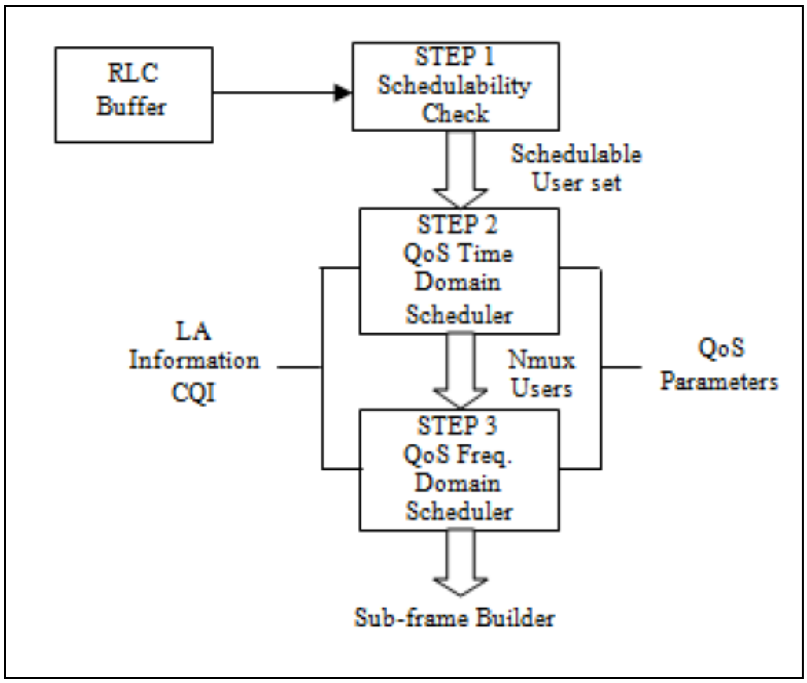


Figure 2.1 Diagramme de l'ordonnanceur des paquets
Tirée de Dhameliya, Bhoomarker et Zafar (2014)

Étape 1: les utilisateurs sont sélectionnés en fonction du délai et de la valeur du buffer afin d'être planifiés à l'instant t actuel. Les autres utilisateurs sont retenus au prochain instant.

Étape 2: le but principal de cette étape est de fournir la QoS appropriée aux utilisateurs qui ont une faible tolérance au délai. Les utilisateurs sont divisés en fonction du délai d'attente

en deux groupes de priorité. Si la valeur de la latence est inférieure à un seuil de temps prédéfini, les utilisateurs sont groupés dans le groupe 1 et ont la priorité la plus faible. Dans le cas opposé, les utilisateurs sont groupés dans le groupe 2 et ont la priorité la plus élevée.
En fonction de la priorité, (Nmax +1) utilisateurs sont sélectionnés pour l'ordonnancement dans le domaine fréquentiel, car il est supposé que certains utilisateurs sont dans l'incapacité de recevoir les données.

Étape 3: L'ordonnanceur du domaine fréquentiel sélectionne les PRB puis les utilisateurs selon leurs priorités. Les PRB sont allouées de façon à maximiser le débit de chaque utilisateur.

Les résultats de simulation montrent que le nouvel algorithme *Proportional Fair* améliore le débit pour les utilisateurs qui ont un SINR faible en comparant avec l'algorithme *Proportional Fair* classique.

Dans (Safwat, El-Badawy,Yehya et El-motaafy, 2014) les auteurs proposent une solution pour surmonter le problème de dégradation de la QoS aux bords de la cellule causé par les interférences avec les cellules voisines**.** La solution se fonde en premier lieu sur une nouvelle politique de délimitation d'appels nommée *New Call Bounding* dont le principe est de rejeter un nouvel appel quand le nombre des nouveaux appels admis dans la cellule dépasse un certain seuil M. L'appel *handoff* est rejeté uniquement lorsque tous les canaux de la cellule sont utilisés**.** La deuxième phase de la solution est d'allouer les ressources aux utilisateurs se trouvant aux bords de la cellule selon le schéma *Software Frequence Reuse* qui divise la cellule en deux parties, la partie bords de la cellule et la partie cœur de la cellule. En revanche, les PRB et les utilisateurs sont pareillement subdivisés en PRB et utilisateurs de bords et PRB et utilisateurs de cœur. Les utilisateurs de bords se font allouer les sous fréquences que dans le cas où toutes les bandes de fréquences sont suffisantes pour servir les utilisateurs de cœur.

Les résultats de simulation montrent que le système se comporte différemment selon l'emplacement des utilisateurs. De plus, la probabilité de blocage s'accroît quand le taux d'arrivées des nouveaux utilisateurs et le nombre maximum des nouveaux appels admis augmentent. Par ailleurs, La probabilité de résiliation des appels augmente quand la valeur du seuil M augmente où M désigne le nombre des nouveaux utilisateurs à accepter.

Dans (Vassilakis, Moscholiosy, Bontozoglouz, et Logothetisx, 2015), les chercheurs introduisent la technique *Software-defined networking* (SDN) dans la gestion de ressources radio des réseaux LTE-A et des réseaux de future génération 5G. Le but de cette proposition est de permettre aux macros stations de base (macro-BS) d'allouer les ressources requises aux micro-stations de base (sc-BS) et d'assurer la QoS aux utilisateurs *handoff* lors du changement de la micro cellule.

Le modèle considéré est un réseau hétérogène *HetNet* avec une macro cellule contrôlée par une macro-BS et couvrant Si petites cellules dont chacune est contrôlée par une sc-BS. La macro cellule est équipée d'un contrôleur SDN qui alloue dynamiquement les ressources aux micros cellules. Le modèle considère aussi K classes de service indépendantes représentant chacune le niveau de la QoS requis pour les utilisateurs mobiles MU. Les utilisateurs MU sont de deux types, soit de nouveaux utilisateurs tentant d'établir une connexion initiale, soit des utilisateurs *handoff* dont les appels sont transférés d'une autre cellule. Le contrôleur SDN facilite le *handoff* entre les micros cellules en allouant les ressources radio appropriées. Quand les ressources dans la micro cellule ne sont pas suffisantes, la macro cellules attribue ses ressources inoccupées ou les ressources d'autres micros cellules, si les ressources disponibles demeurent insuffisantes, l'utilisateur est bloqué. Par ailleurs, la macro cellule est dotée d'un mécanisme de control d'admission pour que la probabilité de blocage des appels *handoff* soit moins élevée que la probabilité de blocage des nouveaux appels.

Les résultats de simulation coïncident avec les résultats analytiques. La probabilité de blocage des appels *handoff* est moins élevée que la probabilité de blocage des nouveaux appels en raison de la politique d'ordonnancement adoptée qui favorise les appels *handoff*.

De nouvelles approches ont été considérées avec l'émergence de la technologie LTE-A. L'algorithme (Chen, Hsu, Tsai, Liao, Lin, 2014) dans la liaison montante décrit une gestion de ressources entre les utilisateurs LTE et LTE-A afin de minimiser les interférences entre les porteuses. Les utilisateurs LTE-A sont alloués des PRB des différentes composantes porteuses, alors que les PRB dédiés aux utilisateurs LTE appartiennent à une seule composante. Cependant, la contrainte de sous porteuses adjacentes doit être maintenue dans l'attribution des PRB aux utilisateurs LTE-A car le système SC-FDMA localisé est adopté.

L'allocation des PRB se base sur le concept *Gale-Sharply matching* qui consiste à localiser le PRB approprié comme un point de départ pour l'allocation continue des PRBs. En outre, l'approche définit les concepts suivants dans la conception de la solution:

- bien choisir le RB de départ pour les utilisateurs LTE et LTE-A parmi tout le spectre de toutes les composantes;
- les utilisateurs LTE détiennent la plus haute priorité;
- un utilisateur LTE-A ne peut pas préempter un PRB déjà assigné à un utilisateur LTE. Par contre, un utilisateur LTE peut préempter un utilisateur LTE-A car les utilisateurs LTE-A peuvent être alloués des PRBS d'autres composantes porteuses.

Les performances de la solution ont été évaluées en considérant trois modèles de trafics, à savoir la VoIP, le vidéo streaming et le FTP où la VoIP et le vidéo streaming spécifient les services GBR. La quantité de la bande passante sollicitée par chaque utilisateur a été bien satisfaite sauf pour les utilisateurs LTE-A demandant le service vidéo streaming lorsque la cellule est saturée.

2.4 Ordonnancement dans la liaison descendante

Les algorithmes d'ordonnancement dans le lien descendant ont pour objectif d'optimiser les performances du système. Plusieurs recherches ont été fondées sur le concept *prorptional fair* dont le but est de maximiser le débit global du système en augmentant le débit de chaque

utilisateur. En outre, le procédé vise à garantir l'équité entre les utilisateurs. Dans (Kausar, Chen et Chai, 2011) un mécanisme d'ordonnancement de paquets dans la liaison descendante a été présenté afin de satisfaire les exigences incompatibles de la QoS des différents types de trafic. Le mécanisme proposé se déroule en trois étapes principales, gestion de la file d'attente, ordonnancement dans le domaine temporel puis ordonnancement dans le domaine fréquentiel.

Le trafic reçu est d'abord séparé dans quatre files d'attente relatives aux trafics de haute priorité, de faible latence, de haut débit et de faible priorité. Ces quatre types de trafics correspondent respectivement aux trafics de contrôle, de la voix, du vidéo streaming et de *best effort*. L'algorithme de tri d'utilisateurs par type de service est appliqué au niveau des files d'attente dans le but d'améliorer la QoS. Le trafic de contrôle d'information est le plus important. Néanmoins, il est transmis selon la technique de Round Robin afin d'assurer l'équité entre les utilisateurs. Le temps d'attente est pris comme référence pour trier les utilisateurs RT et NRT. Le trafic *best effort* n'a pas d'exigences en termes de paramètres de QoS. Par contre, l'algorithme de *Proportional Fair* est utilisé lors de tri des utilisateurs afin de maintenir l'équité. La deuxième étape de l'algorithme est l'ordonnancement dans le domaine temporel qui consiste à déterminer le nombre d'utilisateurs à servir en fonction de la quantité des ressources réservées aux trafics RT et NRT. La dernière étape de l'algorithme est l'ordonnancement dans le domaine fréquentiel où les PRB sont allouées aux utilisateurs sélectionnés.

Les résultats de simulation montrent une diminution au niveau du délai du trafic RT. En outre, le débit moyen du trafic NRT est maintenu à un bon niveau.

Une autre variante d'ordonnancement dans le lien descendant est proposée dans (Iturralde, Martin et Ali Yahiya, 2013). Le but de l'approche est de garantir la QoS aux utilisateurs du réseau LTE. Le scénario d'allocation de ressources est modélisé comme un processus de prise de décision multicritères *Multi-Criteria Decision Making* où une seule entité est

chargée de prendre la décision en tenant compte de plusieurs paramètres lorsqu'il y'a différents objectifs incompatibles de la QoS.

Le nœud B est considéré le décideur qui vise à atteindre certains objectifs tels que la réduction du délai, l'augmentation du débit et la réduction de la perte de paquets lors de la transmission de données. Tandis que, les paramètres utilisés sont les conditions du canal, la taille de la file d'attente du paquet, la modulation et la valeur du schéma de codage MSC. De ce fait, plusieurs ensembles sont définis en fonction des paramètres dans le but de dériver la portion des ressources blocs à allouer pour chaque utilisateur.

L'algorithme *Pondering Parameters scheduling* a été comparé avec l'algorithme *First Maximum Expansion* qui se base sur le concept *proportional fair* dans l'ordonnancement. Les classes de trafic simulées sont citées suivant un ordre décroissant de priorité, le vidéo streaming, la VoIP et le trafic NRT dont les portions de la bande passante affectée aux divers trafics sont 40%, 40% et 20% respectivement. Le nouveau procédé possède le meilleur indice d'équité. En outre, il améliore le délai quand le nombre d'utilisateurs augmente dans le système d'où la perte de paquets diminue et le débit s'accroît.

L'agrégation des porteuses est l'une des nouveautés apportées par la technologie 4G et qui a grandement influencé la gestion des ressources. Dans (Miao, Min, Jiang, Jin et Wang, 2014), le problème de charges déséquilibrées entre les composantes porteuses est traité afin de soutenir les services temps réels. La nouvelle solution de gestion de ressources appelée *Cross-CC User Migration* se déroule en trois étapes. D'abord l'ordonnanceur de haut niveaux définit la quantité des données que chaque utilisateur temps réel doit transmettre afin de satisfaire la contrainte du délai. Il calcule pour chaque utilisateur deux types de transmission trame-par-trame et diverses longueurs de files d'attente. Les transmissions calculées sont quota des paquets pour *Head-of-line* (HOL) et quota de *Long- Term-Perspective* (LTP).

Ensuite, l'ordonnancement de bas niveau est appliqué pour assurer l'équité entre les utilisateurs, son rôle est d'attribuer les PRB en utilisant l'algorithme *Proportional Fair* pour remplir les quotas de transmission. En outre, les paquets HoL sont servis en premier pour éviter leur perte.

En fin, le nœud B continue à allouer les PRB aux utilisateurs qui ont des quotas incomplets et exécute l'ordonnancement au niveau des composantes porteuses afin de réduire le degré de déséquilibre entre les composantes et augmenter l'efficacité du système.

Le procédé proposé a été mis en œuvre au niveau de la liaison descendante des réseaux LTE-A. Les résultats obtenus sont comparés avec les résultats de l'algorithme *Two-Level Downlink Scheduling* qui a prouvé ses performances par rapport aux autres procédures d'ordonnancement. L'algorithme *Cross-CC User Migration* donne les meilleurs résultats, il améliore le débit de chaque utilisateur et réduit le délai et la longueur moyenne de la file d'attente.

2.5 Algorithme de courtoisie: optimisation de la performance dans les réseaux WIMAX fixes

L'algorithme dans (Kadoch et Tata, 2008) vise à optimiser l'utilisation de la bande passante dans les réseaux WIMAX fixes en réduisant le délai et la perte de paquets des classes de faible priorité. Il définit deux classes de QoS, à savoir la classe rtPS et la classe nrtPS relatives au trafic VoIP et au trafic FTP respectivement. Toutefois, la solution a été étendue à un système de N classes.

Dans le cas de deux classes, les paquets qui arrivent dans le réseau sont retenus dans deux files d'attente de tailles finies pour être servies selon la valeur de leur priorité. La classe rtPS possède la plus haute priorité. Cependant, l'algorithme de courtoisie est appliqué pour garantir l'équité de service, les paquets de la classe nrtPS sont servis au lieu des paquets rtPS à condition que les paquets rtPS courtois aillent suffisamment le temps pour qu'ils soient servis sans que leur QoS soit affectée. Les arrivées des classes rtPS et nrtPS dans les files

d'attente suivent la loi de Poisson de taux $\lambda_1$et de taux λ_2 et reçoivent un service de traitement suivant la loi exponentielle avec un temps moyen 1/μ paquet/sec. Le taux total d'arrivées est $\lambda = \lambda_1 + \lambda_2$.

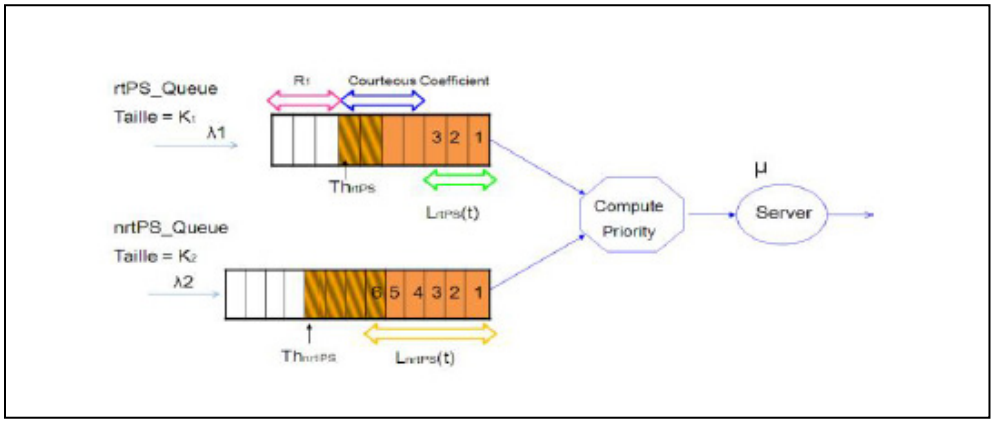


Figure 2.2 Système de file d'attente M/G/1 de l'algorithme de courtoisie
Tirée de Kadoch et Tata (2008)

La figure.2.2 représente le système de file d'attente M/G/1 avec priorité non préemptive. rtPS_Queue et nrtPS_Queue désignent les files d'attente des classes rtPS et nrtPS, ThrtPS et ThnrtPS représentent les seuils de taux de remplissage des files d'attente. La solution est ornée d'un mécanisme de calcul de priorité qui permet de déterminer le paquet à servir en premier.

Le système d'équations suivant détermine le nombre de paquets dans le système, le nombre de paquets dans les files d'attente et le délai d'attente de chaque classe après avoir appliqué l'algorithme de courtoisie. Tandis que, L1 et L2 indiquent successivement le nombre moyen de paquets rtPS et nrtPS dans le système avant la courtoisie, l_{q1} et l_{q2} représentent le nombre de paquets dans les files d'attente et finalement w_{q1} et w_{q2} désignent le temps d'attente des paquets rtPS et nrtPS :

$$L_{\text{rtPS}} = L_1 + coeff_{courtois} \tag{2.2}$$

$$L_{\text{qrtPS}} = l_{q1} + coeff_{courtois} \tag{2.3}$$

$$W_{\text{qrtPS}} = w_{q1} + \xi_1 \quad (2.4)$$

$$L_{\text{nrtPS}} = L_2 - coeff_{courtois} \quad (2.5)$$

$$L_{\text{qnrtPS}} = l_{q2} - coeff_{courtois} \quad (2.6)$$

$$W_{\text{qnrtPS}} = w_{q2} - \xi_1 \quad (2.7)$$

Avec ξ_1 le temps de tolérance d'attente supplémentaire des paquets prioritaires dans leur file d'attente rtPS_Queue, il est défini comme:

$$\xi_1 = coeff_{courtois} * \mu \quad (2.8)$$

Et $coeff_{courtois}$est le nombre de paquets attendus pour atteindre le seuil Th_{rtPS} de remplissage de la file rtPS_Queue:

$$coeff_{courtois} = \left(K_1 - \left((\lambda_1 + \sigma_{\lambda 1}) * \left(w_{q1} + \sigma_{wq1}\right)\right)\right) - L_1 \quad (2.9)$$

Où $\sigma_{\lambda 1}$ est la variance de taux d'arrivées λ_1 et σ_{wq1} est la variance de temps d'attente moyen des paquets rtPS

La solution a été implémentée au niveau de la liaison montante. Les résultats de simulation ont été comparés avec les résultats obtenus pour les politiques PQ et WFQ. L'algorithme de courtoisie se comporte de façon similaire que WFQ dans le cas où le taux de trafic voix dépasse considérablement le taux de trafic FTP. Par contre, l'algorithme de courtoisie a pu privilégier 80% des paquets de basse priorité lorsque le taux de trafic FTP est assez élevé par rapport au taux de trafic voix d'où une réduction du délai et de perte de paquets pour la classe de basse priorité.

2.6 Synthèse et limites des solutions existantes

L'objectif de la gestion de ressources est d'améliorer les performances des systèmes LTE/LTE-A. De ce fait, maintes approches ont été développées et sont axées sur différents aspects tels que la gestion de files d'attente ou la gestion des priorités. Certaines recherches prennent en compte la tolérance au délai dans la gestion des files d'attente afin d'offrir un traitement différencié aux classes de trafic. Les classes de faible tolérance ont la plus haute priorité de service. Cependant, plusieurs travaux de recherche considèrent la gestion des priorités comme un calcul de priorités sans définir un mécanisme qui permet d'attribuer des priorités variables aux différentes classes de trafic. Par conséquent, les classes de faible priorité peuvent expérimenter un taux de blocage et de perte de paquets assez élevé.

D'autres études privilégient le trafic RT par rapport au trafic NRT, notamment l'algorithme de *propotional fair* présenté dans (Kausar, Chen et Chai, 2011) où le trafic RT est avantagé au détriment du trafic NRT. Bien que dans (Dhameliya, Bhoomarker, Zafar, 2014) une version avancée de l'algorithme a été développée. Toutefois, aucune amélioration n'a été apportée en faveur de trafic de basse priorité.

D'autre part, plusieurs procédés de gestion de ressources favorisent les appels *handoff* par rapport aux nouveaux appels. L'approche présentée dans (Ben Ali, Zarai et Kamoun, 2010) se fonde sur le concept de réservation de ressources pour réduire la probabilité de blocage des appels *handoff* en particulier dans la situation de congestion. Ainsi que la solution proposée dans (Vassilakis, Moscholiosy, Bontozoglouz, et Logothetisx, 2015) qui introduit un mécanisme pour que la probabilité de blocage des appels *handoff* soit moins élevée que la probabilité de blocage des nouveaux appels dans le cas où les ressources sont insuffisantes.

La solution exposée dans (Safwat, El-Badawy,Yehya et El-motaafy, 2014) a tenté de soutenir le trafic de basse priorité, d'où elle consiste à limiter le nombre de nouveaux appels acceptés quand il dépasse un certain seuil, les taux de blocage des appels *handoff* et de nouveaux appels peuvent être décidés en fonction de la valeur du seuil. Néanmoins, cette solution reste

limitée, car si la probabilité de blocage est améliorée, la probabilité de résiliation des appels *handoff* est dégradée et vice versa.

La gestion des priorités devrait également être introduite dans **(**Chen, Hsu, Tsai, Liao et Lin, 2014**)** afin d'assurer une gestion efficace de ressources. Cependant, dans l'étude les utilisateurs LTE détiennent la plus haute priorité, alors que les utilisateurs LTE-A sont les premiers à être bloqués quand la cellule est surchargée. Les approches dans (Iturralde, Martin et Ali Yahiya, 2013) et (Miao, Min, Jiang, Jin et Wang, 2014) ne définissent pas de mécanisme de gestion des priorités. La première étude offre une solution pour calculer le nombre de ressources à allouer pour chaque utilisateur en tenant compte des exigences de la QoS. Tandis que dans la deuxième approche, une nouvelle gestion de composante porteuse a été développée visant à optimiser le débit et du délai des utilisateurs RT.

L'algorithme d'optimisation de la performance dans les réseaux WIMAX fixes proposé par Michel KADOCH et Chafika TATA est doté d'un mécanisme de courtoisie permettant une gestion dynamique des priorités afin d'assurer l'équité de service. Les paquets de haute priorité cèdent leurs tours de transmission aux paquets de basse priorité tant que la QoS des paquets de haute priorité n'est pas affectée.

Dans le prochain chapitre, une nouvelle étude axée sur la courtoisie est présentée avec le cas de trois classes de trafic dans le but d'optimiser la gestion de ressources et d'améliorer les performances des systèmes LTE/LTEA.

CHAPITRE 3

ALGORITHME DE COURTOISIE: ORDONNANCEMENT DANS LA LIAISON MONTANTE DES RESEAUX LTE-ADVANCED

3.1 Introduction

Les réseaux LTE/LTE-A offrent des services différentiés en fonction des priorités attribuées aux différents types de trafics. L'attribution des priorités est basée sur plusieurs critères tels que la tolérance au temps d'attente dans les files ou si le trafic est nouveau ou transféré. Habituellement, le trafic RT est prioritaire par rapport au trafic NRT et les appels *handoff* sont privilégiés par rapport aux nouveaux appels, ce qui engendre des taux de blocage et de perte de paquets élevés expérimentés par les trafics de basse priorité, en particulier dans le cas de congestion du réseau. De ce fait, l'algorithme de courtoisie a été développé afin d'assurer l'équité de service aux diverses classes de trafic à travers une nouvelle gestion des priorités qui permet au trafic moins prioritaire d'être privilégié à condition que la qualité de service du trafic de haute priorité ne soit pas affectée. En outre, il vise également à optimiser le débit dans le système, à minimiser le taux de blocage et de perte de paquets et à réduire le délai d'attente des classes défavorisées.

Les différentes étapes de conceptions de l'algorithme de courtoisies sont détaillées dans les prochaines sections de ce chapitre.

