



THESE

Présentée devant

L' ECOLE SUPERIEURE DE COMMUNICATION DE TUNIS (SUPCOM)

Pour obtenir

Le grade de : **DOCTEUR DE L'ECOLE SUPERIEURE DE
COMMUNICATION DE TUNIS (SUPCOM) ET L'UNIVERSITE
OCCIDENTALE DE BRETAGNE (UBO)**

Mention : **TECHNOLOGIE DE L'INFORMATION
ET DE LA COMMUNICATION**

Par

Sonia BEN REJEB

Titre de la thèse :

Gestion de la QoS dans un Réseau Radio Mobile Multimédia

A soutenir le 28 Février 2006 devant la commission d'examen

Composition du Jury :

Mme Sihem GUEMARA
M. Farouk KAMOUN
Mme Noémie SIMONI
M. Pierre VILBE
M. Sami TABBANE
M. Zièd CHOUKAIR

Président de jury
Rapporteur
Rapporteur
Co-Directeur de thèse
Directeur de thèse
Co-Directeur de thèse

SUPCOM Tunis
ENSI Tunis
ENST Paris
UBO France
SUPCOM Tunis
SUPCOM Tunis

Abstract

In the general context of the mobile networks planning of third generation and by taking account of the mobility and the multiplicity of the services, the study of the behaviour of a such network proves to be necessary to compensate for a heterogeneous traffic load from the point of view of the characteristics of flows on the one hand and the quality of service on the other hand. The choice of a policy of resources allocation and management of the quality of service for a type of service given must be the subject of impact study on the total performances of the system in term of flow offered for each type of service on the one hand and in term of respect of the constraints of service on the other hand.

We propose an approach of QoS resources allocation management based mainly on parameters of quality of service and a mechanism of call admission control. The objective is to increase the acceptance level of the calls within the UMTS cellular access network /DiffServ core network while respecting the constraints of service quality. This mechanism allows to adapt the behaviour of the traffic of the characteristics of the network with respect to the parameters of quality of service such as : the rate of handovers, the traffic load, the dropping probability call, the grade of service (GoS), etc.

In this approach, the admission control is done with two levels: firstly on the level of the Capacity of Acceptance of New Calls per Cell (CANCC) recorded to the level of the SLA (Service Level Agreement) radio operator mobile, where the mobility of user is considered as a major constraint for call acceptance. The second stage takes place on the level of the core network of the operator. The acceptance of quality of service is defined by the contract SLA which specifies the parameters and the various levels of acceptance. At the global level, the acceptance of a call handover or a new call is based on a collegial and consensual decision between the various participating domains. Such a consensus allows a coherent decision making to hold a resource according to one of the levels of preset service by the SLA of the applicant then to allow to the caller to renegotiate the level of QoS of the required service according to the fluctuation of the traffic load or to negotiate new resources for new applications required during a session. The other participating domains can also initiate a renegotiation of the resources allocated with this session for the same reasons.

Résumé

Dans le contexte général de planification des réseaux mobiles de troisième génération (3G) et en tenant compte de la mobilité et de la multiplicité des services, l'étude du comportement d'un tel réseau s'avère nécessaire pour compenser une charge de trafic hétérogène du point de vue des caractéristiques des flux d'une part et la QoS (*Quality of Service*) d'autre part. Le choix d'une politique d'allocation des ressources et de gestion de la qualité de service pour un type de service donné doit faire l'objet d'étude d'impact sur les performances globales du système en terme de débit offert pour chaque type de service d'une part et en terme de respect des contraintes de qualité de service d'autre part.

Nous proposons une approche de gestion de QoS de l'allocation des ressources basée principalement sur des paramètres de qualité de service et un mécanisme de contrôle d'admission d'appel. L'objectif est d'augmenter le niveau d'acceptation des appels au sein d'un réseau d'accès cellulaire UMTS/réseau cœur DiffServ tout en respectant les contraintes de qualité de service. Ce mécanisme permet d'adapter le comportement du trafic aux caractéristiques du réseau vis à vis des paramètres de qualité de service tels que : le taux de handovers, la charge de trafic, la probabilité de rejet d'appel, le GoS (Grade de Service), etc.