3.2 Algorithme de courtoisie: Ordonnancement dans la liaison montante

Trois types de classes sont définies représentant le trafic *handoff*, et les nouveaux trafics RT et NRT. La classe *handoff* détient la plus haute priorité et toujours servie en premier afin d'éviter de préempter les porteurs en cours de service dans les situations de congestion, alors que la classe NRT possède la priorité la plus faible. Ensuite, le mécanisme de courtoisie est appliqué pour réorganiser la gestion d'accès aux ressources et accorder aux deux autres classes la plus haute priorité dans le but d'assurer l'équité de service. De ce fait, deux

conceptions de l'algorithme de courtoisie ont été développées, la première conception consiste à privilégier les paquets RT lorsque leur temps d'attente moyen est atteint. La durée de transmission des paquets RT ne doit pas dépasser le temps d'attente moyen des paquets *handoff* afin de maintenir la QoS de la classe *handoff*.

Dans la deuxième conception, les paquets RT cèdent la plus haute priorité aux paquets NRT, qui seront favorisés également quand leur temps d'attente moyen sur une valeur *x* choisie est abouti et durant une période incluse dans le même intervalle de transmission des paquets RT selon PQ. La conception du procédé est basée sur un système Markovien M/M/S/K décrit dans ce qui suit.

3.2.1 Description du système M/M/S/K

Le système M/M/S/K considéré est Markovien, il est composé de *S* serveurs analogues et de buffer unique de capacité finie K partagé entre trois files d'attente virtuelles, à savoir Q_HD, Q_RT et Q_NRT relatives aux classes de trafics *handoff*, RT et NRT. La plus haute priorité est donnée à Q_HD grâce à la politique d`ordonnancement *Priority Queueing* (PQ) non préemptive qui permet de traiter les paquets *Handoff* temps réel et non-temps réel avant les nouveaux paquets sans interruption de service lorsqu'un nouveau paquet est en cours de transmission. Les nouveaux paquets dans Q_RT et Q_NRT sont prévus d'être servis selon la technique *Class Based* WFQ (CBWFQ) avec le poids w_1 de Q_RT est supérieur au poids w_2 de Q_NRT dans le but de favoriser le trafic RT. Les paquets de la même classe sont servis suivant l'ordre – premier arrivée premier servis ou FCFS.

Les arrivées dans les trois files d'attente suivent la loi de poisson de moyenne λp, avec $p = 0, 1,2$, et nécessitent un temps de service exponentiel $1/s\mu$ où $\mu = \mu_p$, $w_1\mu_1$ ou $w_2\mu_2$, et *S* représente le nombre de serveurs considéré et qui doit être inférieur ou égale à la capacité du buffer K.

Lorsque le buffer atteint sa taille maximale, les nouvelles demandes d'accès au système sont rejetées. Par ailleurs, il est exigé que la discipline du service dans chaque lien *S* avec *S*=1,2,…K, soit avec conservation de service où le lien ne peut être en mode repos (IDLE) tant qu'il y'a des paquets en attente dans les files. La seconde contrainte à satisfaire est d'empêcher plusieurs serveurs de travailler sur le même paquet, en d'autres termes, un paquet est servi par un seul serveur**.** (Gautam, 2012).

Les arrivées λ_0 dans la queue Q_HD, peuvent être de type RT ou NRT, la gestion au sein de cette file suit la politique CBWFQ, la raison de représenter le trafic *handoff* par une seule file d'attente Q_HD est que la superposition de deux processus de Poisson de paramètres λ_{01} et λ_{02} est un processus de poisson de paramètre λ_0 où λ_{01}, λ_{02} indiquent successivement les arrivées des paquets *handoff* RT et *handoff* NRT. (Chaudet, 2015).

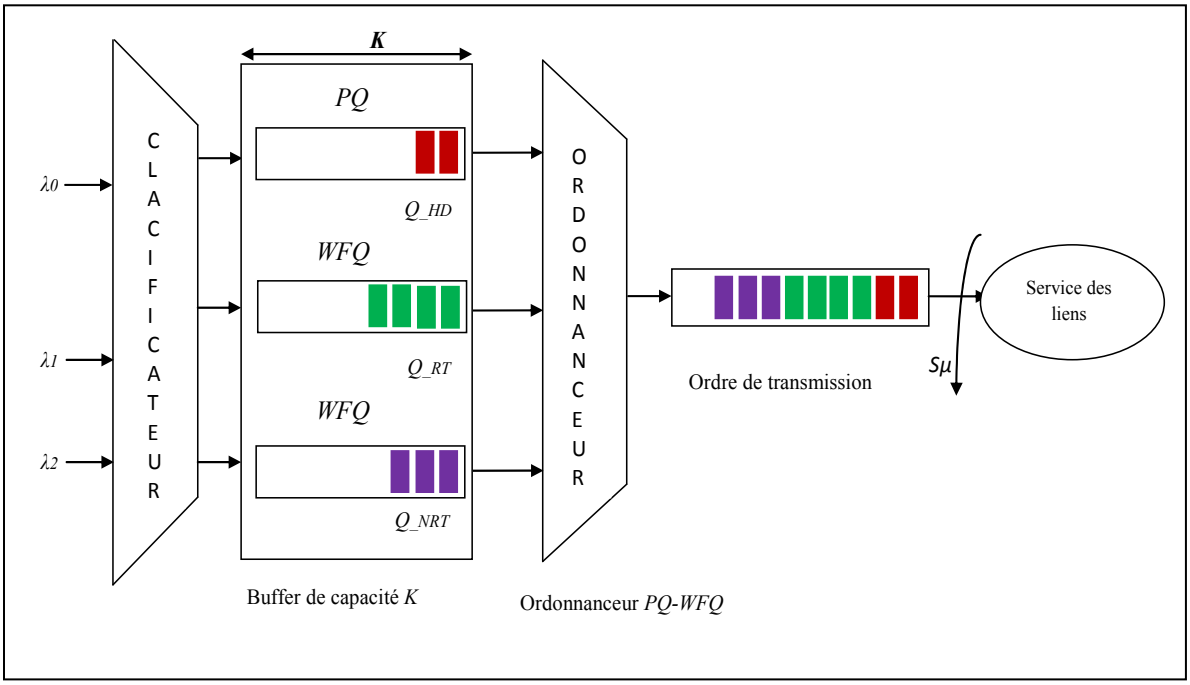


Figure 3.1 Schéma du système M/M/S/K avec PQ-CBWFQ

La figure 3.1 représente le schéma du système M/M/S/K avec PQ-CBWFQ, $S\mu$ désigne les différents taux de services destinés aux trois files d'attente. Bien que cette distinction assure un certain degré de traitement pour chaque classe, les paquets *handoff* peuvent monopoliser

le lien. L'algorithme de courtoisie proposé repose sur ce schéma pour conditionner le service des différents flux.

3.2.2 Conditions d'application de l'algorithme de courtoisie

L'algorithme de courtoisie permet aux diverses classes de trafic de détenir alternativement la plus haute priorité dans des périodes bien définies. Les conditions d'application de l'algorithme permettent de délimiter ces périodes en vue de garantir la qualité de service pour chaque type de classe.

3.2.2.1 Condition 1

La première condition vise à déterminer un ordre initial de traitements des classes *handoff*, RT et NRT en leur attribuant successivement les priorités comme suit:

$$PR_1 > PR_2 > PR_3 \tag{3.1}$$

3.2.2.2 Condition 2

Les paquets *handoff* sont servis en premier tant que le délai d'attente moyen dans la file RT n'est pas encore atteint. Les paquets *handoff* sont transmis pendant une période T_{r1} inférieure ou égale au $seuil_{RT}$

$$T_{r1} \leq seuil_{RT} \tag{3.2}$$

Le $seuil_{RT}$ est choisi comme étant le temps d'attente moyen des paquets RT dans la file Q_RT, ce choix se justifie par la possibilité de faire éviter les paquets RT une attente excessive qui peut dépasser le temps d'attente maximal permis dans leur file et par conséquent, la perte de paquets, T_{r1} sera prise comme référence pour débuter le service des paquets RT selon PQ.

3.2.2.3 Condition 3

Les paquets *handoff* sont retenus dans la file d'attente durant la période de transmission T_{r2} des paquets RT qui ne doit pas excéder $seuil_{HD}$, le temps d'attente moyen des paquets *handoff* dans Q_HD. Cette contrainte a pour but d'empêcher la perte de paquets *handoff*.

$$T_{r2} \leq seuil_{HD} \tag{3.3}$$

Des conditions 2 et 3, il en résulte que les paquets RT auront la plus haute priorité à partir de l'instant t égale au $seuil_{RT}$ et pendant une période qui peut durer au maximum un $seuil_{HD}$

$$seuil_{RT} \leq T_{r2} \leq seuil_{HD} + seuil_{RT} \tag{3.4}$$

La figure 3.2 schématise les contraintes à satisfaire dans la deuxième et la troisième condition.

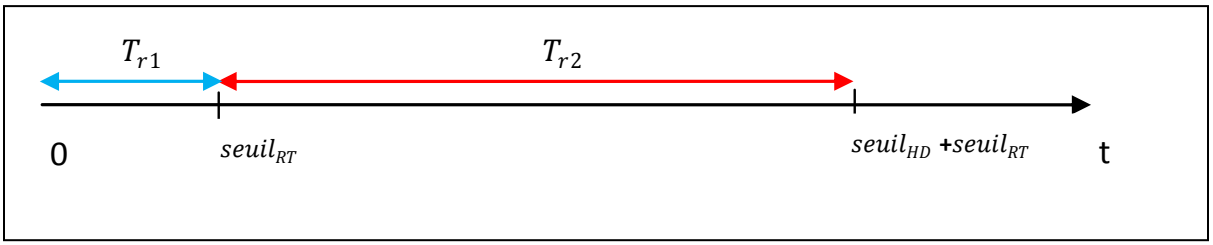


Figure 3.2 Intervalle de service des paquets RT selon PQ

3.2.2.4 Condition 4

Cette condition concerne la deuxième conception de l'algorithme, elle consiste à temporiser le service des paquets RT comme prioritaire et servir les paquets NRT au cours de l'intervalle T_{r2} suivant la politique PQ. Le changement des priorités s'initie lorsque le temps d'attente des paquets NRT aboutit à un $seuil_{NRT}$ équivaut au temps d'attente moyen des paquets NRT divisé par une valeur *x* choisie. Le but de cette division est d'avoir un $seuil_{NRT}$ du même ordre de grandeur que les $seuil_{HD}$ et $seuil_{RT}$ vu que la classe NRT tolère des délais

plus élevés. Ainsi, les paquets RT sont prioritaires durant un temps Tr_3 inclus dans l'intervalle T_{r2} et inférieur ou égal au $seuil_{NRT}$.

$$T_{r3} \leq seuil_{NRT}. \tag{3.5}$$

Avec:

$$seuil_{NRT} = (temps\ d'attente\ moyen\ NRT)/x \tag{3.6}$$

3.2.2.5 Condition 5

Les paquets NRT sont servis selon PQ jusqu'au l'accomplissement de la durée T_{r2} permise.

$$seuil_{NRT} \leq T_{r4} \leq seuil_{HD} + seuil_{RT} \tag{3.7}$$

Où T_{r4} est l'intervalle de service des paquets NRT selon PQ.

La figure 3.3, illustre les conditions 4 et 5

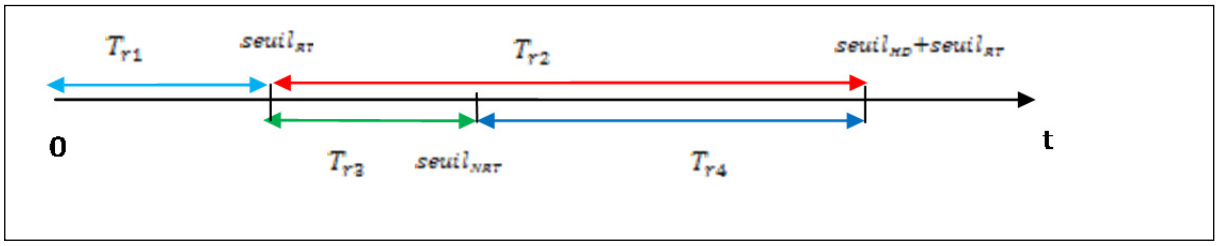


Figure 3.3 Intervalle de service des paquets NRT selon PQ

3.2.3 Modèle analytique du système M/M/S/K avec PQ_CBWFQ

Cette section décrit le modèle analytique du système M/M/S/K avec PQ-CBWFQ cité précédemment qui sera considéré comme le diagramme de base du nouveau procédé. L'état du système M/M/S/K avec PQ_CBWFQ à un instant t est modélisé à l'aide de la chaîne de

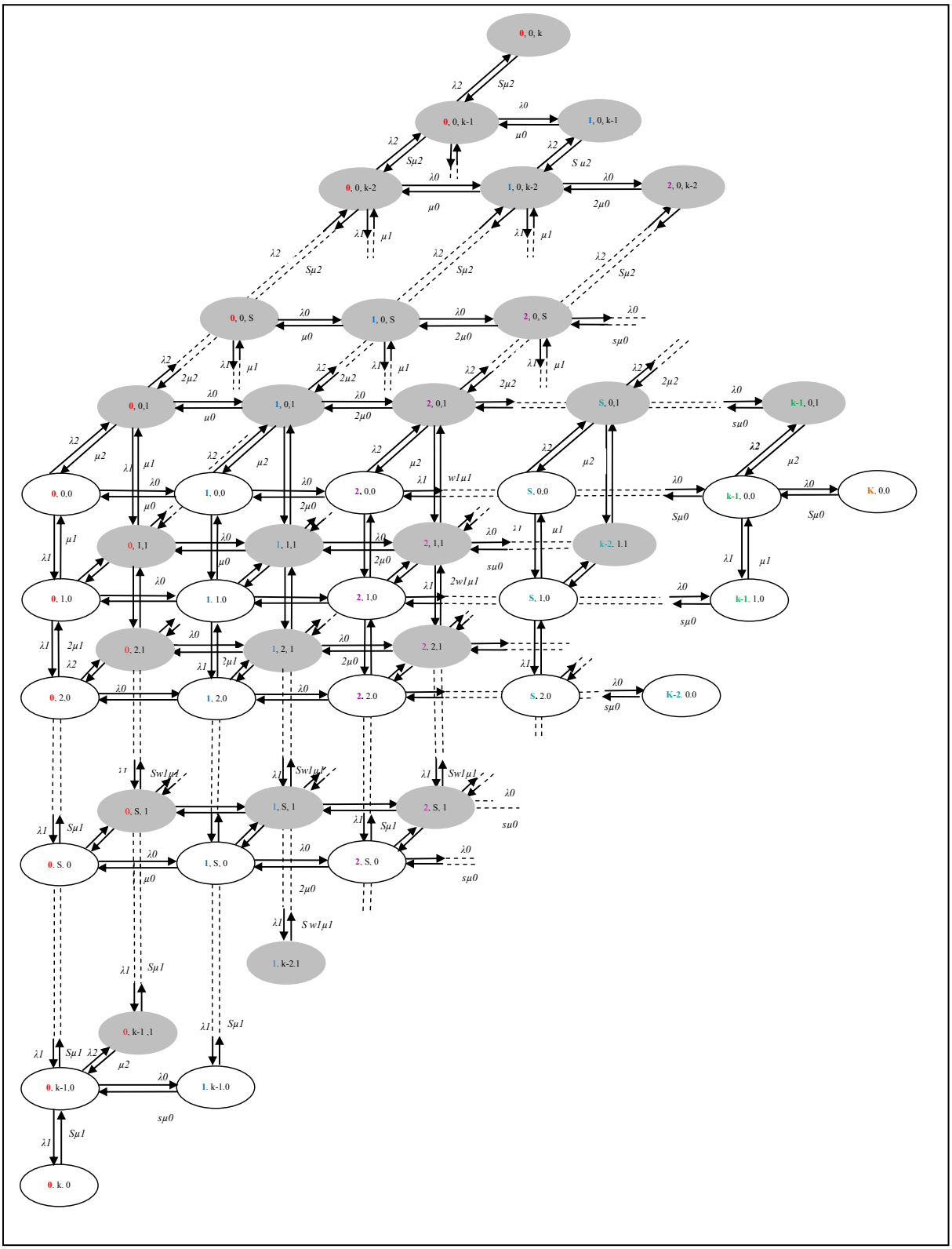


Figure 3.4 Diagramme d'état du système de files d'attente M/M/S/K avec PQ_CBWFQ

Markov discrète au temps continu à trois dimensions avec l'état (i, j, k) où *i, j, k* dénotent le nombre de paquets dans chaque état des files *handoff*, RT et NRT respectivement. Le processus stochastique résultant est représenté par le diagramme de transition de la figure 3.4 où le processus (i, j, k) peut être transféré à l'un des états suivants: $(i+1, j, k)$, $(i, j+1, k)$, $(i, j, k+1)$, $(i-1, j, k)$, $(i, j-1, k)$ et $(i, j, k-1)$.

Le nombre de paquets existant simultanément dans les trois files d'attente ne doit pas dépasser la capacité totale du buffer *K*. Par exemple: lorsqu'il y'a *m* paquets dans Q_HD et *n* paquets dans Q_RT, Q_NRT contiendra au maximum $K-(m+n)$ paquets.

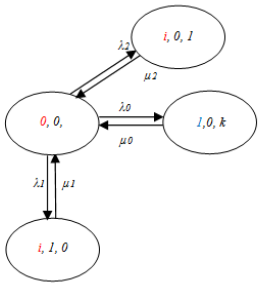
La classe *handoff* détient la plus haute priorité et servie avec un taux $S\mu_0$ suivant la politique PQ, alors que la discipline CBWFQ est appliquée pour servir les paquets RT et NRT. Lorsque Q_NRT est vide, les paquets RT sont traités avec un taux μ_1. Or, dans le cas opposé, les paquets NRT sont servis avec un taux μ_2. Par contre, si les paquets existent dans les deux files en même temps, les taux de service du lien sera distribué sur les deux classes en fonction des poids w_1 et w_2, avec l'attribution du poids le plus élevé w_1 à la classe RT. (Govindarajulu, 2011).
Ainsi, la classe RT est servie avec un taux $Sw_1\mu_1$ et la classe NRT avec un taux $Sw_2\mu_2$.

3.2.4 Équations de l'état d'équilibre

La résolution de la chaîne de Markov en trois dimensions exige d'établir l'ensemble des équations différentielles ci-après en égalisant le trafic entrant et sortant dans chaque état (i, j, k). Ensuite de calculer la probabilité $\pi(i, j, k)$ en utilisant la méthode de matrice géométrique afin de déterminer les diverses mesures de performance et de trouver les seuils définis dans les conditions mentionnées antérieurement.

- $\boldsymbol{i=0, j=0, k=0}$

  $(\lambda_0+\lambda_1+\lambda_2)\pi(0,0,0) = \mu_0\pi(1,0,0) + \mu_1\pi(0,1,0) + \mu_2\pi(0,0,1)$

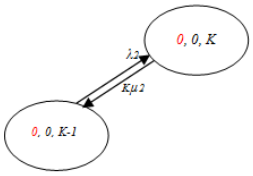


- ***i* = 0, *j* = 0, 0 < *k* < K**

$$(\lambda_0 + \lambda_1 + \lambda_2 + k\mu_2)\pi(0,0,k) = \mu_0\pi(1,0,k) + w_1\mu_1\pi(0,1,k) + \lambda_2\pi(0,0,k-1) + (k+1)\mu_2\pi(0,0,k+1)$$

- ***i* = 0, *j* = 0, *k* = K**

$$K\mu_2(0,0,K) = \lambda_2(0,0,K-1)$$



- ***i* = 0, 0 < *j* < K, *k* = 0**

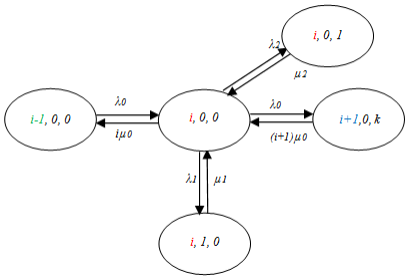
$$(\lambda_0 + \lambda_1 + \lambda_2 + j\mu_1)\pi(0,j,0) = \mu_0\pi(1,j,0) + (j+1)\mu_1\pi(0,j+1,0) + w_2\mu_2\pi(0,j,1) + \lambda_1\pi(0,j-1,0)$$

- ***i* = 0, *j* = K, *k* = *0***

$$K\mu_1(0,K,0) = \lambda_1(0,K-1,0)$$

- **0 < *i* < K, *j* = 0, *k* = 0**

$$(\lambda_0 + \lambda_1 + \lambda_2 + i\mu_0)\pi(i,0,0) = (i+1)\mu_0\pi(i+1,0,0) + \mu_1\pi(i,1,0) + \mu_2\pi(i,0,1) + \lambda_0\pi(i-1,0,0)$$



- ***i* = K, *j* = 0, *k* = *0***

  $K\mu_0(K,0,0) = \lambda_0(K-1,0,0)$

- ***i* = 0, 0 < *j* < K, 0 < *k* < K**

  $(\lambda_0+\lambda_1+\lambda_2+j\mu_1+k\mu_2)\pi(0,j,k) = \lambda_1\pi(0,j-1,0)+\lambda_2\pi(0,0,k-1)$

  $+\mu_0\pi(1,j,k) + (j+1)w_1\mu_1\pi(0,j+1,0)$

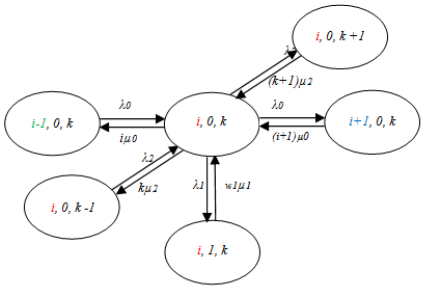
  $+(k+1)w_2\mu_2\pi(0,0,k+1)$

- **0 < *i* < K, *j* = 0, 0 < *k* < K**

  $(\lambda_0+\lambda_1+\lambda_2+i\mu_0+k\mu_2) = \lambda_0\pi(i-1,0,k)+\lambda_2\pi(i,0,k-1)$

  $+(i+1)\mu_0\pi(i+1,0,k) + w_1\mu_1\pi(i,1,k)$

  $+(k+1)\mu_2\pi(i,0,k+1)$

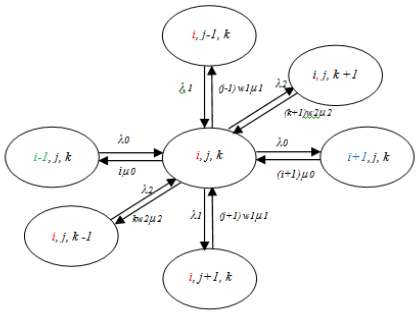


- **$0 < i < K$, $0 < j < K$, $k = 0$**

$$(\lambda_0 + \lambda_1 + \lambda_2 + i\mu_0 + j\mu_1) = \lambda_0\pi(i-1,j,0) + \lambda_1\pi(i,j-1,0) + w_2\mu_2\pi(i,j,1)$$
$$+ (i+1)\mu_0\pi(i+1,j,0) + (j+1)\mu_1\pi(i,j+1,0)$$

- **$0 < i < K$, $0 < j < K$, $0 < j < K$**

$$(\lambda_0 + \lambda_1 + \lambda_2 + i\mu_0 + jw_1\mu_1 + kw_2\mu_2) = \lambda_0\pi(i-1,j,k) + \lambda_1\pi(i,j-1,k)$$
$$+ \lambda_2\pi(i,j,k-1) + (i+1)\mu_0\pi(i+1,j,k)$$
$$+ (j+1)w_1\mu_1\pi(i,j+1,k)$$
$$+ (k+1)w_2\mu_2\pi(i,j,k+1)$$



Pour obtenir la solution du système, il est nécessaire de réécrire les équations de l'état d'équilibre sous la forme appropriée vu qu'il n'est pas simple de trouver la solution directement à partir de ces équations. Le procédé matrice géométrique est l'une des méthodes simplificatrices existantes, il sera adopté comme technique de résolution durant cette étude.

3.2.5 Procédé matrice géométrique

La méthode repose sur le principe de regrouper les états en des niveaux selon la valeur de *i*, et de permettre les transitions subséquentes:

Entre les états du même niveau; de (i,j,k) aux $(i,j+1,k)$, $(i,j-1,k)$, $(i,j,k+1)$ ou $(i,j,k-1)$.