Lors de cette approche, le contrôle d'admission se fait à deux niveaux : premièrement au niveau de la Capacité d'Acceptation de Nouveaux Appels par Cellule (CANAC) enregistrée au niveau du SLA (*Service Level Agreement*) radio mobile, là où la mobilité d'utilisateur est considérée comme une contrainte majeure pour l'acceptation d'un appel. La seconde étape a lieu au niveau du réseau cœur de l'opérateur. L'acceptation de la QoS est définie par le contrat SLA qui spécifie les paramètres et les différents niveaux d'acceptation. Au niveau global, l'acceptation d'un appel handover ou d'un nouvel appel se base sur une décision collégiale et consensuelle entre les différents domaines participants. Un tel consensus permet une prise de décision cohérente pour réserver une ressource suivant un des niveaux de service prédéfini par le SLA du demandeur puis par la suite de permettre à l'appelant de renégocier le niveau de QoS du service demandé en fonction de la fluctuation de la charge de trafic ou négocier de nouvelles ressources pour de nouvelles applications demandées au cours de la session. Les autres domaines participants peuvent également initier une renégociation des ressources allouées à cette session pour les mêmes raisons.

Remerciements

Je tiens tout d'abord à remercier mon co-directeur de thèse, Monsieur Sami TABBANE, Professeur à l'Ecole Supérieure des Communications de Tunis (SUPCOM), pour ses orientations, conseils et remarques, et de la qualité de nos rapports professionnels.

Je tiens à remercier Messieurs Zièd CHOUKAIR et Pierre VILBE de m'avoir fait l'honneur de co-diriger et d'encadrer les travaux de cette thèse dans les meilleures conditions lors de mes séjours à l'Ecole Nationale Supérieure des Télécommunications de Bretagne et à l'Université de Bretagne Occidentale. Je leur suis reconnaissante de tout ce qu'il m'ont appris. Qu'ils trouvent ici l'expression de ma profonde gratitude et reconnaissance.

Mes vifs remerciement s'adressent tout particulièrement à Madame Sihem GUEMARA, Maître de conférence à l'Ecole Supérieure des Communications de Tunis (SUPCOM) pour avoir présider le jury de ma thèse.

J'adresse ma reconnaissance à Madame Noémie SIMONI, Professeur à l'Ecole Nationale Supérieure des Télécommunications de Paris (ENST) et à Monsieur Farouk KAMOUN, Professeur à l'Ecole Nationale des Sciences de l'Informatique de Tunis (ENSI) pour m'avoir fait le grand honneur d'être les rapporteurs de ma thèse et ce en dépit de leurs nombreuses occupations et responsabilités.

Je n'oublie pas de remercier tous les membres de l'Unité de Recherche (MEDIATRON) de SUPCOM et du Laboratoire d'Informatique des Télécommunications (LIT) de l'ENST Bretagne de leurs accueils au sein de leurs équipes.

Je tiens à remercier chaleureusement toute ma famille, en particulier mes parents et mon mari, qui n'ont cessé de m'encourager et de m'aider aux moments difficiles pour que je réussisse au mieux.

Enfin, je tiens à remercier vivement toutes les personnes qui de près ou de loin ont contribué au bon déroulement de cette thèse. Qu'elles trouvent ici l'expression de ma profonde gratitude.

Dédicace

à mon père
à ma mère
à mon mari
à ma petite fille Ghofrane
à toute ma famille et
mes amis je dédie ce
travail

Table des matières

Chapitre 1

Introduction générale	10
1.1 Contexte de l'étude	10
1.2 Problématique de la thèse	11
1.3 Contribution de la thèse.....	12
1.4 Plan de la thèse	13

Partie 1 : Contexte

Chapitre 2

Planification d'un réseau radio mobile multimédia	18
2.1 Introduction.....	18
2.2 Processus de planification des systèmes radio mobiles de 3G	19
2.2.1 Objectifs dans la conception cellulaire d'un réseau cellulaire	21
2.2.2 Types de cellules considérés	22
2.2.3 Architecture d'un réseau radio mobile cellulaire	23
2.2.3.1 Le sous système radio	24
2.2.3.2 Le sous système réseau	24
2.2.4 Les mécanismes spécifiques des réseaux mobiles	24
2.2.4.1 Transfert inter-cellulaires (handover)	25
2.2.4.2 Gestion de l'itinérance (<i>roaming</i>)	25
2.2.4.2.1 Systèmes sans localisation	25
2.2.4.2.2 Systèmes utilisant des zones de localisation	25
2.2.5 La mobilité des utilisateurs	26
2.2.6 Le trafic	26
2.2.6.1 Modélisation de la demande	27
2.2.6.2 Techniques de prédiction	27
2.2.6.3 Distribution de la population	28
2.2.6.4 Distribution horaire	28
2.3 La mise en place d'un réseau UMTS	28
2.3.1 La topologie d'un réseau UMTS	29
2.3.2 La planification d'un réseau UMTS	30
2.3.2.1 Dimensionnement en trafic	31
2.3.2.2 Estimation de la demande	31
2.3.2.2.1 Mode circuit	31
2.3.2.2.2 Mode paquet	32
2.3.3 Le modèle de l'environnement considéré	33
2.4 Conclusion	33