D'un état à un autre appartenant à un niveau supérieur, de (i,j,k)au $(i+1,j,k)$.

Vers un état d'un niveau inférieur adjacent, de (i,j,k) au $(i-1,j,k)$.

Dans la figure 3.4, les états du même niveau sont représentés par une couleur identique de *i*. Comme la chaîne de Markov est à trois dimensions, un niveau *i* est à deux dimensions de *j* et de *k*. (Qi-Ming, 2014).

La figure 3.5 illustre le diagramme de transition des états (*j, k*) pour un niveau de *i* donné, les transitions suivantes sont permises:

Entre les états du même niveau, de (j,k) aux $(j,k+1)$ ou $(j,k-1)$.

D'un état à un autre appartenant à un niveau supérieur, de (j,k) au $(j+1,k)$.

Vers un état d'un niveau inférieur adjacent, de (j,k) au $(j-1,k)$.

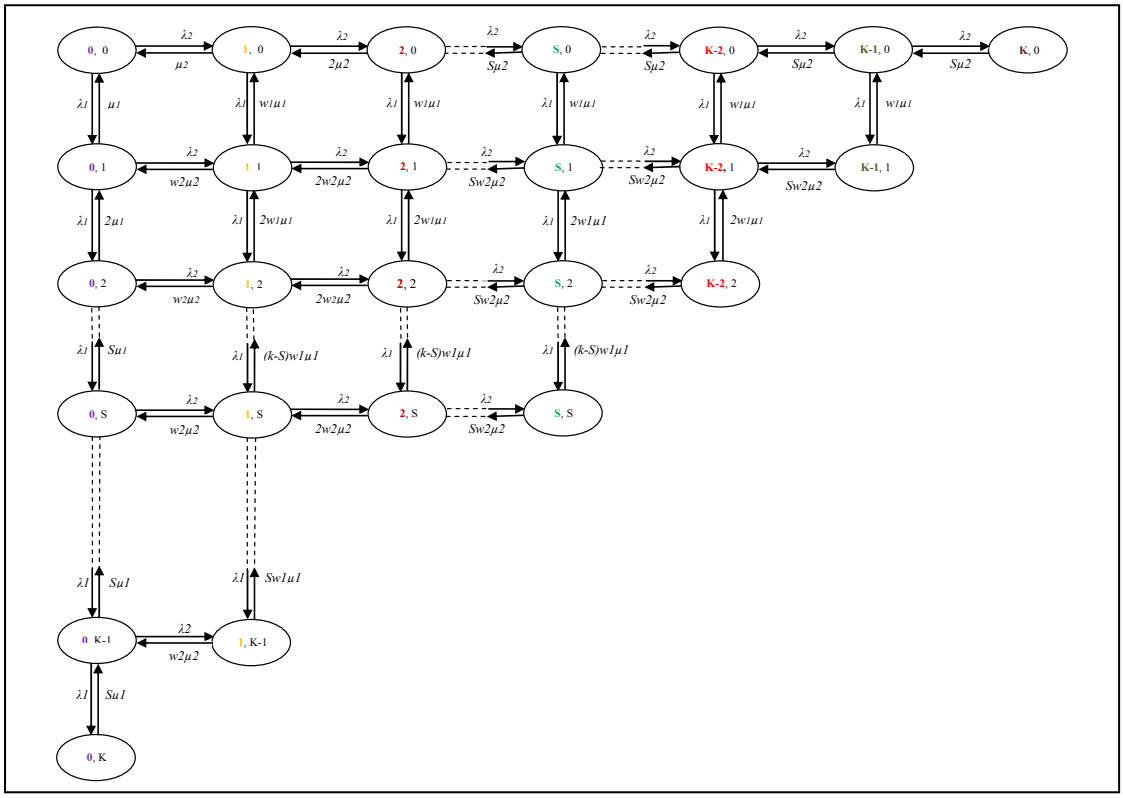


Figure 3.5 Diagramme de transition des états (*j*, *k*) par niveau *i*

Le générateur infinitésimal du processus est donné par:

$Q(i,j,k)(i+1,j,k) = \lambda_0$

$Q(i,j,k)(i-1,j,k) = \mu_0, i \geq 1$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\mu_0 + jw_1\mu_1 + kw_2\mu_2), i+j+k < K$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\mu_0 + j\mu_1), k = 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\mu_0 + k\mu_2), j = 0$

$Q(i,j,k)(i,j,k) = -(i\mu_0 + jw_1\mu_1 + kw_2\mu_2), i+j+k = K$

$Q(i,j,k)(i,j,k+1) = \lambda_2$

$Q(i,j,k)(i,j,k-1) = k\mu_2, j = 0$

$Q(i,j,k)(i,j,k-1) = kw_2\mu_2, j \neq 0$

$Q(i,j,k)(i,j+1,k) = \lambda_1$

$Q(i,j,k)(i,j-1,k) = j\mu_1, k = 0$

$Q(i, j, k)(i, j-1, k) = jw1\mu1, k \neq 0$

La matrice géométrique de la chaîne de Markov M/M/S/K avec PQ_CBWFQ est:

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & \overline{0} & \cdots & \cdots & \overline{0} \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots & \cdots & \overline{0} \\ \overline{0} & Q_{2,1} & Q_{2,2} & & Q_{2,3} & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ \vdots & & Q_{K-1,K-2} & Q_{K-1,K-1} & & Q_{K-1,K} \\ \overline{0} & \cdots & & \overline{0} & Q_{K,K-1} & Q_{K,K} \end{pmatrix}_{(j,\,k)} \quad (3.1)$$

Où $\overline{0}$ représente la matrice nulle. Les matrices $Q_{i,i-1}$, $Q_{i,i+1}$ sont des matrices rectangulaires, alors que les matrices $Q_{i,i}$ le long de la diagonale sont des matrices carrées non singulières ayant la même forme que la matrice Q.

Les niveaux au sein des matrices $Q_{i,i}$ sont des niveaux de *j* d'une dimension pour un *i* donné.

$$Q_{i,i}(j,k) = \begin{pmatrix} q_{0,0} & q_{0,1} & \overline{0} & \cdots & \overline{0} \\ q_{1,0} & q_{1,1} & q_{1,2} & \cdots & \overline{0} \\ \overline{0} & q_{2,1} & q_{2,2} & q_{2,3} & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & q_{K-1,K-2} & q_{K-1,K-1} & q_{K-1,K} & \\ \overline{0} & \cdots & \overline{0} & q_{K,K-1} & q_{K,K} \end{pmatrix}_{(k)} \quad (3.2)$$

Pour i = 0,…. K, les matrices $q_{j,j}$ ont les éléments de la diagonale négatifs en raison de la condition d'équilibre du système en chaque état (i, j, k). (Gautam, 2012).

Pour j = 0, quel que soit i:

$$q_{0,0} = \begin{pmatrix} -\sum(\lambda_p + i\mu_0) & \lambda_2 & \cdots & \cdots & 0 \\ \mu 2 & -(\sum \lambda_p + i\mu 0 + \mu 2) & \lambda_2 & \cdots & 0 \\ 0 & 2\mu 2 & -(\sum \lambda_p + i\mu 0 + 2\mu 2) & \lambda_2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \ddots & \lambda_2 \\ 0 & \cdots & k\mu 2 & -(i\mu 0 + k\mu 2) & \end{pmatrix} \quad (3.3)$$

Pour $j \neq 0$, quel que soit i:

$$q_{j,j} = \begin{pmatrix} -\sum(\lambda_p + i\mu_0 + j\mu_1) & \lambda_2 & 0\cdots & 0 \\ w_2\mu 2 & -(\sum \lambda_p + i\mu 0 + jw1\mu 1 + w2\mu 2) & \lambda_2 & \cdots & 0 \\ 0 & 2w_2\mu 2 & -(\sum \lambda_p + i\mu 0 + jw1\mu 1 + 2w2\mu 2) & \lambda_2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \lambda_2 \\ 0 & \cdots & kw_2\mu 2 & -(i\mu 0 + jw1\mu 1 + kw2\mu 2) & \end{pmatrix} \quad (3.4)$$

Lorsque $i = K$, $Q_{K,K}(j,k) = (-K\mu_0)$ car la taille des matrices $q_{j,j}$ change de $(K+1) \times (K+1)$ au $(1) \times (1)$ pour *j* allant de 0 à *K*.. Quant aux matrices $q_{j,j-1}$ et $q_{j,j+1}$ de chaque niveau de *i*, avec *j* =1,…K, les dimensions varient de $(K) \times (K+1)$ aux $(1) \times (2)$ et de $(K+1) \times (K)$ aux $(2) \times (1)$ successivement.

$$q_{j,j-1} = \begin{pmatrix} j\mu_1 & 0 & 0 & \dots & 0 \\ 0 & jw_1\mu_1 & 0 & \dots & 0 \\ 0 & 0 & jw_1\mu_1 & & \vdots \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & & jw_1\mu_1 \end{pmatrix} \tag{3.5}$$

De la même manière les matrices $Q_{i,i-1}$, et $Q_{i,i+1}$ ont la même configuration que les matrices $q_{j,j-1}$, et $q_{j,j+1}$ à l'exception que les transitions entre les niveaux dépendent uniquement des taux $i\mu_0$ et λ_0 .A noter que pour *i*= *K*, $Q_{K,K-1} = (K\mu_0)$

$$Q_{i,i+1} = \begin{pmatrix} \lambda_0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_0 & & \vdots \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & & \lambda_0 \end{pmatrix} \tag{3.6}$$

$$Q_{i,i-1} = \begin{pmatrix} i\mu_0 & 0 & 0 & \dots & 0 \\ 0 & i\mu_0 & 0 & \dots & 0 \\ 0 & 0 & i\mu_0 & & \vdots \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & & i\mu_0 \end{pmatrix} \tag{3.7}$$

3.2.6 Résolution des matrices

Les probabilités d'état stationnaire $\pi(i, j, k)$ peuvent être estimées à partir de $\pi Q = 0$ avec l'équation de normalisation $\sum \pi i = 1$ où π est un vecteur de probabilité partitionné en sous vecteurs π_i, $\pi = [\pi_1, \pi_2, \pi_3, \ldots, \pi_K]$.(Essia, Elhafsi et Molle, 2007)

La résolution de $\pi Q = 0$ conduit au système de formules ci-dessous :

$$\pi_0 Q_{0,0} + \pi_1 Q_{1,0} = 0 \quad (3.8)$$

$$\pi_0 Q_{0,1} + \pi_1 Q_{1,1} + \pi_2 Q_{2,1} = 0 \quad (3.9)$$

$$\pi_1 Q_{1,2} + \pi_2 Q_{2,2} + \pi_3 Q_{3,2} = 0 \quad (3.10)$$

$$\pi_{i-2} Q_{i-2,i-1} + \pi_{i-1} Q_{i-1,i-1} + \pi_i Q_{i,i-1} = 0, \text{ si } 2 < i <= K\text{-}1 \quad (3.11)$$

$$\pi_{K-1} Q_{K-1,K} + \pi_K Q_{K,K} = 0 \quad (3.12)$$

Les probabilités stationnaires sont:

$$\pi_1 = -\pi_0 Q_{0,0} \, {Q_{1,0}}^{-1} = \pi_0 R_1 \quad (3.13)$$

$$\pi_2 = -\pi_0 (Q_{0,1} + R_1 Q_{1,1}) {Q_{2,1}}^{-1} = \pi_0 R_2 \quad (3.14)$$

$$\pi_3 = -\pi_0 (R_1 Q_{1,2} + R_2 Q_{2,2}) {Q_{3,2}}^{-1} = \pi_0 R_3 \quad (3.15)$$

$$\pi_i = -\pi_0 (R_{i-2} Q_{i-2,i-1} + R_{i-1} Q_{i-1,i-1}) {Q_{i,i-1}}^{-1} = \pi_0 R_i \quad (3.16)$$

Les formules des probabilités stationnaires peuvent être généralisées de la forme suivante:

$$\pi_\xi = \pi_0 R_\xi, \text{ pour } \xi \geq 1 \quad (3.17)$$

Les matrices R_ξ sont calculées selon la procédure ci-dessous:

$$\begin{cases} R_0 = I & (3.18) \\ R_1 = -Q_{0,0}\, Q_{1,0}{}^{-1} & (3.19) \\ R_\xi = -\left(R_{\xi-2} Q_{\xi-2,\xi-1} + R_{\xi-1} Q_{\xi-1,\xi-1}\right) Q_{\xi,\xi-1}{}^{-1} & (3.20) \end{cases}$$

Où *I* représente la matrice identité

L'équation (3.12) peut être réécrite comme suit:

$$\pi_0 (R_{K-1} Q_{K-1,K} + R_K Q_{K,K}) = 0 \qquad (3.21)$$

Ainsi π_0 est obtenu en résolvant l'équation (3.21) avec l'équation de normalisation précédente exprimée en fonction de π_0

$$\pi_0 \sum_{d=0}^{K} R_d\, e_d = 1. \qquad (3.22)$$

Où e_d sont des vecteurs colonnes de 1.

3.2.7 Indicateurs de performances

Il s'agit de déterminer les expressions analytiques en fonction de π(*i, j, k*) de la taille moyenne des trois files d'attente Lq_p, du temps d'attente moyen dans chaque file Wq_p, de la probabilité de blocage globale dans le systèmeπ^*, du débit de chaque classe T_p avec
p =0, 1, 2 et des probabilités de perte de paquets par type de classe. (Wang, Min, Kouvatsos et Jin, 2009).

Le nombre moyen des paquets dans les trois queues est défini par:

$$Lq_0 = \sum_{i=0}^{K} \sum_{j=0}^{K-i} \sum_{k=0}^{K-i-j} i\pi(i,j,k) \quad (3.23)$$

$$Lq_1 = \sum_{j=0}^{K} \sum_{i=0}^{K-j} \sum_{k=0}^{K-i-j} j\pi(i,j,k) \quad (3.24)$$

$$Lq_2 = \sum_{k=0}^{K} \sum_{j=0}^{K-j} \sum_{i=0}^{K-k-j} k\pi(i,j,k) \quad (3.25)$$

Les expressions du temps d'attente moyen peuvent être directement dérivées de la formule de Little:

$$Wq_P = \frac{Lq_P}{\lambda_P (1 - \pi^*)} \quad (3.26)$$

π^*: Est la fraction des arrivées détournée par manque d'espace dans le buffer, π^*est également la probabilité de blocage globale dans le système.

$$\pi^* = \pi(i,j,k), \text{ avec } i+j+k = K \quad (3.27)$$

Le débit est défini comme le taux moyen des paquets passant par le système à l'état stationnaire, ou par le nombre moyen des paquets prévus d'être servis:

$$T_p = \lambda_p (1 - \pi^*) \text{ , avec } p = 0, 1, 2 \quad (3.28)$$

Afin d'éviter la saturation du système, les débits atteints par serveur doivent satisfaire la condition suivante:

$$T_p \leq \mu \quad (3.29)$$

Où $\mu = \mu_p$, $w_1\mu_1$ ou $w_2\mu_2$. et p =0, 1, 2.

En tenant compte de la stationnarité du système, la probabilité de perte de paquets par type de classe est définie par:

$$\pi_{LP} = \frac{\lambda_p}{\lambda_0 + \lambda_1 + \lambda_2}\ \pi^* \qquad \text{Avec } p = 0, 1, 2 \tag{3.30}$$

3.2.8 Estimation des seuils et des intervalles T_{r1}, T_{r2} et T_{r3}

Le $seuil_{RT}$ correspond au délai d'attente moyen d'un paquet RT dans la file Q_RT. De ce qui précède, le $seuil_{RT}$ est équivalent au Wq_2

$$seuil_{RT} = \frac{Lq_1}{\lambda_1\,(1-\pi^*)} \tag{3.31}$$

Le $seuil_{HD}$ est l'intervalle de rétention des paquets *handoff* dans la file Q_HD, il est équivalent au délai d'attente moyen Wq_1

$$seuil_{HD} = \frac{Lq_0}{\lambda_0\,(1-\pi^*)} \tag{3.32}$$

Le $seuil_{RT}$ est définie comme le délai d'attente moyen des paquets NRT dans leur file divisé par x, il est égal à Wq_3

$$seuil_{NRT} = \frac{Lq_2}{x\lambda_2\,(1-\pi^*)} \tag{3.33}$$

Par ailleurs, les intervalles T_{r1}, T_{r2} et T_{r3} , sont déterminés selon l'analyse de L. Kleinrock des files d'attente de multiples priorités. (Kleinrock, 1975), qui consiste à étiqueter tout paquet accédant au système. Ensuite, d'estimer le temps d'attente total dans chaque file qui dépendra du temps de service du paquet en cours de transmission, car le système est non préemptif, du temps de service des paquets de priorité égale ou supérieure présents dans les files d'attente et qui seront servis avant le paquet tagué, et finalement, du temps de service des paquets détenteurs de la plus haute priorité qui arrivent dans le système après le paquet tagué et qui seront traités en premier.

Le temps d'attente moyen Wp du paquet étiqueté de priorité PR_p peut s'écrire sous la forme. (Bergida et Shavitt, 2006):

$$Wp = Wres_p + \frac{1}{s\mu}(n_p + m_p) \ , p = 0, 1, 2 \qquad (3.34)$$

Où:

$Wres_p$ est la durée moyenne à partir de l'instant aléatoire d'arrivée d'un paquet jusqu'à l'achèvement du service courant.

$\frac{1}{s\mu}$: Temps de service moyen d'un paquet de la classe de priorité PR_p, $\mu = \mu_p$, $w_1\mu_1$ ou $w_2\mu_2$. et S représente le nombre de serveurs.

n_p: Nombre de paquets appartenant à la classe de priorité *p*, existant déjà dans la file et recevant un service avant le paquet étiqueté

m_p : Nombre de parquets de classe *p,* arrivant pendant l'attente du paquet tagué et transmis en premier

La même méthode de L. Kleinrock sera utilisée dans le calcul de T_{r1}, T_{r2} et T_{r3}. Le temps résiduel dépend du paquet en cours de transmission qui peut être soit *handoff*, RT ou NRT**,** le temps résiduel est donné par l'expression:

$$Wres_p = \frac{1}{S\mu}(1 - \pi(0,0,0)) \qquad (3.35)$$

Où:

$\pi(0,0,0)$ représente la probabilité de trouver le système IDLE, d'où $(1 - \pi(0,0,0))$ est la probabilité de trouver le système occupé.

Afin de faciliter l'application de l'algorithme et sans perte de généralités, le taux de service moyen μ considéré dans le calcul de $Wres_p$, est choisi égale à μ_0 car la possibilité de trouver un paquet *handoff* en cours de transmission est la plus élevée, vue que la classe *handoff* est servie selon la politique PQ avec la supposition qu'un paquet *handoff* est présent à tout instant *t* dans le système.

Par ailleurs, en plus du temps résiduel, le temps d'attente d'un paquet RT dépend également du temps de service des paquets *handoff* présents dans le système et des paquets *handoff* qui arrivent après le paquet RT et qui seront favorisés. Sachant que tous les paquets sollicitant un service doivent d'abord être temporisés dans les files, le nombre moyen des paquets *handoff* existant dans la file Q_HD peut être aisément tiré à partir du calcul de la taille de la file d'attente Q_HD, le nombre moyen des paquets *handoff* inclus *n0 et m0* simultanément, ce qui permet de conclure l'expression de T_{r1}:

$$T_{r1} = \frac{1}{\mu_0} \left(\frac{1-\pi(0,0,0)}{i} + \sum_{i=0}^{K} \sum_{j=0}^{K-i} \sum_{k=0}^{K-i-j} \pi(i,j,k) \right) \tag{3.36}$$

Pendant que les paquets RT sont servis, les paquets *handoff* sont retenus dans Q_HD durantT_{r2}. Donc T_{r2} correspond au temps d'attente supplémentaire des paquets *handoff* dans Q_HD, et au même temps, c'est le temps de service des paquets RT comme prioritaires. T_{r2} peut être calculée de la manière suivante:

$$T_{r2} = \frac{1}{\mu_1} \sum_{j=0}^{K} \sum_{i=0}^{K-j} \sum_{k=0}^{K-i-j} \pi(i,j,k) \tag{3.37}$$

Un paquet NRT doit attendre un temps résiduel, et le temps de livraison des paquets de plus hautes priorités *handoff* et RT, actuels et nouveaux; le nombre des paquets *handoff* et RT sont obtenus par le calcul des longueurs des files Q_{HD} et Q_{RT} respectivement. L'équation (3.38) permet d'estimer la période de l'intervalle d'attente T_{r3} des paquets NRT:

$$T_{r3} = \frac{1-\pi(0,0,0)}{i\,\mu_0} + \frac{1}{\mu_0} \sum_{i=0}^{K} \sum_{j=0}^{K-i} \sum_{k=0}^{K-i-j} \pi(i,j,k) + \frac{1}{\mu_1} \sum_{j=0}^{K} \sum_{i=0}^{K-j} \sum_{k=0}^{K-i-j} \pi(i,j,k) \tag{3.38}$$

3.3 Générateurs infinitésimaux de l'algorithme de courtoisie

L'estimation de la période T_{r1} permet de déterminer les valeurs de *i, j, k* correspondantes au début de transmission de la classe RT selon PQ. Durant la période T_{r3}, la classe RT est servie avec le taux μ_1, alors que les classes *handoff* et NRT partagent le débit du lien suivant les taux $w_0^{`}$ et $w_2^{`}$. Les nouvelles matrices générées après l'application de l'algorithme de

courtoisie dans ses deux conceptions ont les mêmes formes et dimensions que les matrices obtenues dans le schéma de base. Néanmoins, les taux de services changent. Le générateur infinitésimal du processus de Markov relatif à la première conception de l'algorithme de courtoisie est donné par:

$Q(i,j,k)(i+1,j,k) = \lambda_0$

$Q(i,j,k)(i-1,j,k) = i\mu_0, k = 0$

$Q(i,j,k)(i-1,j,k) = iw_0^{`}\mu_0, k \neq 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + iw_0^{`}\mu_0 + j\mu_1 + kw_2^{`}\mu_2), i+j+k < K$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\mu_0 + j\mu_1), k = 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + iw_0^{`}\mu_0 + kw_2^{`}\mu_2), j = 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + j\mu_1 + k\mu_2), i = 0$

$Q(i,j,k)(i,j,k) = -(iw_0^{`}\mu_0 + j\mu_1 + kw_2^{`}\mu_2), i+j+k = K$

$Q(i,j,k)(i,j,k+1) = \lambda_2$

$Q(i,j,k)(i,j,k-1) = k\mu_2, i = 0$

$Q(i,j,k)(i,j,k-1) = kw_2^{`}\mu_2, i \neq 0$

$Q(i,j,k)(i,j+1,k) = \lambda_1$

$Q(i,j,k)(i,j-1,k) = j\mu_1$

La classe NRT est servie avec le taux μ_2.pendant l'intervalle T_{r4}. Tandis que les classes *handoff* et RT partagent le débit du lien selon les taux $w_0^{``}$ et $w_1^{``}$. Le générateur infinitésimal du processus de Markov relatif à la deuxième conception de l'algorithme de courtoisie est le suivant:

$Q(i,j,k)(i+1,j,k) = \lambda_0$

$Q(i,j,k)(i-1,j,k) = i\mu_0, j = 0$

$Q(i,j,k)(i-1,j,k) = i\, w_0^{``}\mu_0, j \neq 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\, w_0^{``}\mu_0, + j\, w_1^{``}\mu_1 + k\,\mu_2), i+j+k < K$

$Q(i,j,k)(i,j,k) = -(i\, w_0^{``}\mu_0, + j\, w_1^{``}\mu_1 + k\,\mu_2), i+j+k = K$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\,w_0^{``}\mu_0, + j\,w_1^{``}\mu_1), k = 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + i\mu_0 + k\mu_2), j = 0$

$Q(i,j,k)(i,j,k) = -(\sum \lambda_p + j\mu_1 + k\mu_2), i = 0$

$Q(i,j,k)(i,j,k+1) = \lambda_2$

$Q(i,j,k)(i,j,k-1) = k\mu_2$

$Q(i,j,k)(i,j+1,k) = \lambda_1$

$Q(i,j,k)(i,j-1,k) = j\mu_1, i = 0$

$Q(i,j,k)(i,j-1,k) = j\,w_1^{``}\mu_1, i \neq 0$

3.4 Stationnarité du système M/M/S/K avec PQ_CBWFQ

Les différents seuils et périodes de transmission, ont été déterminés en se basant sur le schéma de la figure-A I-1. Néanmoins, il est important de recalculer la probabilité $\pi(i,j,k)$ dans chaque état (i,j,k) après avoir appliqué l'algorithme et avant de découler les nouvelles valeurs des mesures de performances, ce calcul est indispensable afin d'assurer l'équilibre du système qui exige de satisfaire la condition exprimée dans l'équation de normalisation:

$$\sum \pi i = 1 \tag{3.39}$$

3.5 Structure générale de l'algorithme de courtoisie

La structure générale de l'algorithme de courtoisie est donnée dans ce qui suit, la distinction entre les différentes priorités est illustrée à travers les poids affectés à chaque classe, le poids donné à la classe *handoff* dans le scénario de base est égal à 1**.**

1. Début du scénario de base.

1.1. Calcul et attribution des taux $w_1$et w_2.

$w_1 + w_2$=1.