Chapitre 3

Analyse des modèles de trafic	36
3.1 Introduction	36
3.2 Classe des services	37
3.2.1 Classe des services conversationnels	37
3.2.2 Classe des services à flux continu	37
3.2.3 Classe des services interactifs	38
3.2.4 Classe des services en mode téléchargement	39
3.3 Modèles de trafic	39
3.3.1 Modèle de communication téléphonique	39

3.3.2	Modèle de flux multimédia	40
3.3.3	Modèle de flux Internet	41
3.3.4	Modèle de transfert de fichier	42
3.3.4.1	Le modèle FUNET	42
3.3.4.2	Le modèle RAILWAY	42
3.3.5	Modèle de flux Web sur Internet	42
3.4	Conclusion	45
Chapitre 4		
	Caractéristiques de QoS d'un réseau radio mobile multimédia	47
4.1	Introduction	47
4.2	Critères de performance réseau	48
4.3	Critères de classification des paramètres de QoS	49
4.3.1	Caractéristiques principales de QoS	49
4.3.2	Classes de QoS	51
4.3.3	Profil d'usage	51
4.3.4	Les besoins en QoS	51
4.3.5	Le champ d'étude de QoS	52
4.3.6	Les défis posés par QoS	52
4.3.7	Gestion de QoS	52
4.4	Suivi et supervision de QoS	54
4.4.1	Indicateurs de QoS	54
4.4.2	Concept de QoS	54
4.4.3	Paramètres de réseau	55
4.4.4	L'OMC	55
4.5	Conclusion	56
Chapitre 5		
	Configuration d'un réseau radio mobile de bout en bout	58
5.1	Introduction	58
5.2	Evolution des réseaux radio mobiles vers le multimédia	59
5.3	Approvisionnement en ressources en UMTS-NGN	60
5.3.1	Le sur-approvisionnement	60
5.3.2	Le contrôle souple	60
5.3.3	L'admission strict	60
5.3.3.1	Gestion de la bande passante	60
5.3.3.2	Signalisation dans la bande	61
5.4	Correspondance de QoS IP et QoS UMTS	61
5.4.1	Paramètres de QoS bout en bout	63
5.4.2	QoS de bout en bout : plan de transmission	63
5.4.3	Délai bout en bout dans 'UMTS/IP	64
5.5	Architecture générale d'un réseau UMTS/NGN à base de règle	64
5.5.1	Le serveur d'allocation des ressources	65
5.5.2	Le serveur de règle	66
5.5.3	Le protocole COPS	67
5.5.4	IntServ, le protocole RSVP	69
5.5.5	DiffServ au sein des réseaux UMTS	72
5.6	Conclusion	73
<i>Partie 2 : Etat de l'art</i>		
Chapitre 6		
	Etat de l'art de la gestion de QoS de l'allocation des ressources dans un RRM.....	77
6.1	Introduction	77
6.2	Dimensionnement des ressources du système	78

6.2.1 Capacité d'un système cellulaire	78
6.2.2 Mobilité des utilisateurs.....	79
6.2.3 Impact d'une QoS dégradée	80
6.3 Différentes techniques d'allocation des ressources radio	82
6.3.1 <i>Fixed channel</i> allocation (FCA)	82
6.3.2 <i>Dynamic channel</i> allocation (DCA)	83
6.3.3 <i>Hybrid channel</i> allocation (HCA)	84
6.4 Processus de gestion de QoS de l'allocation des ressources de bout en bout.....	85
6.5 Travaux sur l'allocation des ressources et la gestion de la QoS	86
6.5.1 Approches et travaux divers	86
6.5.2 Projets et travaux divers	87
6.5.3 Discussion	88
6.5.4 Contrôle d'admission d'appel	88
6.6 Conclusion	89