$w_1 > w_2$

$w_{rt} = w_1$

$w_{nrt} = w_2$

1.2. Déterminer matrice génératrice Q

1.3. Calcul_Probabilités

1.4. Vérification de la stationnarité du système

Si somme Calcul_Probabilités ≤ 1

Système stationnaire

Sinon

Changer les valeurs de w_1 et w_2

Refaire 1.3 et 1.4

Fsi

Fin de scénario de base

2. Début de la première conception

2.1. Calcul_Seuils ($seuil_{HD}$,$seuil_{RT}$, $seuil_{NRT}$)

2.2. Calcul_Periodes (T_{r1}, T_{r2})

2.3. Tant que $T_{r1} > seuil_{RT}$ et $seuil_{RT} \leq T_{r2} \leq seuil_{HD} + seuil_{RT}$

2.4. Calcul et attribution des taux $w_0^{`}$ et $w_2^{`}$.

$w_1 = 1$

$w_0^{`} + w_2^{`} = 1$

$w_0^{`} > w_2^{`}$

$w_{hd} = w_0^{`}$

$w_{nrt} = w_2^{`}$

2.5. Refaire 1.2 et 1.3.

2.6. Vérification de la stationnarité du système

Si somme Calcul_Probabilites ≤1

Système stationnaire

Sinon

Changer valeurs de $w_0^{`}$ et $w_2^{`}$

Refaire 1.3 et 2.6

Fsi

Fin Tant que

Fin de la première conception

3. Début de la deuxième conception

3.1. Calcul_Période T_{r3} ,

3.2. Tant que $T_{r3} > seuil_{NRT}$ et $seuil_{RT} \leq T_{r2} \leq seuil_{HD} + seuil_{RT}$

3.3. Calcul et attribution des taux $w_0^{``}$ et $w_1^{``}$

$w_2 = 1$

$w_0^{``} + w_1^{``} = 1$

$w_0^{``} > w_1^{``}$

$w_{hd} = w_0^{``}$

$w_{rt} = w_1^{``}$

3.4. Refaire les étapes de la première conception

3.5. Refaire 1.2 et 1.3.

3.6. Vérification de la stationnarité du système

Si somme Calcul_Probabilites ≤1

Système stationnaire

Sinon

Changer les valeurs de $w_0^{``}$ et $w_1^{``}$

Refaire 1.3et 3.6

Fsi

Fin tant que.

Fin de la deuxième conception

3.6 Conclusion

La solution développée s'intéresse à la gestion des priorités pour servir équitablement les utilisateurs et augmenter le nombre de clients acceptés et servis. La réalisation du procédé est fondée sur un système Markovien M/M/S/K comprenant *S* serveurs équivalents et trois files d'attentes relatives aux trafics *handoff*, RT et NRT, la gestion des files d'attente est effectuée suivant les politiques PQ pour traiter les paquets *handoff* et CBWFQ pour traiter les paquets RT et NRT. L'état du système M/M/S/K avec PQ_CBWFQ à un instant t est modélisé par une chaîne de Markov à trois dimensions qui permet de dériver les différentes équations d'équilibres, la résolution de ces équations est effectuée à l'aide du procédé de matrice géométrique dans le but de découler les formules relatives au temps d'attente moyen et de déterminer les seuils nécessaires pour la conception de l'algorithme.

Ensuite, l'algorithme de courtoisie est appliqué afin d'attribuer consécutivement la plus haute priorité aux deux autres classes de trafics. Quand le temps d'attente moyen d'un paquet RT tagué est atteint, la classe RT devient prioritaire et elle est servie suivant PQ, alors que les classes *handoff* et NRT sont servis suivant CBWFQ. Pareillement, lorsque le temps d'attente d'un paquet NRT étiqueté t équivalent au temps d'attente moyen dans la file, la classe NRT est servie selon PQ. Tandis que les classes *handoff* et RT sont servis suivant CBWFQ. L'intervalle de service des classes RT et NRT selon PQ simultanément ne doit pas dépasser le temps d'attente moyen des paquets *handoff* afin d'éviter de dégrader la QoS de cette dernière.

Afin d'évaluer la robustesse de la solution, différents scénarios de simulations ont été réalisés dont les résultats sont présentés dans le prochain chapitre.

CHAPITRE 4

SIMULATIONS ET RÉSULTATS

4.1 Introduction

L'algorithme de courtoisie a été mis en œuvre dans le but de mesurer l'impact des différentes conceptions sur les indicateurs de performances et de vérifier si les nouvelles approches permettent d'améliorer l'ordonnancement dans la liaison montante sans dégrader la qualité de service de la classe de haute priorité. De ce fait, trois scénarios ont été réalisés basés sur les arrivées de Poisson et le temps de service exponentiel. Le premier scénario représente le scénario de base où la classe *handoff* détient en continu la plus haute priorité, le deuxième et le troisième scénario illustrent le cas de transmission des nouveaux paquets RT et NRT selon PQ dans des portions de temps bien déterminées.

4.2 Modèle et paramètres de simulation

Le modèle de simulation est constitué d'une cellule de large bande égale à 80Mhz établie suite à l'agrégation de 4 composantes porteuses LTE (version 8) de 20Mhz utilisant chacune la technique de modulation SC-FDMA en liaison montante. Chaque composante admet l'allocation de 90 PRB simultanément durant un intervalle de temps (TTI) de 1 ms. Les utilisateurs *handoff* sont illustrés par 30 clients *handoff* VoLTE et 15 clients *handoff* FTP.

Le nombre d'éléments de ressources dans un PRB est 144 REs car il est considéré qu'un PRB est constitué de 12 sous porteuses de 12 symboles durant 1 TTI, 120 REs sont dédiés au transfert de données. Tandis que, 24 REs (2 symboles) sont réservés au signal de référence qui est utilisé dans l'estimation du canal afin de réduire le taux d'erreur de transmission. (Don, 2013). En outre, il est également considéré que la qualité du canal de transmission est moyenne et que la valeur de l'indicateur de la qualité du canal rapportée par l'équipement utilisateur au nœud B est égal à 7 (CQI**7**). Ainsi, le taux de codage effectif relatif à cette valeur est égal à 0.369 (378/1024).

D’autre part, CQI7 utilise le schéma de modulation 16 QPSK qui permet à un élément de ressource de transporter 4 bits codés, d’où le nombre de bits de données à transmettre par éléments de ressources est 4×0.369, soit 1.47 bits. Par conséquent, 120×1.47, soit 177 bits par PRB.

Les stations VoLTE utilisent le codec AMR-WB12.65 et génèrent des paquets de taille 300 bits chaque 20ms. [23]. Ainsi, un paquet VoLTE est transmis sur 2 PRB, ce qui permet d’envoyer 15 paquets/ms, d’où λ_{01} = 15000 paquets/s.

Les paquets *handoff* FTP sont de taille 512 octets. De ce fait, le nombre de PRB nécessaire pour envoyer un paquet FTP est 512×8/177, soit 24 PRB. Si chaque utilisateur *handoff* FTP est alloué un seul PRB pendant un TTI, le nombre de PRB dédié au trafic *handoff* FTP pendant 1 seconde est 15000 PRB. Donc λ_{02} = 625 paquets/s (15000/24). Comme le processus de Markov est commutatif, λ_0 = 15625 paquets/s.

Quant aux nouveaux utilisateurs, ils sont constitués de 100 clients VoLTE et 80 clients FTP durant un TTI, les stations VoLTE produisent λ_1 = 50000 paquets/s et les stations FTP émettent des paquets de taille 512 octets par seconde avec la possibilité d’allouer un PRB par utilisateur. De la même manière que précédemment, il en résulte λ_2 = 3333.4 paquets/s.

L’agrégation de 4 composantes porteuses offre un débit de 200 Mbits/s car il est supposé que chaque composante a un débit de 50 Mbits/s. Les serveurs cités précédemment représentent un groupement de porteurs où chaque porteur est réservé à un seul utilisateur en fonction des paramètres de la QoS sollicités. Le nombre de serveurs par composante porteuse est égal à 3 ce qui donne un total de 12 serveurs, le débit agrégé est subdivisé sur ces 12 serveurs. Par conséquent, le taux de service approximatif par serveur est égal à 17 Mbits/s. Les paquets émis par les utilisateurs sont retenus dans le buffer K de capacité maximale de 12 paquets pour qu’ils soient classés avant d’être servis.

Le tableau 4.1 résume les paramètres de simulation communs entre les trois scénarios.

Tableau 4.1 Paramètres globaux de simulation

Paramètre	**Valeur**
Configuration d'agrégation	4 CCs de 20 MHz, (180 kHz par PRB)
Nombre de macro cellule	1
Nombre d'antennes	1×1
Nombre de PRB en liaison montante	(90 PRBs, 180 kHz par PRB)
Durée PRB	1 TTI
Nombre de sous porteuse par ressources block	12
Nombre de symboles par sous porteuse	7
Inter-arrivées *handoff* (λ_0)	15626 paquets/s
Inter-arrivées RT (λ_1)	50000 paquets/s
inter-arrivées NRT (λ_2)	3333.4 paquets/s
Taille du system k	12 paquets
Taille paquet_VoLTE	300 bits
Taille paquet_FTP	512 bytes
Index CQI	7
Modulation	16 QAM
Taux de codage	378/1024
Durée de simulation	10000 TTI

4.3 Évaluation des performances

Cette section analyse les résultats des indicateurs de performances relatifs au scénario de base et aux deux conceptions de l'algorithme de courtoisie décrites dans le chapitre précédent. L'évaluation a été développée en utilisant le langage MATLAB version R2014a installée sur une hôte munie d'un CPU Intel Celeron de 1.9 GHZ et d'une mémoire RAM de 4 Gb.

Afin d'analyser le comportement du système avant et pendant les situations de congestion, les taux de services sont maintenus inchangés durant la réalisation des tests. Par contre, les valeurs des inters arrivées se font augmenter d'une unité λ_P. Autrement dit, une unité de temps de simulation correspond à une unité λ_P. En outre, il est important de mentionner que la congestion est observée à partir de t = 2 s.

4.3.1 Scénario 1: La classe *Handoff* servie selon PQ

Ce cas représente le schéma de base où La classe *handoff* est continuellement prioritaire, la bande passante est entièrement dédiée à la transmission des paquets *handoff*, alors qu'elle est partagée entre les classes RT et NRT en fonction des taux $w_1$et w_2. Les nouveaux paquets qui arrivent dans le réseau sont emmagasinés dans le buffer tant qu'il n'a pas atteint sa capacité maximale K, autrement, ils sont bloqués.

4.3.1.1 Calcul des taux de service

L'estimation du taux μ_0 est fonction de μ_1 et $\mu_2$car les arrivées λ_0 sont hétérogènes:

$$\mu_0 = w_{HD1}\mu_1 + w_{HD2}\mu_2 \qquad (4.1)$$

Les taux w_{HD1} et w_{HD2} déterminent la portion de la bande passante destinée au service des flux *handoff*, RT et NRT, les valeurs de w_{HD1} et w_{HD2} dépendent en premier lieu de la stabilité du système et en second lieu du nombre d'utilisateurs VoLTE et FTP. Le poids le plus élevé est toujours attribué aux utilisateurs VoLTE. En outre, w_{HD1} et w_{HD2} doivent aussi satisfaire la condition de la politique CBWFQ du moment où les deux classes partagent le même serveur:

$$w_{HD1} + w_{HD2} = 1 \tag{4.2}$$

Le nombre d'utilisateurs VoLTE constitue 2/3 du nombre total des clients *handoff*, alors que les utilisateurs FTP constituent 1/3. Par conséquent:

$$w_{HD1} \simeq 0.7 \quad \text{et} \quad w_{HD2} \simeq 0.3$$

Les taux de services μ_1 et μ_2 sont calculés come suit:

$$\mu_1 = \frac{Taux\ de\ service\ \left(\frac{bit}{s}\right)}{La\ taille\ du\ paquet\ RT\ \ (bit)} \tag{4.3}$$

Donc:

$$\mu_1 = 56667 \text{ Paquets/s}$$

Et:

$$\mu_2 = \frac{Taux\ de\ service\ \left(\frac{bit}{s}\right)}{La\ taille\ du\ paquet\ NRT\ \ (bit)} \tag{4.4}$$

Ce qui implique:

$$\mu_2 = 40912 \text{ paquets/s.}$$

De (4.3) et (4.4):

$$\mu_0 = 4150.4 \text{ Paquets/s}$$

Les nouveaux utilisateurs partagent le même serveur et sont transmis suivant la technique CBWFQ, le calcul des taux $w_1$et $w_2$des utilisateurs VoLTE et FTP suit la même méthode utilisée antérieurement, le nombre d'utilisateurs VoLTE constitue 3/5 du nombre total des nouveaux utilisateurs. Tandis que les utilisateurs FTP constituent 2/5. Ainsi $w_1 \tilde{} 0.6$ et $w_2 \simeq 0.4$

Avec les valeurs de w_{HD1} et w_1, le débit minimum garanti (GBR) aux trafics VoLTE *handoff* et nouveau est calculé comme suit:

$$GBR_{HD\_RT} = w_{HD1}\mu_1 \quad (4.5)$$

D' où:

$$GBR_{HD\_RT} = 39666.9 \text{ paquets/s}$$

Et:

$$GBR_{RT} = w_1\mu_1 \quad (4.6)$$

$$GBR_{RT} = 34000.2 \text{ paquets/s}$$

4.3.1.2 Résultats numériques

Tableau 4.2 Résultats numériques du scénario 1 à t = 3 s

Paramètre	***Q_HD***	***Q_RT***	***Q_NRT***
Nombre de paquets dans la file (paquets)	0.93647	4.1175	3.1237
Délai moyen dans la file ×1.0e-04 (s)	0.22620	0.31080	3.53682
Débit ×1.0e+05 (paquets/s)	0.41399	1.32477	0.08832

Le tableau 4.2 donne un aperçu sur les valeurs de quelques indicateurs de performance calculées à t = 3 s. Les résultats indiquent que la classe *handoff* servie selon PQ a le nombre moyen de paquets et le délai les plus faibles.

4.3.1.3 Résultats graphiques

Étude de la longueur des files d'attente

La figure.4.1 représente la variation de la longueur des trois files d'attente en fonction du temps, elle montre que la classe *handoff* possède continuellement la longueur Lq_0 la plus étroite, car elle détient la plus haute priorité. Par contre, la longueur de la file NRT Lq_2 est supérieure à la longueur de la file RT Lq_1 durant l'intervalle approximatif de 1 s à 2.3 s car les paquets VoLTE sont servis avec un taux plus grand que le taux de service des paquets FTP. Ensuite, Lq_1 devient plus élevée que Lq_0 et Lq_2 pendant tout le temps de simulation restant, car le nombre d'utilisateurs VoLTE acceptés dans le système est très grand.

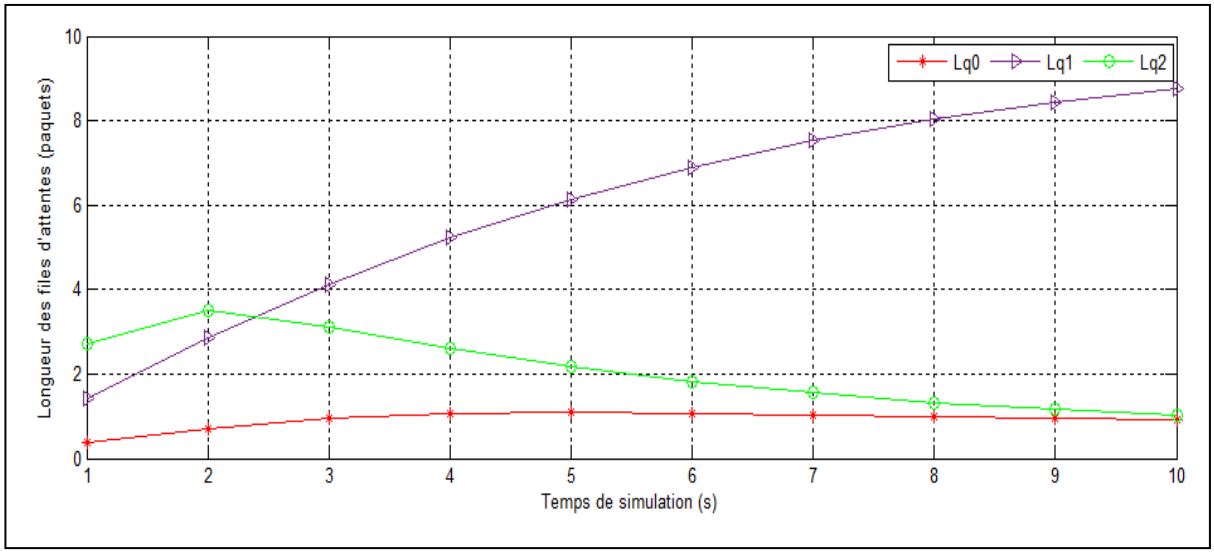


Figure 4.1 Longueur des files d'attentes VS le temps de simulation

Étude du débit

Le débit représente le taux du trafic accepté et servi par le système, il s' accroît au fur et à mesure que le nombre d'inters arrivées augmente. Selon la figure 4.3, la valeur du débit la plus élevée a été constatée pour le trafic RT au moment où le nombre d'inter-arrivées VoLTE est le plus grand, suivie de T0, débit du trafic *handoff* et enfin T3 débit de la classe NRT.

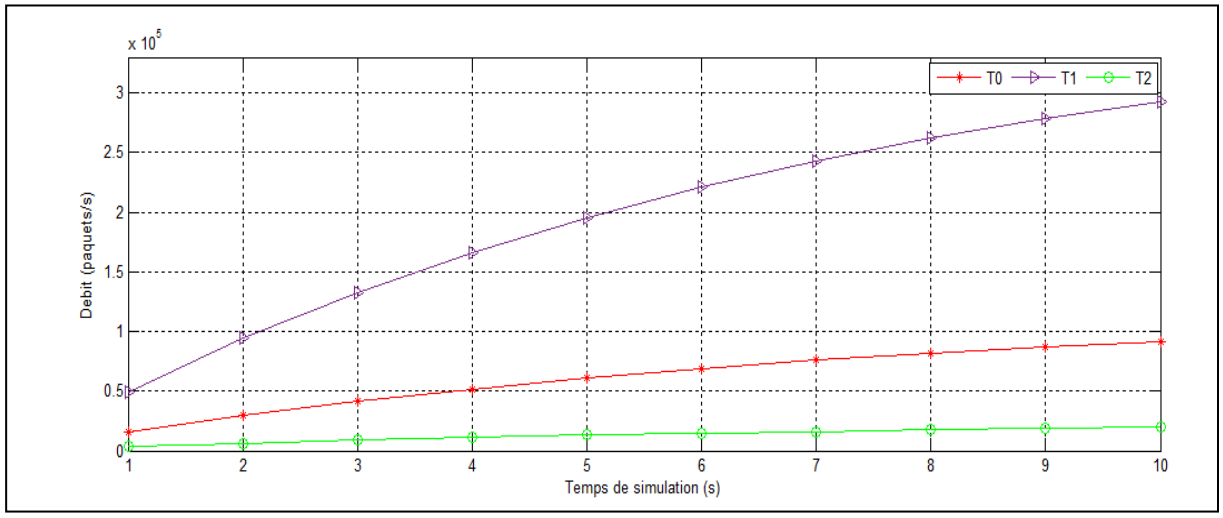


Figure 4.2 Débits Tp VS le temps de simulation

Étude du délai moyen dans les files d'attente

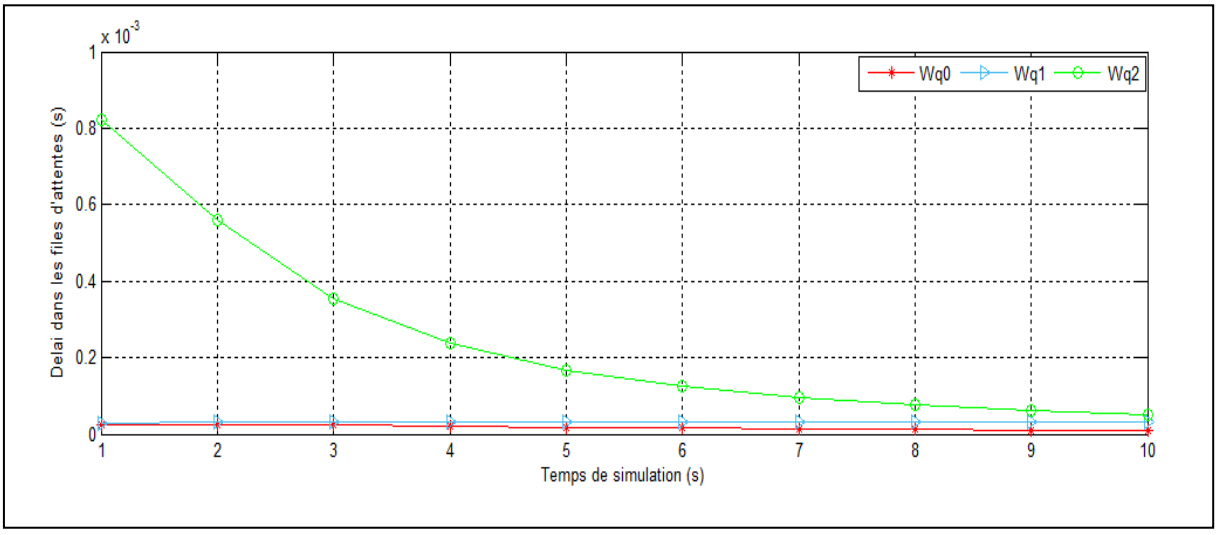


Figure 4.3 Délai dans les files d'attentes VS le temps de simulation

La figure 4.2 illustre la variation du délai moyen dans les trois files d'attente, le délai le plus court est observé dans la file *handoff* (Wq_0). Le délai Wq_1 dans la file RT est inférieur au délai Wq_2 dans la file NRT malgré Lq_1 est supérieure à lq_2 pour t entre 1 s et 2.3 s car le nombre d'utilisateurs VoLTE acceptés dans le système est important (voir figure 4.3). De plus, ils sont servis avec un taux plus grand que le taux de service des utilisateurs FTP. Le

délai dans la file NRT atteint sa valeur la plus élevée dans la première seconde (Wq_2 = 0.00082 s) puis commence à diminuer dans la dixième seconde jusqu'au Wq_2 = 0.00005 s car le nombre de paquets FTP acceptés augmente constamment, ce qui a permis de réduire le délai.

Étude de la probabilité de blocage

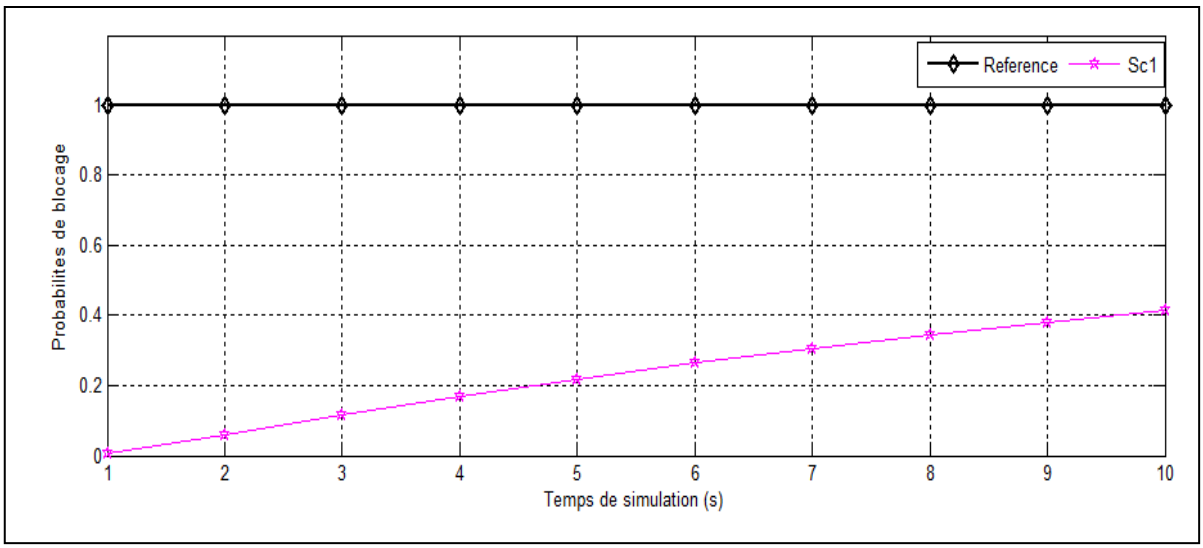


Figure 4.4 Probabilité de blocage VS le temps de simulation

La probabilité du blocage π^* détermine la portion du trafic rejetée par le système quand il excède la capacité de traitement des serveurs. Plus le nombre d'inters arrivées est grand, plus la probabilité du blocage est supérieure, pour t = 10 s, π^*= 0.41499.

Étude de la perte de paquets

La perte de paquets dépend essentiellement de la probabilité de blocage dans le système et du taux d'inters arrivées de chaque classe de trafic. Dans la figure.4.5, HD, RT et NRT désignent π_{L0}, π_{L1} et π_{L2} respectivement. La classe RT expérimente le taux de perte le plus haut. Ensuite la classe *handoff* et enfin, la classe NRT qui a le taux le plus faible.

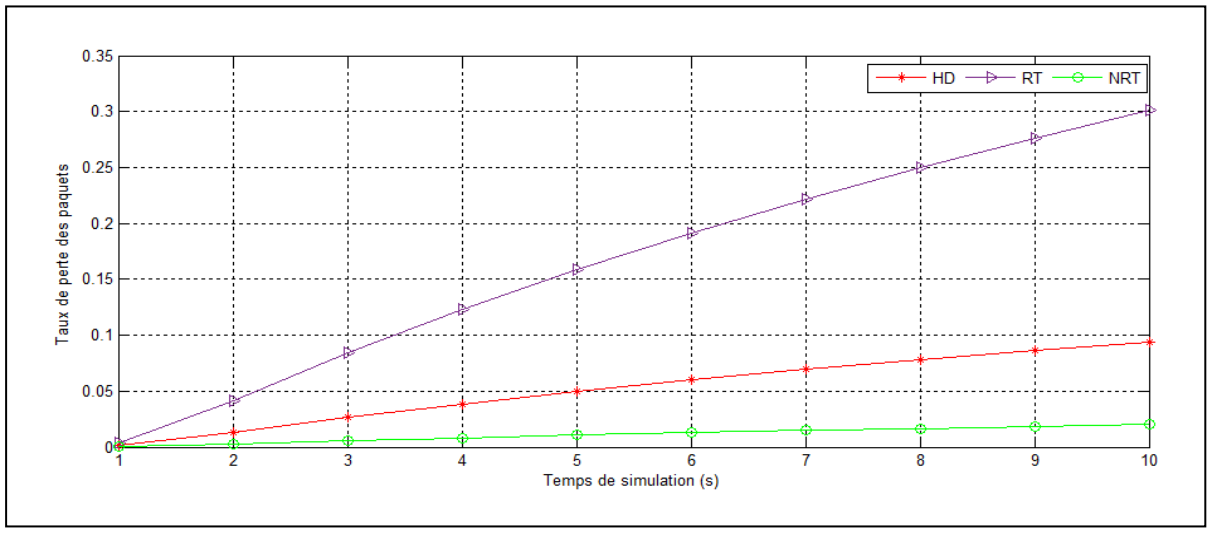


Figure 4.5 Taux de perte des paquets VS le temps de simulation

4.3.2 Scénario 2: La classe RT servie selon PQ

La classe RT est prioritaire durant la période équivalente au temps d'attente moyen des paquets *handoff*, les paquets RT bénéficient du débit total de la bande passante, alors que la politique d'ordonnancement CBWFQ est appliquée dans la transmission des paquets *handoff* et NRT, et le débit est subdivisé sur les deux classes suivant les taux $w_0^{`}$ et $w_2^{`}$. Les valeurs de $w_0^{`}$ et $w_2^{`}$ sont fixées à 0.9 et à 0.1 afin d'assurer la stabilité du système.

4.3.2.1 Résultats numériques

Le tableau.4.3 affiche les statistiques recueillies de certaines grandeurs de performance à t = 3 s, en les comparants avec les valeurs du tableau.4.2, le nombre moyen de paquets *handoff* et le délai correspondant ont augmenté. En revanche, une réduction de la valeur de ces deux paramètres est perçue dans les files RT et NRT. En outre, le scenario 2 offre les meilleurs résultats en termes de débit et de perte de paquets.

Tableau 4.3 Résultats numériques du scénario 2 à t =3 s

Paramètre	**Q_HD**	**Q_RT**	**Q_NRT**
Nombre de paquets dans la file (paquets)	1.00820	3.90831	2.47210
Délai moyen dans la queue ×1.0e-04 (s)	0.23465	0.28427	2.69702
Débit ×1.0e+05 (paquets/s)	0.42965	1.37487	0.0916
Taux des paquets perdus	0.01890	0.06048	0.00403

Le tableau 4.4 indique les valeurs des seuils des états de transitions correspondants à T_{r2}, la période de transmission des paquets RT selon PQ. Ainsi, (i_{RT}, j_{RT}, k_{RT}) et (i_{HD}, j_{HD}, k_{HD}) désignent les composantes (i, j, k) des états de transition relatifs au début et à la fin de T_{r2}, sachant que (i_{RT}, j_{RT}, k_{RT}) sont obtenus lorsque t = $seuil_{RT}$ et (i_{HD}, j_{HD}, k_{HD}) sont calculés à t = $seuil_{RT}$ + $seuil_{HD}$. Les résultats du tableau 4.4 montrent que (i_{RT}, j_{RT}, k_{RT}) demeurent inchangés à tout instant t, alors que (i_{HD}, j_{HD}, k_{HD}) changent à partir de t = 6 s et offrent un intervalle de transmission plus restreint au paquets RT afin d'éviter un taux de blocage important dans le système.

Tableau 4.4 Valeurs des seuils relatifs à la période de transmission des paquets RT selon PQ

Paramètre	**Résultat**									
Temps de simulation (s)	1	2	3	4	5	6	7	8	9	10
i_{RT}	1	1	1	1	1	1	1	1	1	1
j_{RT}	1	1	1	1	1	1	1	1	1	1
i_{HD}	11	11	11	11	11	1	1	1	2	1
j_{HD}	11	11	11	11	11	5	7	7	8	7
k_{HD}	11	11	11	11	11	4	2	1	0	0

Le tableau 4.5. Présente les valeurs des grandeurs suivantes: les taux des débits, la probabilité de blocage et la perte de paquets de chaque classe.

Tableau 4.5 Valeurs des débits, de la probabilité de blocage et de la perte de paquets du scénario 2

Paramètre	Résultat									
Temps de simulation (s)	1	2	3	4	5	6	7	8	9	10
T0-SC2×1.0e+05 (Paquets/s)	0.15562	0.30020	0.42965	0.54743	0.65404	0.74951	0.83115	0.9004	0.95833	1.00644
T1-SC2 ×1.0e+05 (Paquets/s)	0.49796	0.96064	1.37487	1.75176	2.09294	2.39845	2.65968	2.88130	3.0666	3.22062
T2-SC2 ×1.0e+04 (Paquets/s)	0.33198	0.64044	0.91660	1.16787	1.39532	1.59900	1.77316	1.92091	2.04490	2.14712
Probabilite de bocage	0.0041	0.03935	0.08342	0.12412	0.16282	0.20051	0.24009	0.27967	0.31852	0.35587
Perte de paquets HD	0.00092	0.00891	0.01890	0.02812	0.03689	0.04543	0.05440	0.06337	0.07217	0.08064
Perte de paquets RT	0.00296	0.02853	0.06048	0.08999	0.11805	0.14538	0.17408	0.20278	0.23095	0.25804
Perte de paquets NRT	0.00019	0.00190	0.00403	0.00599	0.00787	0.00969	0.01160	0.01352	0.01539	0.01720

4.3.2.2 Résultats graphiques

Étude de la longueur des files d'attente

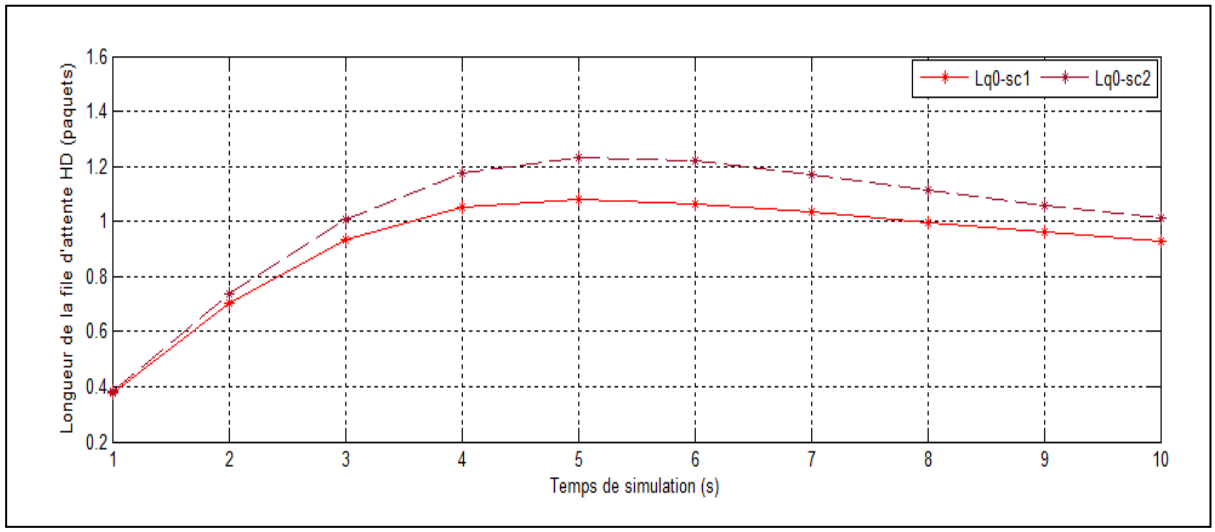


Figure 4.6 Longueurs de la file d'attente *handoff* VS le temps de simulation

Le nombre moyen de paquets dans les trois files d'attente a été inventorié, la figure.4.6 montre que la longueur de la file *handoff* Lq_0-sc2 dans le deuxième scenario a augmenté car le taux de service de la classe *handoff* a baissé de 0.1 afin de permettre la transmission des paquets NRT au sein des porteurs du même serveur. La valeur maximale de Lq_0-sc2 est

1.23341 paquet obtenu pour t = 5 s, subséquemment et comme appuyé par la figure 4.7, la différence du nombre moyen de paquets RT entre les deux scénarios a la valeur optimale de 0.3175 paquet à t = 5 s car plus l'intervalle de transmission des paquets RT suivant PQ est large plus Lq_0-sc2 est importante, d'où Lq_1-sc2 est moins significative.

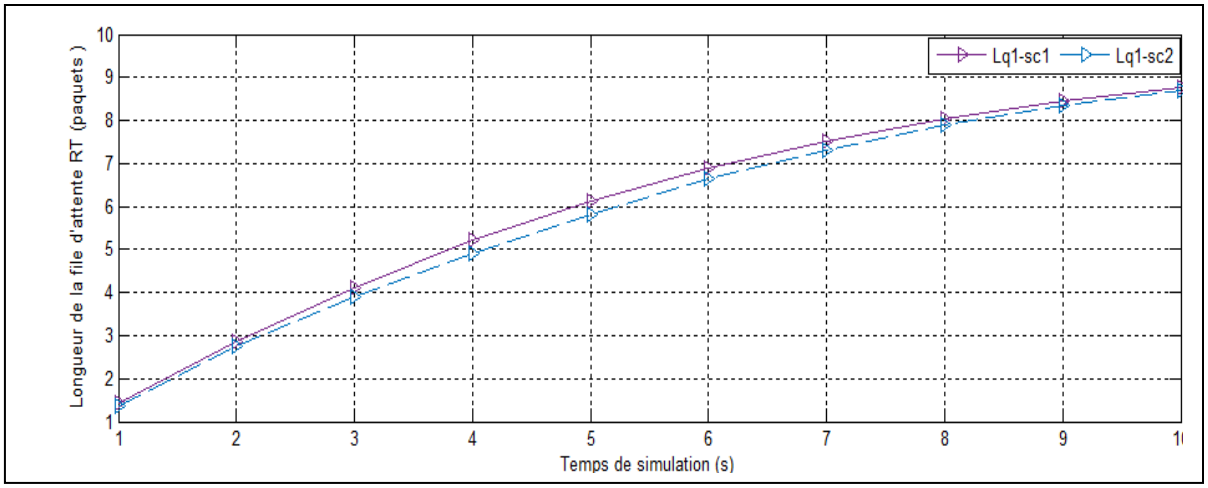


Figure 4.7 Longueurs de la file d'attente RT VS le temps de simulation

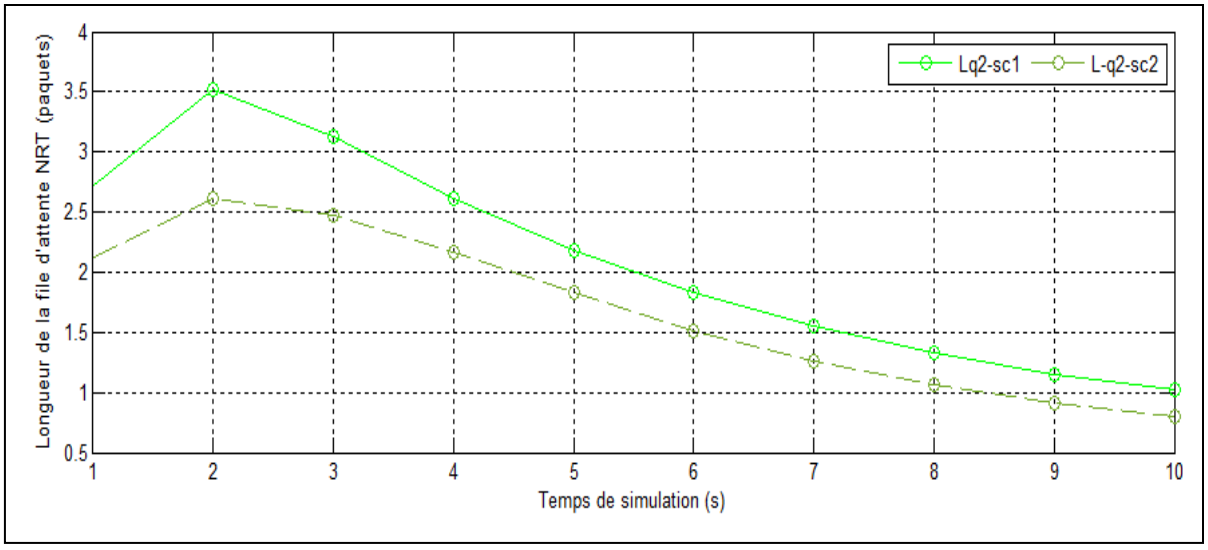


Figure 4.8 Longueurs de la file d'attente NRT VS le temps de simulation

La figure 4.8, représente la longueur de la file d'attente NRT Lq_2 dans les deux scenarios. Bien que les paquets NRT détiennent la plus basse priorité. Cependant, Lq_2-sc2 a également diminué car dans le scénario de base, les paquets FTP et VoLTE sont servis selon la

technique CBWFQ. Or, la classe RT est prioritaire dans la période équivalente au temps d'attente moyen des paquets *handoff*, ce qui permet de transmettre un plus grand nombre de paquets VoLTE durant cet intervalle. Ainsi, les paquets FTP n'attendent aussi longtemps que dans le premier scenario et sont servis plus rapidement.

Étude du débit

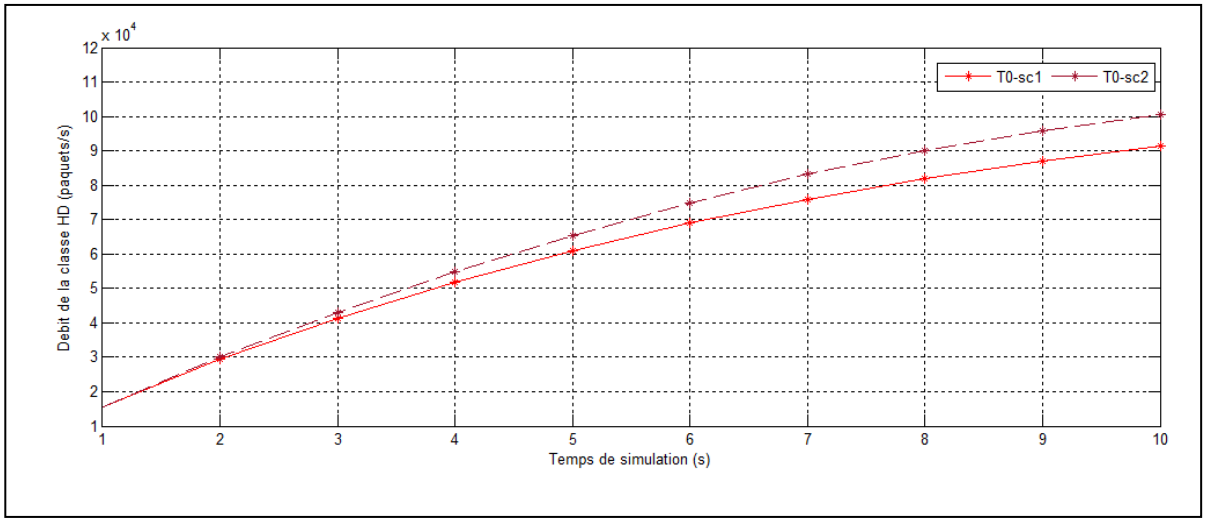


Figure 4.9 Débits *handoff* VS le temps de simulation

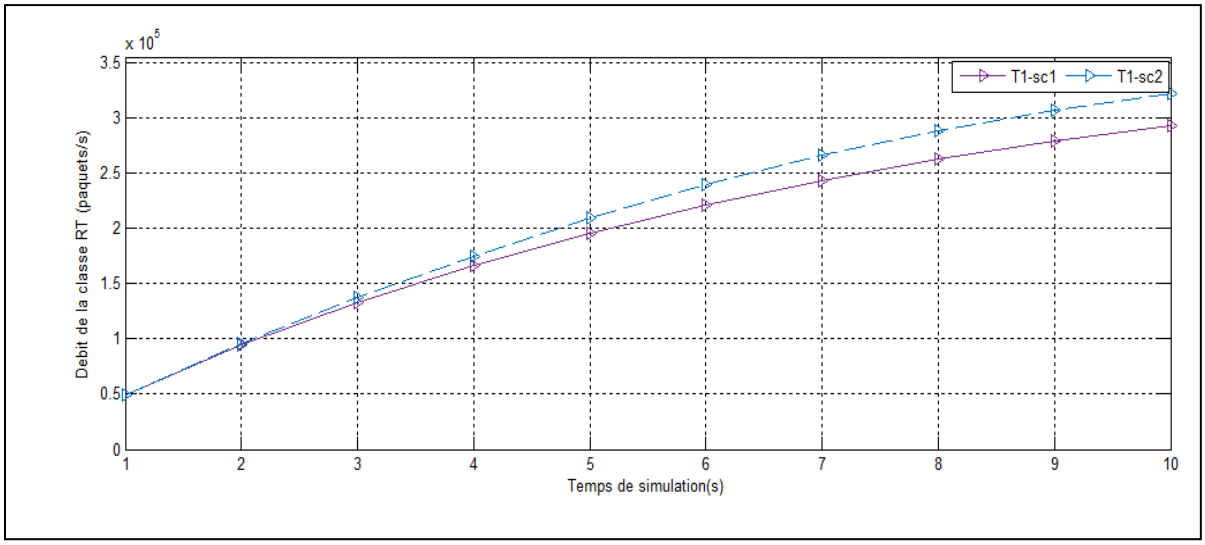


Figure 4.10 Débits RT VS le temps de simulation

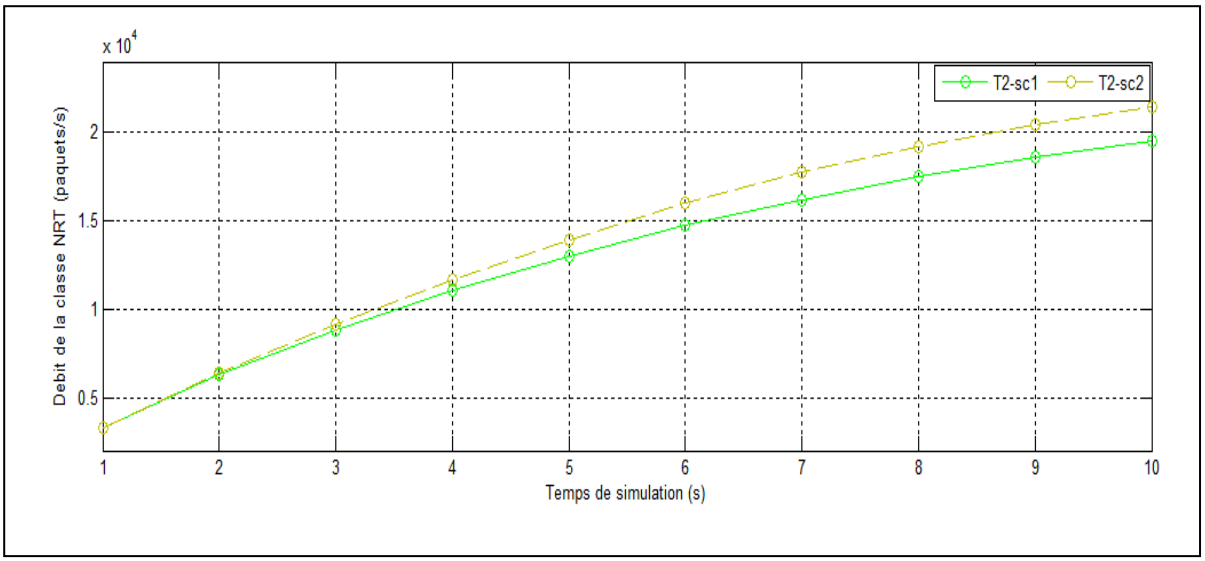


Figure 4.11 Débits NRT VS le temps de simulation

Les figures 4.9, 4.10 et 4.11 représentent les débits des trois classes de trafics .Les débits de la classe *handoff* sont presque similaires dans les deux scenarios durant l'intervalle entre 1 s et 3 s (figure 4.9). Puis, le débit T0-sc2 commence à s'accroitre d'une façon plus importante pendant le temps de simulation restant.

Les débits T1-sc2 et T2-sc2, sont légèrement supérieurs par rapport au T1-sc1 et T2-sc1 respectivement pendant les trois premières secondes. Par contre, les deux débits ont considérablement augmenté dans la période de 3s à 10 s (figure 4.10 et 4.11). Donc, plus le nombre d'inters arrivées dans le réseau est important, plus les débits sont améliorés

Étude du délai dans les files d'attente

Comme montré par la figure 4.12, le délai moyen expérimenté par les paquets *handoff* dans le deuxième scenario est plus élevé que dans le premier et demeure supérieure durant tout l'intervalle de simulation, à partir de t = 5 s, Wq_0-sc2 se rapproche de plus en plus de Wq_0-sc1, tel que à t = 10 s, Wq_0-sc1 = 0.11033 s et Wq_0-sc2 = 0.11038 s car la variation du délai est directement proportionnelle à la variation de la longueur de la file d'attente Lq_0-sc2 qui a augmenté dans le deuxième scenario (figure 4.6). Par contre, le délai évolue d'une façon

opposée à la variation du débit T0-sc2 qui a connu un accroissement plus significatif à partir de t = 5 s.