Partie 3 : Allocation des ressources multimédia

Chapitre 7

Modèle de mobilité dans un réseau radio mobile multimédia	94
7.1 Introduction	94
7.2 Développement de modèle de mobilité	95
7.2.1 Les modèles graphiques	95
7.2.2 La définition du graphe de rues	96
7.2.3 Répartition des flux d'utilisateurs aux cellules	97
7.2.4 Flux de transition entre cellules	98
7.2.5 Taux de transition entre cellules	99
7.2.6 Taux de transition global de la cellule j vers les cellules voisines	100
7.2.7 Taux de handover	100
7.2.8 Charge de trafic	100
7.3 Modèle d'environnement considéré	101
7.4 Calibration de modèle de mobilité	102
7.5 Expérimentations et interprétations des résultats	104
7.6 Dimensionnement réseau et modèles de trafic	108
7.7 Conclusion	112

Chapitre 8

Gestion de QoS dans un réseau radio mobile cellulaire.....	115
8.1 Introduction	115
8.2 Performance réseau – paramètres de QoS	116
8.2.1 La probabilité de rejet d'appel handover	116
8.2.2 Le degré de service	117
8.2.3 Pré-réservation dynamique des canaux (DPCr)	118
8.3 Contrôle d'admission d'appel local.....	120
8.3.1 Application à notre approche	121
8.3.2 Gestion d'allocation des ressources locales	124
8.4 Conclusion	125

Chapitre 9

Algorithmes de consensus et approche consensuelle.....	129
9.1 Introduction	129
9.2 Les modules de consensus	130
9.2.1 Consensus centralisé asynchrone avec échéance de temps (ex. <i>DWD</i>)	130
9.2.2 Consensus distribué asynchrone (<i>Losange S</i>)	133
9.2.3 Consensus de type <i>frame work</i>	134
9.3 Phases d'exécution de services	137

9.3.1 Phase de négociation d'un média	137
9.3.2 Les caractéristiques d'un flux multimédia	138
9.4 Contrôle d'admission d'appel bout en bout	139
9.4.1 Table de routage	140
9.4.2 Problématique	141
9.4.3 Approche consensuelle	142
9.4.4 Discussion	143
9.4.5 Application de notre approche consensuelle	144
9.4.5.1 <i>Losange S</i>	144
9.4.5.2 <i>DWD</i>	145
9.5 Conclusion	146
Chapitre 10	
Gestion de QoS dans un réseau radio mobile de bout en bout.....	148
10.1 Introduction	148
10.2 CAC de bout en bout	148
10.2.1 Mise en place de CAC et approche consensuelle.....	149
10.2.2 Exemple de CAC de bout en bout.....	151
10.3 Mise en œuvre du consensus	152
10.3.1 Théorie de graphe	152
10.3.2 Exécution de l'algorithme de consensus	153
10.3.2.1 RM correct - pré-réservation des ressources immédiates	154
10.3.2.2 RM défaillant - négociation des ressources intra et inter-domaines	154
10.3.3 Expérimentation et interprétation des résultats	155
10.3.3.1 Votes successifs – chemins alternatifs – convergence de votes	155
10.3.3.2 Combinaison de consensus et de chaîne de Markov	157
10.4 Conclusion	165
Conclusion générale	168
Annexe A	175
Annexe B	177
Annexe C	178
Annexe D	180
Annexe E	182
Annexe F	184