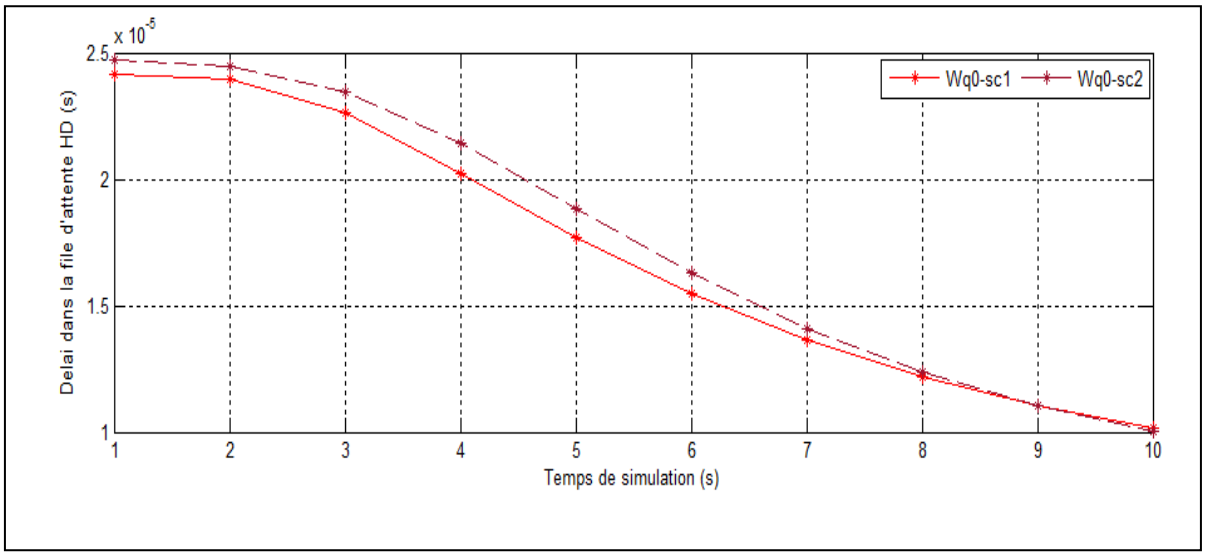


Figure 4.12 Délais moyens dans la file d'attente *handoff* VS le temps de simulation

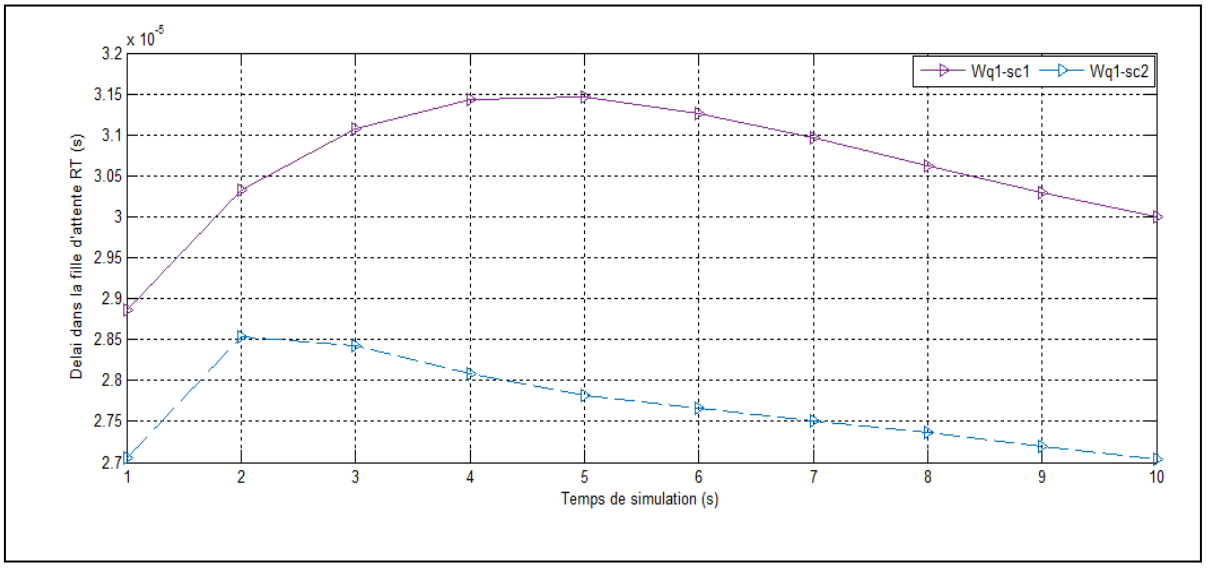


Figure 4.13 Délais moyens dans la file d'attente RT VS le temps de simulation

Quant aux délais dans les files RT et NRT, il convient de remarquer qu'ils ont diminué dans le deuxième scenario (figure.4.13 et 4.14) suite à la réduction du nombre moyen de paquets VoLTE et FTP dans les deux files, en particulier Wq_1-sc2, qui a connu un abaissement

considérable. La variation des délais Wq_1 et Wq_2 à chaque instant t est aussi fonction des valeurs des débits T1-sc2 et T2-sc2.

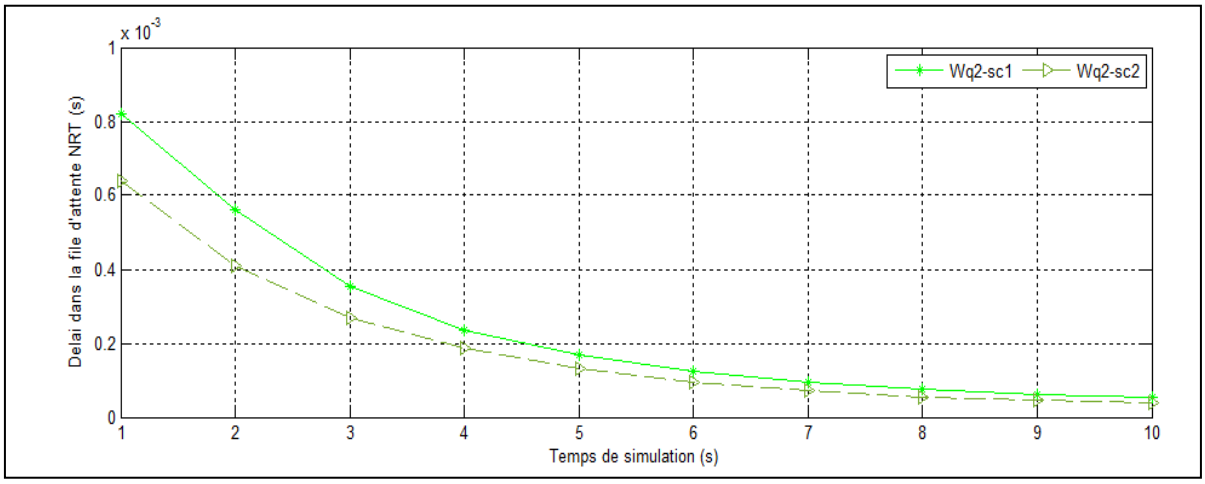


Figure 4.14 Délais moyens dans la file d'attente NRT VS le temps de simulation

Étude de la probabilité de blocage

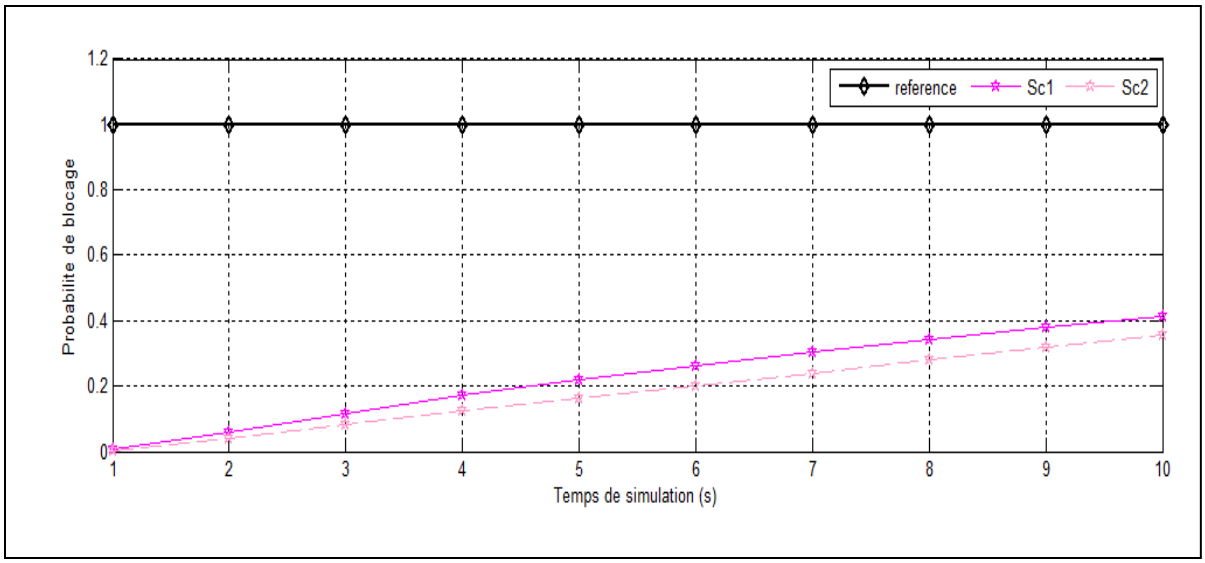


Figure 4.15 Probabilités de blocage VS le temps de simulation

La figure 4.15 montre que la probabilité de blocage dans le deuxième scenario est inférieure à la probabilité de blocage dans le premier scenario. Autrement dit, le nombre d'utilisateurs rejetés par le réseau est moins élevé, ce qui a permis d'améliorer les débits de chaque classe.

Étude de la perte de paquets

Les figures 4.16, 4.17 et 4.18 affirment que le scenario 2 révèle un taux de perte de paquets plus petit que le scenario1, l'algorithme de courtoisie a permis de réduire le nombre de paquets rejetés de 0.0134 dans la file *handoff*, de 0.0429 dans la file RT et de 0.0286 dans la file NRT à t = 10 s.

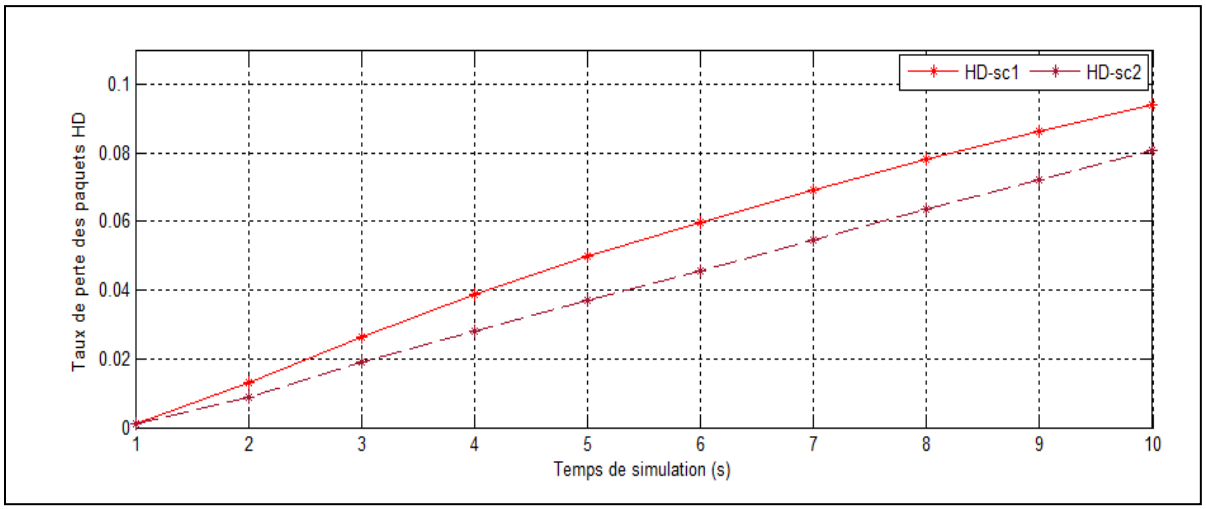


Figure 4.16 Taux de perte de paquets *handoff* VS le temps de simulation

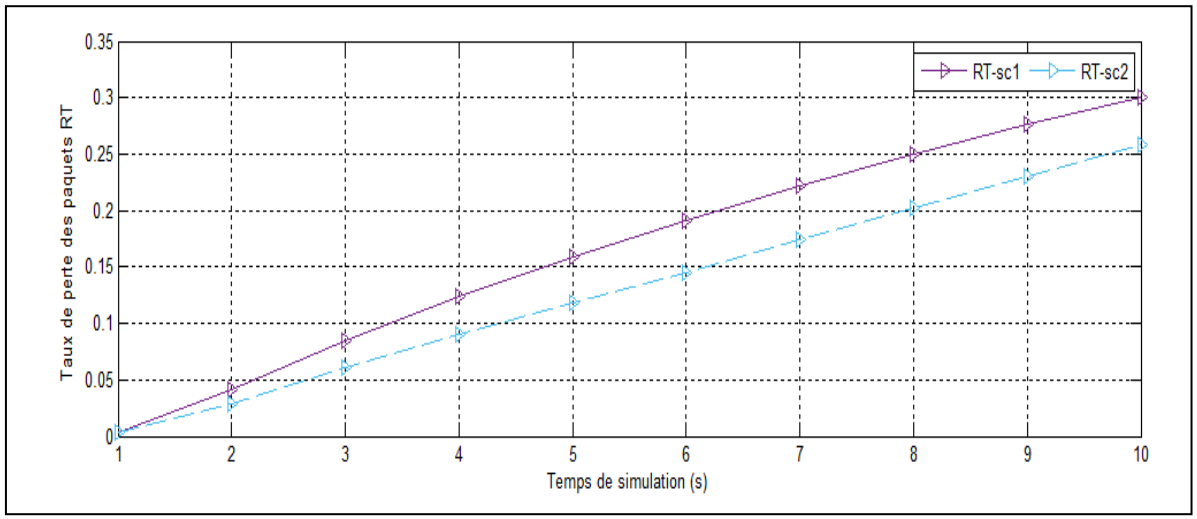


Figure 4.17 Taux de perte de paquets RT VS le temps de simulation

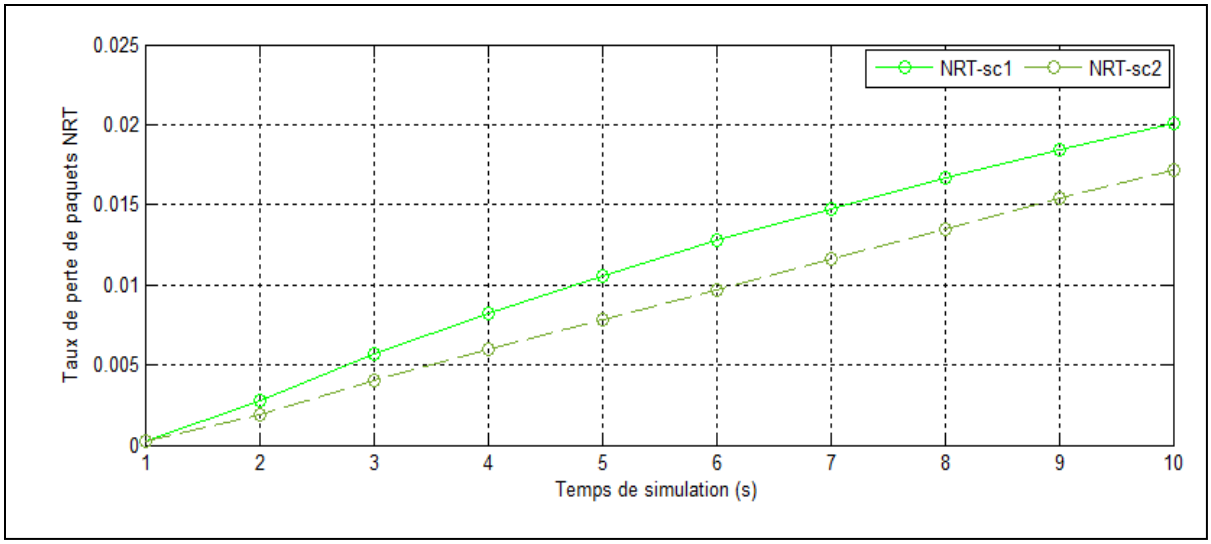


Figure 4.18 Taux de perte de paquets NRT VS le temps de simulation

4.3.3 Scénario 3: La classe NRT servie selon PQ

Les paquets NRT possèdent la plus haute priorité, ils sont transmis avec le taux μ_2 pendant l'intervalle T_{r4} inclus dans la période de transmission des paquets RT du deuxième scenario, Les paquets *handoff* et RT partagent le débit de chaque serveur avec des taux $w_0^{``}$ et $w_1^{``}$.
La valeur de $w_0^{``}$ est choisie égale à $w_0^{`}$ afin d'isoler l'impact du changement des taux sur les résultats. Ainsi, $w_1^{``}$.= 0.1 et $w_0^{``}$= 0.9.

La valeur moyenne de x déterminant le $seuil_{NRT}$ est fixée à $\frac{2Wq_2}{3Wq_1}$ afin que la durée correspondante au temps d'attente moyen des paquets *handoff* soit partagée entre les paquets RT et NRT puisque les seuils peuvent influencer considérablement les résultats. L'intervalle de transmission des paquets RT en mode PQ est choisi égal à $Wq_1/2$. De ce fait, la période d'attente des paquets NRT avant d'être transmis selon PQ correspond au $seuil_{NRT}$ qui est égale à ($T_{r1} + Wq_1/2$), d'où $3Wq_1/2$. D'autre part, $seuil_{NRT}$ a été défini comme:

$$seuil_{NRT} = (temps\ d'attente\ moyen\ NRT)/x$$

Ainsi, *x* doit satisfaire:

$$\frac{Wq_2}{x} = \frac{3Wq_1}{2} \tag{4.7}$$

D'où:

$$x = \frac{2Wq_2}{3Wq_1} \tag{4.8}$$

4.3.3.1 Résultats numériques

Tableau 4.6 Résultats numériques du scénario 3 à t = 3s

Paramètre	***Q_HD***	***Q_RT***	***Q_NRT***
Nombre de paquets dans la file (paquets)	1.04326	3.97020	2.4060
Délai moyen dans la queue × 1.0e-04 (s)	0.24204	0.28785	2.61660
Débit ×1.0e+05 (paquets/s)	0.43102	1.37927	0.09195
Taux des paquets perdus	0.01824	0.05836	0.00389

Les valeurs des paramètres citées dans le tableau 4.6 ont été prises à t = 3 s, en les comparant avec les valeurs du tableau.4.3, il s'en déduit que le nombre de paquets dans Q_HD a encore augmenté. En contrepartie, cette grandeur a diminué dans Q_RT et Q_NRT. Bien que le scenario3 cause une légère élévation des délais *handoff* et RT. Néanmoins, il offre les meilleurs débits, ce qui a permis de réduire le taux de paquets perdus.

La période de service des paquets NRT suivant PQ débute lorsque t = $seuil_{NRT}$ et termine à t = $seuil_{RT}$ + $seuil_{HD}$**,** cette période correspond aux états de transitions dont les indices *i, j, k* sont cités dans le tableau 4.7.

Tableau 4.7 Valeurs des seuils relatifs à la période de transmission des paquets NRT selon PQ

Paramètre	**Résultat**									
Temps de simulation (s)	1	2	3	4	5	6	7	8	9	10
i_{RT}	1	1	1	1	1	1	1	1	1	1
j_{RT}	1	1	1	1	1	1	1	1	1	1
k_{RT}	0	0	0	0	0	0	0	0	0	0
i_{NRT}	1	1	1	1	1	1	1	1	1	1
j_{NRT}	1	1	1	1	1	1	1	1	1	1
k_{NRT}	1	1	1	1	1	1	1	1	1	1
i_{HD}	11	11	11	11	11	0	0	0	0	0
j_{HD}	11	11	11	11	11	5	7	7	8	7
k_{HD}	11	11	11	11	11	4	2	1	0	0

Selon les statistiques collectées, les valeurs de (i_{NRT}, j_{NRT}, k_{NRT}) restent fixes durant le temps de simulation, les valeurs obtenues des seuils permettent une période de transmission suffisante pour analyser davantage le comportement du système après l'application de la deuxième conception de l'algorithme.

Tableau 4.8 Valeurs des débits, probabilité de blocage et pertes de paquets du scénario 3

Paramètre	**Résultat**									
Temps de simulation (s)	1	2	3	4	5	6	7	8	9	10
T0-SC2 ×1.0e+05 (Paquets/s)	0.15530	0.30094	0.43102	0.54864	0.65475	0.74942	0.83109	0.90037	0.95833	1.00644
T1-SC2 ×1.0e+05 (Paquets/s)	0.49699	0.96300	1.37927	1.75564	2.09520	2.39814	2.65950	2.88120	3.06660	3.22059
T2-SC2 ×1.0e+04 (Paquets/s)	0.33133	0.64201	0.91953	1.17045	1.39683	1.59879	1.77303	1.92084	2.04444	2.14710
Probabilite de bocage	0.00602	0.03699	0.08048	0.12218	0.16192	0.20062	0.24014	0.27969	0.31853	0.35588
Perte de paquets HD	0.00136	0.00838	0.01824	0.02768	0.03668	0.04546	0.05441	0.06337	0.07217	0.08063
Perte de paquets RT	0.00436	0.02682	0.05836	0.08859	0.11740	0.14546	0.17412	0.20280	0.23095	0.2580
Perte de paquets NRT	0.00029	0.00179	0.00389	0.00590	0.00783	0.00970	0.01161	0.01352	0.01540	0.01720

Le tableau 4.8 présente les résultats des grandeurs de débits, de la probabilité de blocage et du taux de perte de paquets**.** Les valeurs optimales sont observées pour le troisième scénario

lorsque t est compris entre 2 s et 5 s. Par contre, le deuxième scenario est meilleur pendant le temps de simulation restant (voir tableau.4.5).

4.3.3.2 Résultats graphique

Étude de la longueur des files d'attente

Les figures 4.19, 4.20 et 4.21 comparent successivement les résultats des longueurs des files d'attente Lq_0, Lq_1 et Lq_2 des trois scenarios, la figure 4.19 montre que pour t entre 1 s et 2 s, les tracés des graphes Lq_0-sc2 et Lq_0-sc3 sont à peu près semblables avec une légère diminution de Lq_0-sc3. Par contre, pour t entre 2 s et 6 s, Lq_0-sc3 devient supérieure à Lq_0-sc2. Bien que les paquets *handoff* soient servis avec un taux inchangé **(**$w_0^{``}$= $w_0^{`}$= 0.9**)** et pendant le même intervalle dans les deux scenarios. Cependant, la longueur de la file d'attente Lq_0 a changé dans le troisième scenario afin de maintenir la stationnarité du système.

La diminution de Lq_0-sc3 durant la deuxième seconde, a été compensée par un accroissement au niveau de Lq_1-sc3 dans le même intervalle (figure 4.20), cette augmentation est dû à la réduction de la période de transmission des paquets VoLTE suivant PQ afin de céder le privilège aux paquets FTP d'être prioritaires. Ensuite, Lq_1-sc3 commence à diminuer et atteint approximativement la valeur de Lq_1-sc2 à t= 6 s grâce à la transmission des paquets NRT suivant PQ qui a permis de réduire le nombre de paquets FTP qui sont supposés d'être transmis suivant CBWFQ. Ainsi, les paquets VoLTE sont servis plus rapidement.

Quant à Lq_2-sc3, la figure 4.21 montre que l'application de la deuxième conception de l'algorithme de courtoisie a permis d'obtenir la longueur la plus étroite lorsque t est inclus entre 1 s et 2 s. Ensuite, cette diminution devient moins importante à partir de t = 2 s jusqu'au t = 6 s, tel que à t = 6 s, Lq_2-sc2 = 1.51279 paquets et Lq_2-sc3 = 1.5087 paquets.