Liste des figures

Figure 2.1	Cycle de vie d'un réseau cellulaire radio mobile	20
Figure 2.2	Processus de planification.....	21
Figure 2.3	Diversité de taille de cellule	23
Figure 2.4	Motif cellulaire.....	23
Figure 2.5	Motif cellulaire subdivisé.....	23
Figure 2.6	Architecture détaillée d'un réseau cellulaire radio mobile	24
Figure 2.7	Flux global de mobilité.....	26
Figure 2.8	Synoptique de dimensionnement de réseau.....	27
Figure 2.9	Topologie d'un réseau UMTS.....	29
Figure 2.10	Positionner les sites suivant une maille de type hexagonale	30
Figure 2.11	Modèle d'environnement considéré.....	33
Figure 3.1	Flux de données par une session de Web Browsing.....	41
Figure 3.2	Représentation du flux Web par des processus MMPP.....	43
Figure 4.1	Satisfaction entre la demande, le contrat et la disponibilité des ressources..	52
Figure 4.2	Impact entre les paramètres mesurés au niveau de service et réseau.....	53
Figure 4.3	Qualité de service de bout en bout.....	54
Figure 5.1	Mapping de QoS IP et QoS UMTS.....	62
Figure 5.2	Architecture de bout en bout.....	62
Figure 5.3	Gestion bout en bout de QoS.....	63
Figure 5.4	Délai de transfert de bout en bout sur UMTS/IP.....	64
Figure 5.5	Architecture générale d'un réseau à base de règles.....	65
Figure 5.6	Les modules softwar.....	66
Figure 5.7	Fonctionnalité du serveur de règle	67
Figure 5.8	Exemple de communication entre le PEP et le PDP.....	67
Figure 5.9	Exemple d'un PEP muni du client COPS.....	68
Figure 5.10	Exemple d'un PEP accessible par un consensus.....	69
Figure 5.11	Exemple de réseau InServ.....	70
Figure 5.12	Principe de base du MPLS.....	72
Figure 6.1	Processus d'allocation	84
Figure 6.2	Modèle Tjelta	86
Figure 6.3	Plans de contrôle/application	88
Figure 6.4	Réseaux radio mobiles/DiffServ.....	89
Figure 7.1	Les modèles graphiques pour la planification radio mobile.....	96
Figure 7.2	Graphe de rue dans une zone urbaine.....	96
Figure 7.3	Répartition des flux d'utilisateurs sur les différentes cellules couvrantes...	97
Figure 7.4	Répartition des flux d'utilisateurs entre cellules dans la rue.....	98
Figure 7.5	Le flux de transition de la cellule m vers la cellule n.....	99
Figure 7.6	Configuration 1.....	101
Figure 7.7	Configuration 2.....	101
Figure 7.8	Modèle d'environnement considéré.....	103
Figure 7.9	Modèle de mobilité.....	104
Figure 7.10	Organigramme de la procédure de planification.....	105
Figure 7.11	Flux d'utilisateur en fonction des longueurs des rues.....	106
Figure 7.12	Densité d'utilisateurs en fonction des vitesses.....	107
Figure 7.13	Flux de transitions entre cellules en fonction des vitesses.....	108
Figure 7.14	Taux de handovers en fonction des taux de transitions.....	110
Figure 7.15	Charge de trafic en fonction des taux de transitions.....	111
Figure 7.16	Résultats de simulations.....	112
Figure 8.1	Sortie de simulation – paramètres seuils d'acceptation d'appel.....	116
Figure 8.2	Probabilité de rejet d'appel handover en fonction de charge de réseau.....	117

Figure 8.3	Grade de service en fonction de taux de transition.....	118
Figure 8.4	Diagramme de transition.....	119
Figure 8.5	Réseau radio mobile – CAC.....	120
Figure 8.6	Données statiques récupérées du réseau cellulaire.....	121
Figure 8.7	CAC pour le trafic vidéo – comparaison bibliographie – simulation.....	124
Figure 9.1	Consensus <i>DWD</i>	131
Figure 9.2	Consensus <i>Losange S</i>	134
Figure 9.3	Consensus <i>Frame work</i>	136
Figure 9.4	Phases d'exécution de services le long d'une session.....	138
Figure 9.5	Ordonnancement des media dans la file d'attente.....	139
Figure 9.6	Session multimédia - négociation dynamique temps réel.....	139
Figure 9.7	CAC/SLA – pré-réservation des ressources inter et intra domaines.....	140
Figure 9.8	RM - vue globale au sein de son domaine et les domaines voisins.....	140
Figure 9.9	Plage commune décidée pour tous les RM.....	140
Figure 9.10	Niveaux différents de SLA au même média.....	141
Figure 9.11	Mise en place de consensus <i>Losange S</i>	143
Figure 10.1	Procédure CANAC(SLAI).....	149
Figure 10.2	Mise en place de CAC de bout en bout.....	150
Figure 10.3	Organigramme du CAC de bout en bout.....	150
Figure 10.4	Contrôle d'admission d'appel intra et inter- domaines.....	151
Figure 10.5	Exemple : modèle de réseau.....	152
Figure 10.6	Procédure consensus <i>Losange S</i>	154
Figure 10.7	Exécution du protocole de consensus-ressources disponibles.....	154
Figure 10.8	Exécution du protocole de consensus-ressources indisponibles.....	155
Figure 10.9	Convergence de consensus vers k domaines concernés par l'application... ..	156
Figure 10.10	Comparaison entre <i>Losange S</i> et <i>DWD</i>	157
Figure 10.11	Vue globale sur l'état de ressources au sein des RM corrects.....	157
Figure 10.12	Consensus-ressources à récupérer-recherche des chemins alternatifs.....	158
Figure 10.13	CAC-régulation des paramètres de QoS.....	158
Figure 10.14	Chaîne de Markov modélise un seul type de service.....	159
Figure 10.15	Chaîne de Markov modélise les services faible bandes.....	161
Figure 10.16	Chaîne de Markov modélise les services large bandes.....	161
Figure 10.17	Probabilité de blocage en fonction des taux d'arrivée d'appel.....	162
Figure 10.18	Probabilité de rejet en fonction des taux d'arrivée d'appel.....	163
Figure 10.19	Taux d'utilisation en fonction des taux d'arrivée d'appel.....	163
Figure 10.20	Probabilité de blocage en utilisant le consensus.....	164
Figure 10.21	Probabilité de rejet en fonction des taux d'arrivée d'appel.....	165