Quand le temps de simulation est inclus entre 6 s et 10 s, la période de transmission des paquets RT et NRT selon PQ dans le deuxième et le troisième scenario devient plus courte à cause des seuils qui ont changé (voir tableau 4.7). Par conséquent, lorsque les intervalles de transmission sont moins larges, les deux mécanismes se convergent.

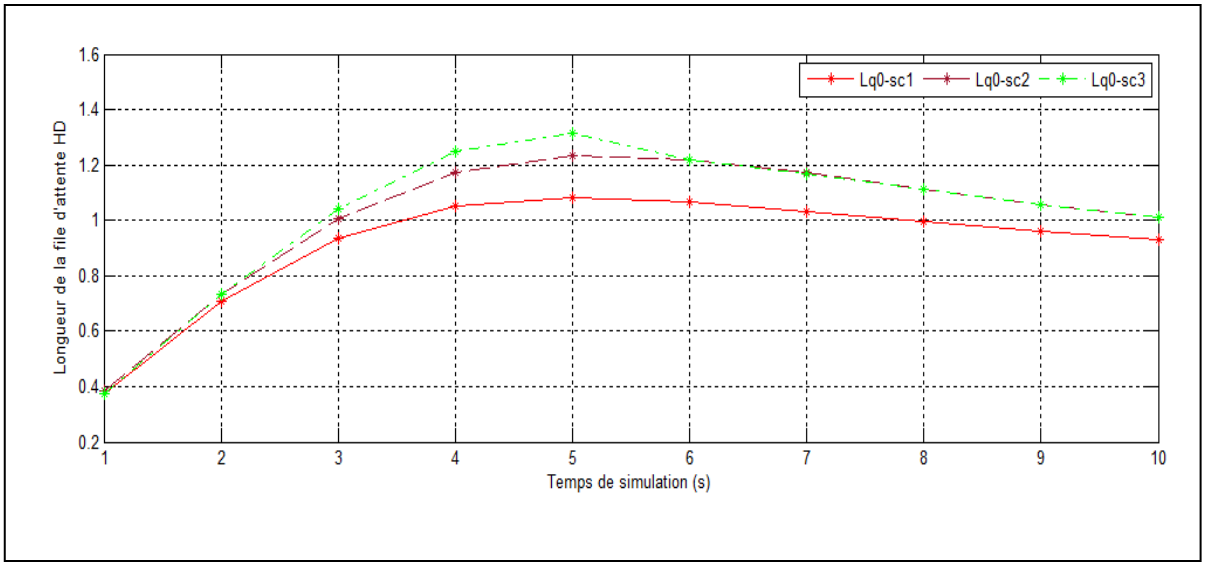


Figure 4.19 Longueurs de la file d'attente *handoff* VS le temps de simulation

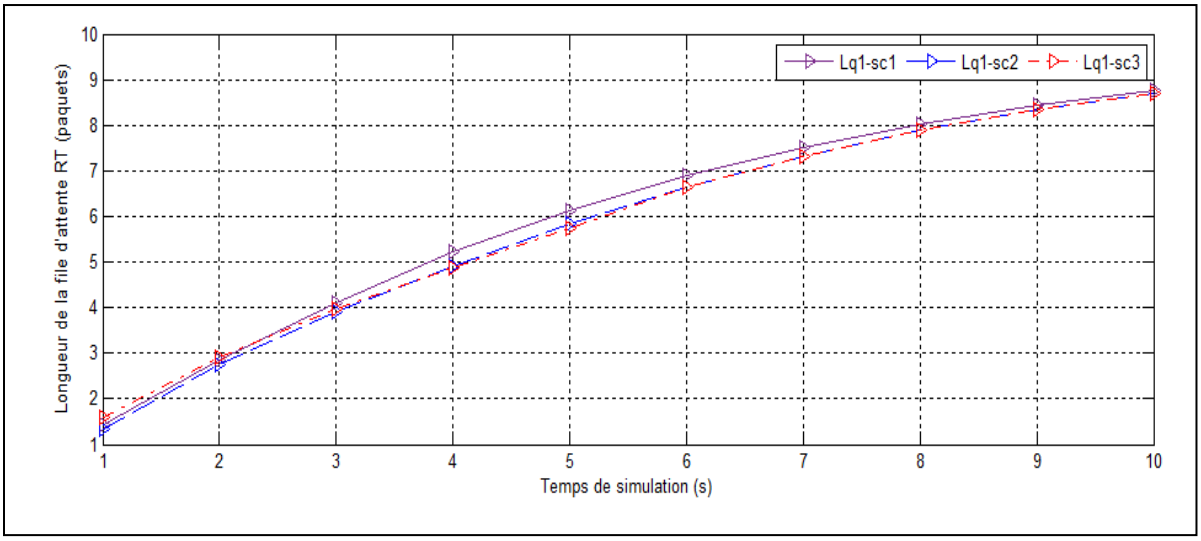


Figure 4.20 Longueurs de la file d'attente RT VS le temps de simulation

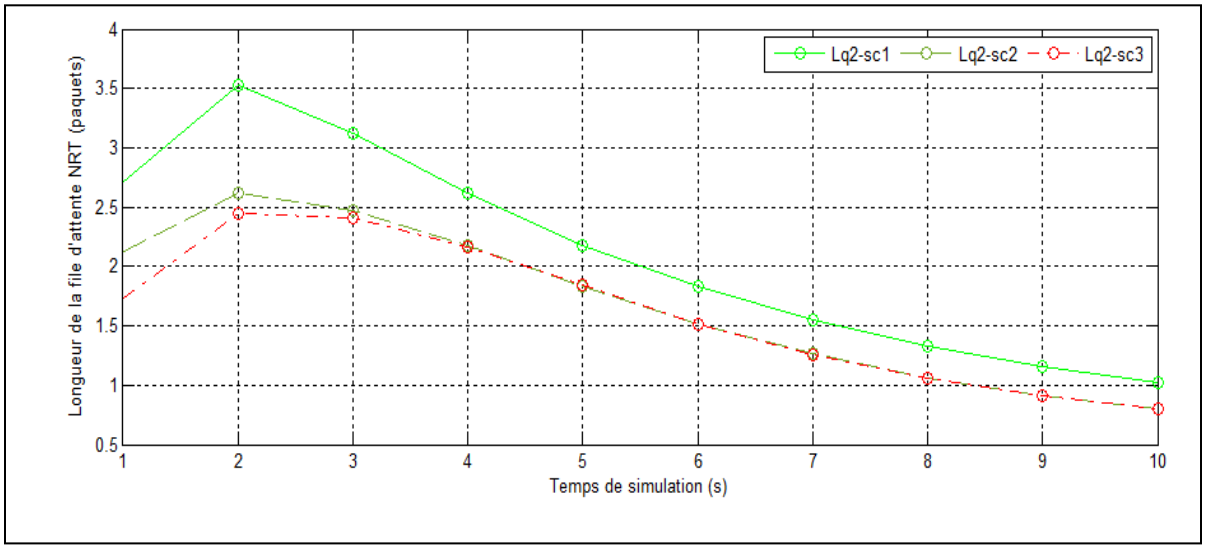


Figure 4.21 Longueurs de la file d'attente NRT VS le temps de simulation

Étude du débit

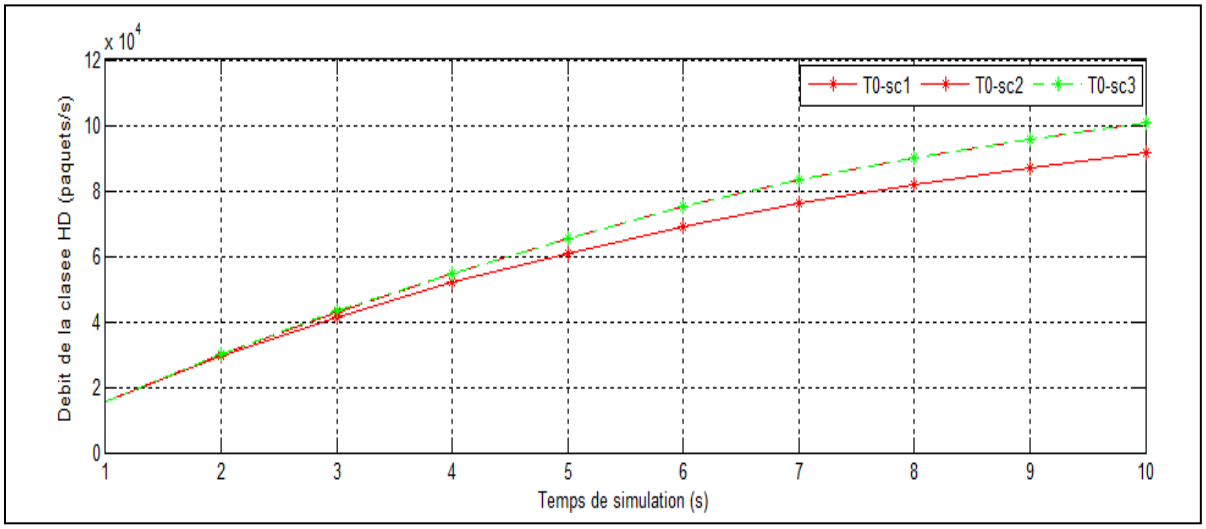


Figure 4.22 Débits *handoff* VS le temps de simulation

Les figures 4.22, 4.23 et 4.24 représentent le débit des classes *handoff*, RT et NRT successivement. Selon les tracés des graphes, il est difficile d'observer l'effet de la deuxième conception sur le nombre d'inters arrivées acceptées et servies. Or, les valeurs du tableau 4.8 montrent que les débits ont légèrement diminués par rapport au second scenario pour t entre

1 s et 2 s, c'est-à-dire, avant la congestion car le service des paquets NRT suivant PQ a causé une augmentation au niveau de la longueur de la file d'attente RT ce qui a conduit à une occupation plus importante du buffer. Par conséquent, les trois classes ont expérimenté des taux de blocage plus élevés. Toutefois, ces taux de blocage demeurent toujours inferieur aux taux obtenus dans le scenario de base.

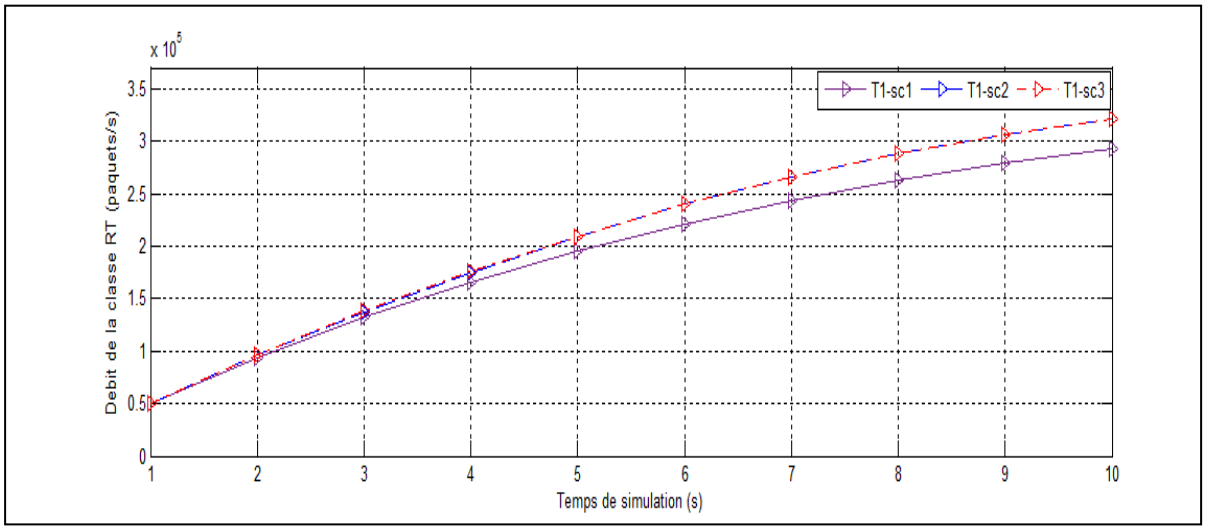


Figure 4.23 Débits RT VS le temps de simulation

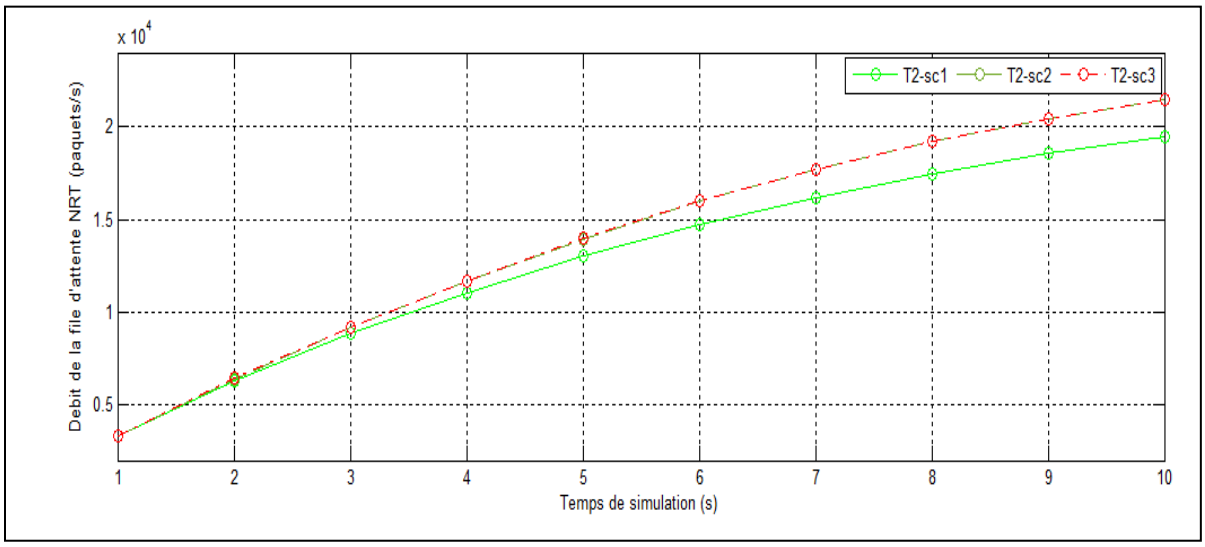


Figure 4.24 Débits NRT VS le temps de simulation

Lorsque t est inclus entre 2 s et 5 s, le scenario 3 offre les meilleurs débits pour les trois classes car les intervalles de transmission des paquets NRT suivant PQ sont plus large.

Dans le temps de simulation restant, les débits obtenus dans le deuxième scenario sont légèrement supérieurs aux débits obtenus du troisième scenario à cause des intervalles T_{r4} qui sont devenus plus étroits.

Étude du délai dans les files d'attente

Le délai moyen des paquets de chaque classe est illustré dans les figures 4.25, 4.26 et 4.27, les variations des délais dépendent principalement des longueurs des files d'attente et des valeurs des débits obtenues.

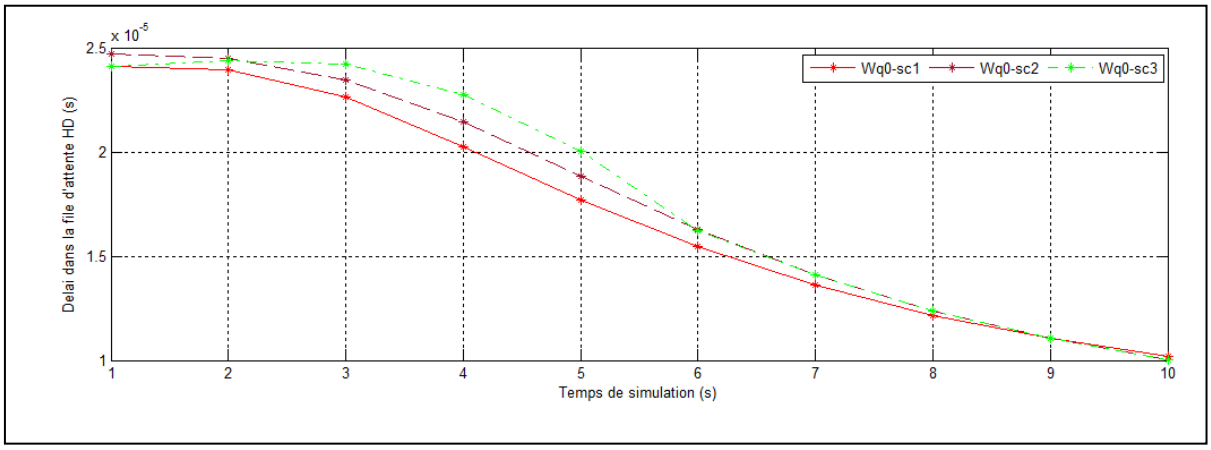


Figure 4.25 Délais moyens dans la file d'attente *handoff* VS le temps de simulation

Dans la figure 4.25, Wq_0-sc3 est inférieure à Wq_0-sc2 quand t est inclus entre 1 s et 2 s car Lq_0-sc3 est inferieure à Lq_0-sc2 et le nombre d'inter-arrivées *handoff* dans les deux scenarios est presque semblable avec une légère diminution de T0-sc3. Ensuite, Wq_0-sc3 devient supérieur à Wq_0-sc1 et Wq_0-sc2 pour t > 2 s jusqu'au t = 6 s à cause de Lq_0-sc3 qui a connu une augmentation assez importante. Or, l'accroissement de T0-sc3 est moins significatif. À partir de t = 6 s, Wq_0-sc3 et Wq_0-sc2 deviennent presque similaires vue que les valeurs des longueurs des files d'attente et des débits sont également presque similaires

avec une légère diminution du délai du troisième scenario, tel que à t = 10 s, Wq_0-sc2 = 0.010031 ms et Wq_0-sc3 = 0.01003 ms.

En ce qui concerne les délais dans la file RT, le troisième scenario repère la valeur la plus élevée parmi toutes les valeurs pour t entre 1 s et 2 s à cause de Lq_1-sc3 qui a augmenté et du débit T1-sc3 qui a légèrement diminué. Ensuite, le délai Wq_1-sc3 décroit et devient inférieur à Wq_1-sc1 et à Wq_1-sc2 dans la période de 4 s à 6 s car Lq_1-sc3 a diminué dans cette période et le débit T1-sc3 a augmenté sauf pour t = 6 s où les débits dans le deuxième et le troisième scenario sont presque similaires. À partir de t = 6 s, Wq_1-sc2 et Wq_1-sc3 sont visuellement semblables.

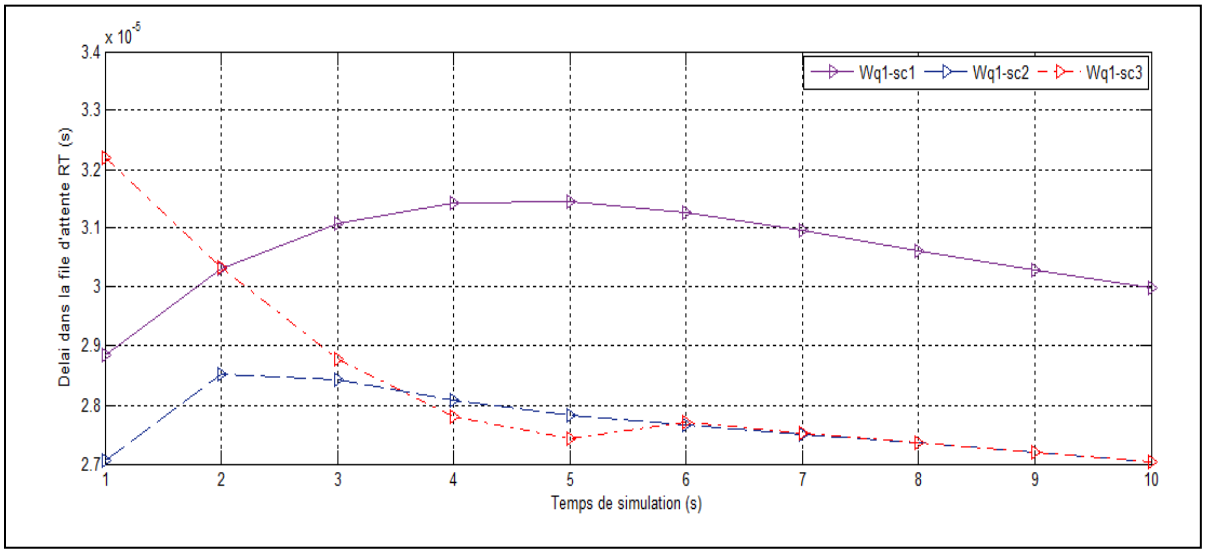


Figure 4.26 Délais moyens dans la file d'attente RT VS le temps de simulation

Lorsque t est compris entre 1 s et 4 s, Wq_2-sc3 est le plus court grâce à Lq_2-sc3 qui a diminué et T2-sc3 qui a augmenté. Ensuite, quand t = 2 s, Wq_2-sc3 se rapproche de Wq_2-sc2 d'où à partir de t = 4 s, les deux délais deviennent approximativement analogues car les longueurs des files d'attentes et les débits sont à peu près équivalents, tel que a t = 10 s, Wq_2-sc3 = 0.03718 s et Wq_2-sc2 = 0.03719 ms.

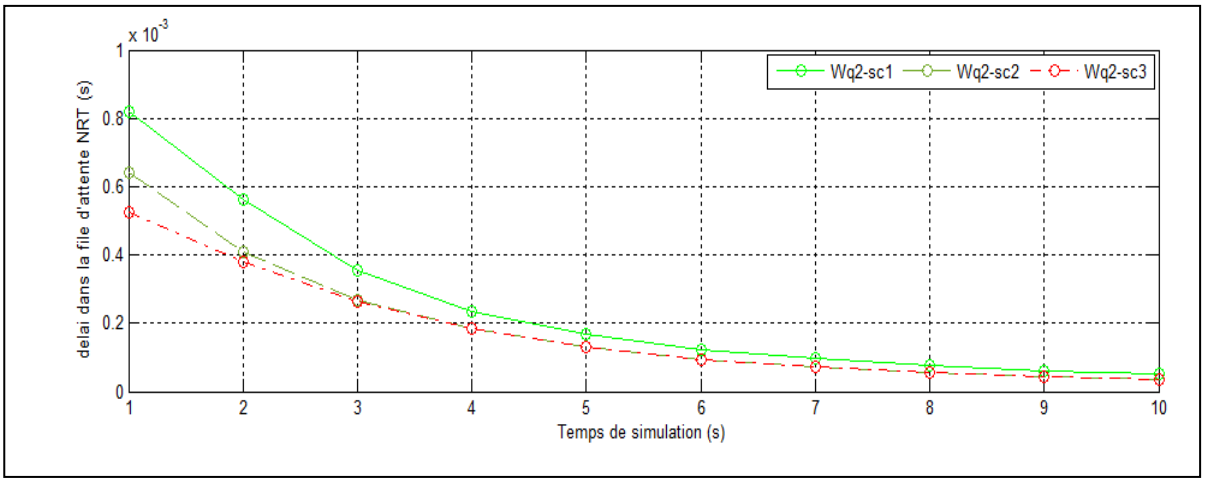


Figure 4.27 Délais moyens dans la file d'attente NRT VS le temps de simulation

Étude de la probabilité de blocage dans le buffer K

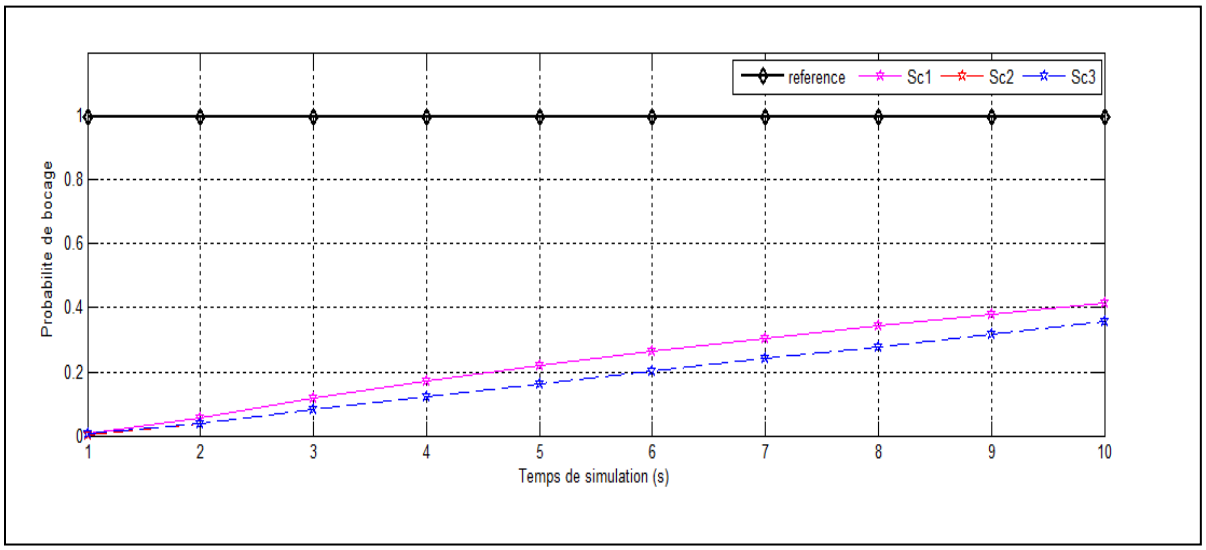


Figure 4.28 Probabilités de blocage VS le temps de simulation

Les tracés des graphes de la probabilité de blocage de sc2 et sc3 sont équivalents. Toutefois, les valeurs différent légèrement dans les deux scénarios, le troisième scenario donne les meilleurs résultats de π^* pour t entre 2 s et 5 s (voir tableau.4.8).

Étude de la perte de paquets

La perte de paquets relative aux trois files est illustrée par les figures 4.29, 4.30 et 4.31. Selon les tracés des graphes résultants, l'application de la deuxième conception de l'algorithme de courtoisie a minimisé le rejet des paquets *handoff*, RT et NRT dans la période comprise entre 2 s et 5 s. Néanmoins, des résultats à peu près semblables sont obtenus dans le deuxième et le troisième scenario pour t allant de 6 s à 10 s.

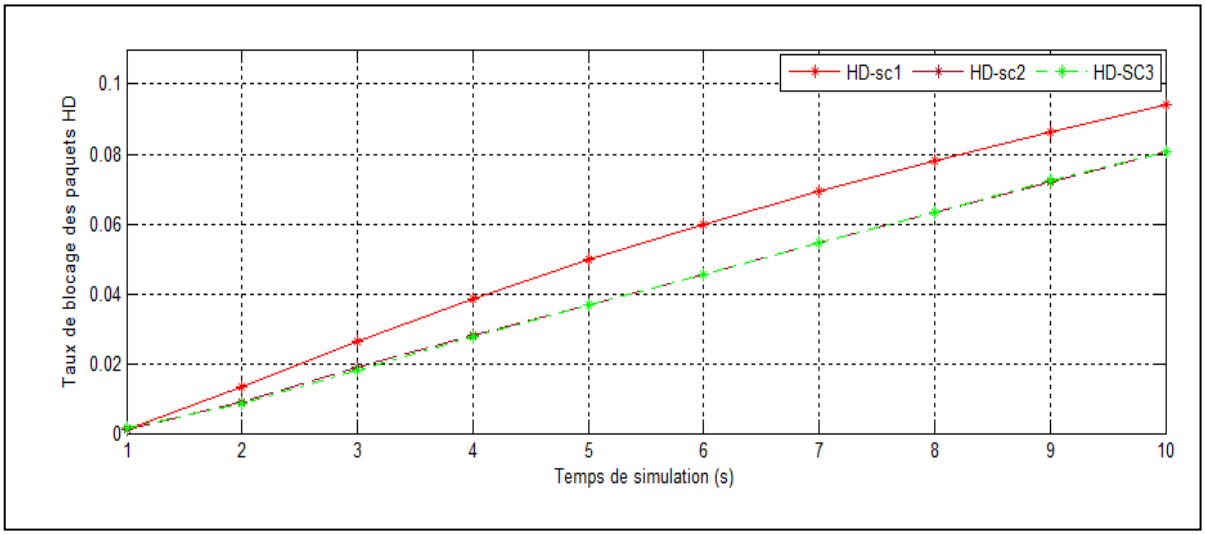


Figure 4.29 Taux de perte des paquets *handoff* VS le temps de simulation

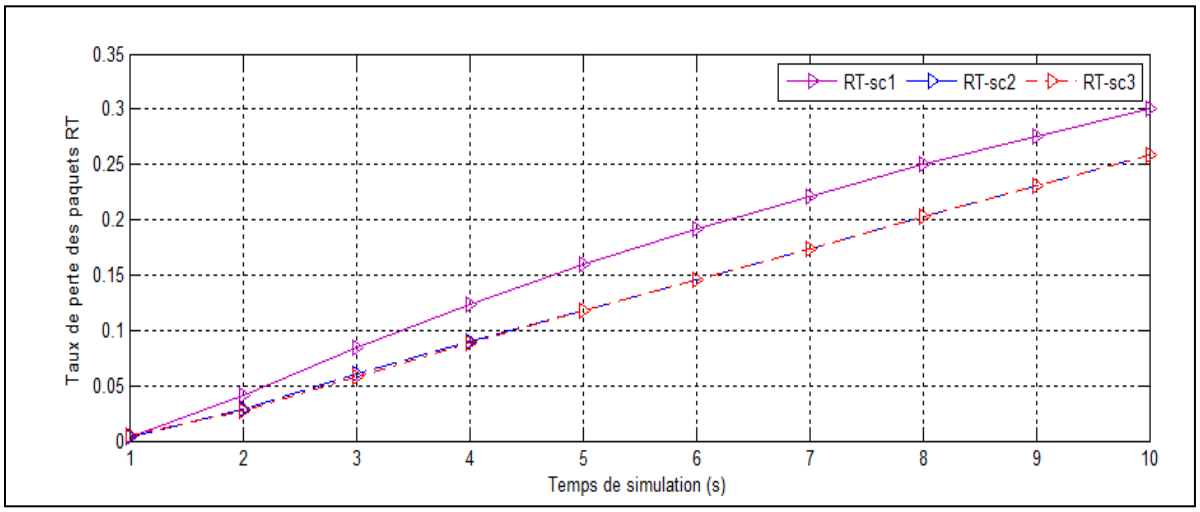


Figure 4.30 Taux de perte des paquets RT VS le temps de simulation

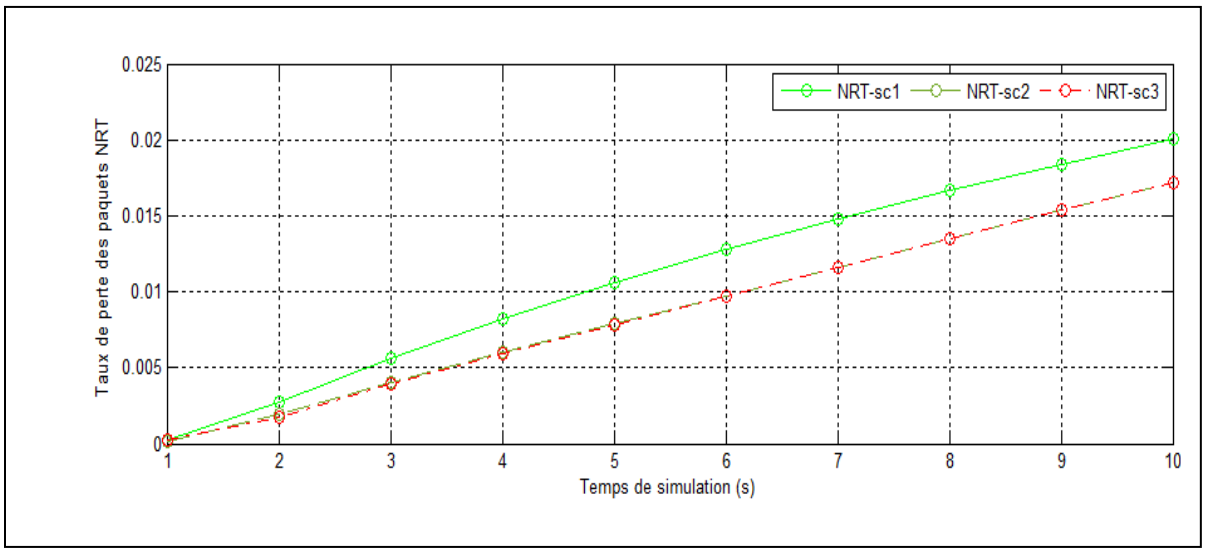


Figure 4.31 Taux de perte de paquets NRT VS le temps de simulation

Malgré la transmission des paquets NRT suivant PQ qui a temporisé le service des paquets moins tolérants au délai (*handoff* et RT). Néanmoins, l'accroissement brutal du taux de perte de paquets a été évité, car chaque augmentation au niveau de la longueur de file d'attente d'une classe, a été compensée par une diminution au niveau de la longueur de file d'attente d'une autre classe.

4.4 Conclusion

Dans ce chapitre, le modèle considéré a été établi suite à l'agrégation de 4 composantes porteuses de 20MHZ, utilisant chacune la modulation SC-FDMA dans la liaison montante. Trois scenarios ont été réalisés afin d'analyser séparément le comportement de chaque conception.

Le premier scénario constitue le scénario de référence où la classe *handoff* est constamment prioritaire grâce à la politique PQ, alors que les nouveaux trafics RT et NRT sont servis selon CBWFQ avec l'attribution du poids le plus élevé à la classe RT. Afin d'illustrer la congestion dans le système, le nombre d'inters arrivés se fait augmenter d'une unité à chaque seconde. Tandis que le taux de service reste fixe.

Il a été prouvé suivant les résultats des simulations que le premier scénario assure en continu la stationnarité du système même durant les situations de congestion avec des taux de blocage et perte de paquets acceptables. Cependant, il est meilleur de réduire le taux de rejet des paquets et d'améliorer les débits. De ce fait, le scénario 2 offre la possibilité au nouveau trafic RT d'être transmis selon PQ durant la période équivalente au temps d'attente moyen des paquets *handoff*, alors que les paquets *handoff* et NRT sont transmis selon CBWFQ en maintenant toujours la classe *handoff* prioritaire par rapport à NRT.

Les valeurs des grandeurs de performance démontrent une augmentation du délai moyen et de la longueur de la file *handoff*. Toutefois, cet accroissement demeure toujours dans les limites permises par les réseaux LTE-A et ne dégrade pas la qualité de service. En outre, les taux de blocage des trois classes ont nettement diminué d'où un gain au niveau des débits.

Le troisième scénario se fonde sur le deuxième et illustre le cas où la classe RT cède le privilège à la classe NRT qui devient prioritaire lorsque le rapport de temps d'attente moyen des paquets NRT sur une valeur x prédéfinie atteint un certain seuil, les paquets RT et NRT partagent l'intervalle correspondant au temps d'attente moyen des paquets *handoff*. Les

statistiques recueillies montrent que le débit a encore augmenté et que le nombre de paquets a également réduit quand les intervalles de transmission des paquets NRT suivant PQ sont plus étendus.

Les résultats expérimentaux ont prouvé la fiabilité du nouveau mécanisme. Le deuxième scénario donne les meilleurs résultats dans la première seconde quand les ressources sont suffisantes pour servir les trois classes. Par ailleurs, dans les situations de congestion, le troisième scénario assure une meilleure gestion des ressources, en particulier quand les intervalles dédiés à la transmission des paquets NRT selon PQ sont plus large. Par contre, le deuxième scénario convient mieux lorsque la congestion s'accentue.

CONCLUSION

La gestion des ressources dans les réseaux mobiles est essentielle afin d'améliorer les performances du système et de satisfaire les demandes des utilisateurs et les exigences des applications et services multimédias en termes de qualité de service. Plusieurs travaux de recherches qui traitent l'ordonnancement ont été préposés pour les réseaux LTE-A. Toutefois, la majorité des solutions privilégient le trafic de haute priorité d'où le trafic de faible priorité expérimente des taux de blocage et de perte de paquets élevés, notamment dans les situations de congestion.

Dans ce mémoire, une nouvelle solution d'ordonnancement dans la liaison montante des réseaux LTE-A a été développée dans le but d'assurer l'équité de service aux différentes classes de trafics et d'augmenter le nombre d'utilisateurs satisfaits dans le réseau. De plus de réduire le délai moyen des classes de basse priorité, d'optimiser le débit et de diminuer le taux de blocage et de perte de paquets avec le maintien du bon niveau de la QoS de la classe de haute priorité.

l'algorithme de courtoisie consiste à attribuer alternativement la plus haute priorité de service aux trois classes de trafic, à savoir la classe *handoff*, la classe RT et la classe NRT suivant leurs temps d'attente moyens. Trois scénarios de simulations ont été réalisés avec MATLAB (version R2014a) afin d'évaluer les performances de l'algorithme. Le premier scénario constitue le scénario de référence où les priorités demeurent inchangées une fois attribuées, la classe *handoff* est prioritaire suivie de la classe RT. Tandis que la classe NRT détient la plus faible priorité. Le deuxième et le troisième scénario ont été réalisés afin d'analyser l'influence du changement des priorités sur les différents indicateurs de performances. Ainsi, dans le deuxième scénario, la classe RT devient prioritaire durant un intervalle équivalent au temps d'attente moyen des paquets *handoff*, alors que la classe NRT possède toujours la priorité la plus faible. Dans le troisième scénario, la plus haute priorité est attribuée successivement à la classe RT puis à la classe NRT durant le même intervalle que précédemment.

Selon les résultats obtenus, le premier scénario maintient le système stationnaire avec des taux de rejets et des débits acceptables, notamment dans les situations de congestions. Le deuxième scénario a permis une réduction des délais et des longueurs des files d'attente des classes RT et NRT. Ainsi une diminution des taux de blocage et de perte de paquets des trois classes, d'où une optimisation des débits. Par ailleurs, une augmentation de la longueur de la file d'attente *handoff* a été constatée. Toutefois, la qualité de service de la classe *handoff* n'a pas été dégradée. Le troisième scénario offre des résultats plus performants que le deuxième scénario lorsque les intervalles de transmission dédiés à la classe NRT s'accroissent.

En conclusion l'algorithme de courtoisie d'ordonnancement dans la liaison montante répond aux objectifs fixés, l'équité de service a été bien assurée et le nombre d'utilisateurs satisfaits a augmenté. En outre, les délais des classes de basse priorité sont réduits et les taux de blocage et de perte de paquets dans le système ont diminué tout en maintenant le niveau de la qualité de service de la classe de haute priorité.

.

RECOMENDATIONS

Bien que l'algorithme de courtoisie ait prouvé sa fiabilité dans la gestion des ressources, il est conseillé d'évaluer ses performances dans le cas où les intervalles dédiés à la classe NRT dans le troisième scénario sont plus étendus.

Un autre scénario est également recommandé lorsque la portion du buffer dédiée à chaque file d'attente est spécifiée. De ce fait, il est important de prendre en considération le taux de blocage des paquets dû au remplissage total de chaque file, en particulier pour les files d'attente de plus haute priorité qui expérimentent une augmentation au niveau de la longueur de la file d'attente et du délai moyen à cause de la temporisation du service des paquets pendant une période équivalente au temps d'attente moyen.

Afin de rendre le mécanisme plus robuste, il est possible de définir un seuil qui permet de diviser la file d'attente en deux portions, l'algorithme de courtoisie est appliqué uniquement sur une seule portion. Ainsi, l'intervalle de temporisation du service des paquets est considéré le temps d'attente moyen de cette portion. La deuxième portion est utilisée pour compenser le nombre de paquets qui devrait arriver dans la file d'attente pendant le temps d'attente moyen de la première portion.

FUTURS TRAVAUX

L'algorithme de courtoisie d'ordonnancement dans la liaison montante ne prend pas en considération le cas où les utilisateurs LTE et LTE-A coexistent dans le réseau. Toutefois, il est possible de développer des solutions basées sur ce mécanisme en vue de concevoir un ordonnanceur qui tient en compte les deux types d'utilisateurs. De ce fait, deux méthodes peuvent être proposées. La première solution consiste à étendre le mécanisme de courtoisie à un système de six files d'attente, les trois files d'attente supplémentaires concernent les utilisateurs *handoff* LTE, RT LTE et NRT LTE, les utilisateurs LTE sont privilégiés par rapport aux utilisateurs LTE-A vu qu'ils ne peuvent être planifiés que sur une seule composante porteuse, l'ordre de priorité entre les trois files d'attente des utilisateurs LTE suit le même ordre défini pour les utilisateurs LTE-A.

La deuxième solution proposée est moins complexe que la première, elle est constituée de deux blocs de files d'attente, le premier bloc comprend deux systèmes de files d'attente $M/M/1/\frac{K}{2}$, chaque système est composé de trois files d'attente relatives aux classes *handoff*, RT et NRT des utilisateurs LTE et LTE-A, l'algorithme de courtoisie d'ordonnancement dans la liaison montante dans les réseaux LTE-A est appliqué au niveau de chaque file afin de déterminer l'ordre de paquets à envoyer au deuxième bloc, car les serveurs considérés dans ce bloc sont les files d'attente du deuxième bloc. Le deuxième bloc est constitué de deux files d'attente relatives aux trafics des utilisateurs LTE et LTE-A résultants de la première étape d'ordonnancement, ces deux trafics peuvent être modélisés avec un système M/M/S/K de deux dimensions. Cependant, les utilisateurs LTE sont servis qu'avec une seule composante porteuse et détiennent la plus haute priorité. L'ordonnancement entre ces deux files se fait soit en appliquant l'algorithme de courtoisie d'ordonnancement dans la liaison montante dans les réseaux LTE-A ou l'algorithme de courtoisie d'optimisation de la planification des ressources proposée par Michel KADOCH et Chafika TATA. L'algorithme qui donne les meilleurs résultats sera considéré dans la solution.

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

Abu-Ali, Taha, Salah M et Hassanein H. 2014. «Uplink scheduling in LTE and LTE-A: Tutorial, Survey and Evaluation Framework». Communications Surveys & Tutorials, IEEE,vol. 16, no 3, p. 1239 – 1265.

Alsawaai, Amina. 2010. «Performance Modelling and Analysis of Weighted Fair Queueing for Scheduling in Communication Networks ». Thèse de doctorat en génie informatique, Bradford, School of Computing, Information and Media, 111 p.

Ben Ali, Khitem,Faouzi, Zarai, Lotfi, Kamoun. 2010 «Reducing handoff dropping probability in 3GPP LTE network». In Communications and Networking (ComNet), 2010 Second International Conference on. (Tozeur, 4-7 Nov. 2010). p 1-8.Tunisia: University of Sfax.

Bendaoud, Fayssal, Marwen, ABDENNEBI et fedoua, DIDI. 2014. «Allocation des ressources radio en LTE». In LTE International Congress on Telecommunication and Application'14. (Bejaia, Apr 23-24 2014). Algeria: University of A.MIRA.

Bergida. En ligne. <http://www.eng.tau.ac.il/~shavitt/pub/EE60.pdf>. Consulté novembre 2015.

Bianchi et Al. 2013. «Wireless Access Flexibility» In First International Workshop, (Kaliningrad,Sep 4-6 2013, p 55-68. Berlin: Springer.

Bouguen, Yannik, Eric Hardouin et Francois-Xavier Wolff. 2012. LTE et les réseaux 4G, 1st ed. Paris: Eyrolles. 548 p.

Chen, Hsu, Tsai, Liao et Lin. 2014 «A Heuristic Design of Uplink Scheduler in LTE-A Networks» In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference. (Kitakyushu , 27-29 Aug 2014). 884 - 888.Taiwan: National Central University.

Chafika,Tata, kadoch, Michel. 2008 « Algorithme de courtoisie: optimisation de la performance des réseaux WIMAX fixes». École de technologie supérieure.

Chaudet. En ligne <http://docplayer.fr/2371246-Introduction-a-la-theorie-des-files-d-attente-claude-chaudet-claude-chaudet-enst-fr.html>. Consulté avril 2015.

Chuck, Semeria. 2001. <http://users.jyu.fi/~timoh/kurssit/verkot/scheduling.pdf>. Consulté septembre 2015

Cox, Chrostopher. 2012. «An introduction to LTE, LTE-Advanced, SAE and 4G Mobile Communications», 1st ed. Londre: John Wiley & Sons, 352 p.

Don. En ligne.<http://lteuniversity.com/get_trained/expert_opinion1/b/donhanley/archive/2013/09/11/how-big-is-a-voice-call.aspx>. Consulté septembre 2014.

Ekström, Ericsson.2009. «QoS control in the 3GPP Evolved Packet System».Communication Magazine, IEEE, vol. 47, no 2, p. 76 – 88.

Essia, Elhafsi et Molle. En ligne <http://alumni.cs.ucr.edu/ ~essia/ research/ LSAA_A_ 241878_O.pdf >. Consulté aout 2014.

Gautam, Natarajan. 2012. « Analysis of queues: methods and applications», 1st ed. USA: CRC press, 802 p.

Govindarajulu,Sree sharanya. 2011. «Reaction mechanism for packet size-based misbehaviour in wireless network». Mémoire de maîtrise en génie électrique, Wichita, Amrita Institute of Technology, 65 P.

Hallahan, Ryan et Jon M. Peha. 2010 «Policies for Public Safety Use of Commercial Wireless Networks» In 38 th Telecommunications Policy Research Conference. (Washington, Oct 1-3 2010). p 1-34. Pittsburgh: Carnegie Mellon University.

Hussain, Nasir. 2011. «Dynamic Admission Control and Scheduling for Efficient Channel Utilization in the LTE Network». Mémoire de maîtrise en génie informatique, Riyadh, King Saud University, 150 p.

Iturralde Mauricio, Steven Martin et Tara Ali Yahiya. 2013 « Resource allocation by pondering parameters for uplink system in LTE Networks» In Local Computer Networks (LCN), 2013 IEEE 38th Conference. (Sydney, NSW, 21-24 Oct 2013). 747 - 750. France: University of Paris-Sud – CNRS.

Kleinrock, 1975. «Queuing Systems: Computer Applications», 1st ed, Vol 2. New York: Wiley. 549 p.

Miao, Min, Jiang, Jin et Wang. 2014 « QoS-aware resource allocation for LTE-A systems with carrier aggregation » In Wireless Communications and Networking Conference (WCNC), 2014 IEEE. (Istanbul, 6-9 April 2014). p 1403 – 1408.

Penttinen, Jyrki. 2012. «The LTE/SAE deployment handbook», 1st ed. Londre: John Wiley & Sons, 407 p.

Qi-Ming, He. 2014. « Fundamentals of Matrix-Analytic Methods », 1st ed.New York: Springer, 346 p.

R. Kausar, Chen.Y et Chai.K. 2011 «QoS aware Packet Scheduling with adaptive resource allocation for OFDMA based LTE-advanced networks» In Communication Technology

and Application (ICCTA 2011), IET International Conference. (Beijing , 14-16 Oct 2011). 207 - 212.London: School of Electronic Engineering and Computer Science.

Safwat, Mahammad, Hesham , El-Badawy, Ahmad, Yehya, H. El-motaafy. 2014 «Performance Analysis for New Call Bounding Scheme with SFR in LTE-Advanced Networks». In High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC,CSS,ICESS), 2014 IEEE Intl Conf. (Paris, 20-22 Aug 2014). 442 - 451.Egypt: University of Cairo.

Sauter, Martin. 2011. «From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadband», 1st ed. Londre: John Wiley & Sons, 452 p.

Tara, Ali-Yahiya. 2011. «Understanding LTE and its performances», 1st ed. New York: Springer-Verlag, 261 p.

Vassilios, Vassilakis, Ioannis, Moscholiosy, Andreas, Bontozoglouz, Michael, Logothetisx. 2015 «Mobility-aware QoS assurance in software-defined radio access networks: An analytical study». In Local Computer Networks (LCN), 2013 IEEE 38th Conference. (London, 21-24 Oct 2013). 747 - 750. United Kingdom: University of Cambridge,.

Wang, Hua; Rosa Claudio et Pedersen, Klaus. 2010. «Performance of Uplink Carrier Aggregation in LTE Advanced Systems». In IEEE Vehicular Technology Conference Proceeding, (Ottawa, Sep 6-9 2010.). p 1-6.Denmark: Nokia Siemens Networks.

Wang, Lan, Geyong Min; Kouvatsos D.et Xiaolong Jin. 2009. «An Analytical Model for the Hybrid PQ-WFQ Scheduling Scheme for WiMAX Networks ».In Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic System Technology. Wireless VITAE, 1st International Conference. (Aalborg,May. 17-20 2009), p. 492-498.IEEE Publishers.