Liste des tableaux

Tableau 2.1	Différents environnements sont définis.....	30
Tableau 2.2	Segmentation des services proposés par l'ETSI.....	31
Tableau 3.1	Classes de services.....	37
Tableau 3.2	Classification des trames par contenu.....	40
Tableau 3.3	Temps d'inter-arrivée des datagrammes.....	42
Tableau 4.1	Mesures effectuées par l'ITU.....	50
Tableau 4.2	Classes de trafic par l'UMTS.....	51
Tableau 5.1	Structures d'un message COPS.....	68
Tableau 5.2	Les routeurs-paramètres de QoS.....	72
Tableau 7.1	Domaine de configuration.....	102
Tableau 7.2	Mesures réalisées dans le réseau.....	102
Tableau 7.3	Hypothèses considérées.....	102
Tableau 7.4	Modèles de trafic	109
Tableau 7.5	Résultats globaux de simulation	111
Tableau 8.1	Pré-réservation dynamique de canaux.....	120
Tableau 8.2	Données statiques et paramètres de QoS.....	122
Tableau 8.3	Paramètres de QoS-Bibliographie	123
Tableau 8.4	Paramètres seuils récupérés du réseau-résultats de simulation.....	125
Tableau 9.1	Les paramètres considérés-protocole <i>DWD</i>	131
Tableau 9.2	Les paramètres considérés-protocole <i>Losange S</i>	134
Tableau 9.3	Les paramètres considérés-protocole <i>frame work</i>	136
Tableau 9.4	Capacité allouée par les interfaces routeurs.....	141
Tableau 9.5	Comparaison des modèles de consensus	144
Tableau 10.1	Markov-pilotage d'accessibilité aux média	165

Chapitre 1

Introduction générale

1.1 Contexte de l'étude

L'évolution du secteur des télécommunications durant la dernière décennie a été marquée par une croissance phénoménale des services mobiles et multimédia concrétisée par l'étendue rapide des réseaux cellulaires mobiles de deuxième génération (GSM, IS95, etc.) d'une part et des réseaux Internet d'autre part. La croissance de ces deux types de services, mobile et multimédia, s'est produite de manière indépendante et à des cadences comparables ce qui favorise la convergence fixe-mobile.

Face à cette évolution rapide, les systèmes radio mobiles actuels (GSM, IS95, etc.) se sont révélés impuissants à répondre aux exigences des nouvelles applications vu qu'ils sont conçus pour satisfaire principalement la transmission des communications téléphoniques de qualité comparables à celle des réseaux fixe et de manière moins commune la transmission de données à bas débit (le débit ne peut pas dépasser 9,6 Kb/s dans le cas du GSM, par exemple).

Pour palier aux limitations des réseaux mobiles de deuxième génération, les constructeurs des systèmes radio mobiles ont opté au début vers l'amélioration des performances de ces réseaux par l'élaboration des techniques d'allocation des ressources plus flexibles qui peuvent se greffer sur les réseaux existants. Tel est le cas du réseau GPRS qui permet un débit utile pouvant atteindre 114 Kb/s et qui peut être facilement implanté sur un réseau GSM opérationnel. Cependant, ces solutions « rapides » ne peuvent constituer qu'une phase transitoire vu que la bande allouée aux

réseaux mobiles de deuxième génération ainsi que les techniques d'allocation des ressources ne permettent pas de supporter les applications haut débit.

La solution radicale qui a été adoptée par la communauté des télécommunications à l'échelle internationale consiste à développer de nouveaux systèmes radio mobiles sur la base d'utilisation d'une bande de fréquence plus large que celle allouée aux réseaux mobiles de deuxième génération [NAP98].

La définition de cette nouvelle génération de système mobile, présentée sous l'acronyme IMT2000, met en valeur les atouts suivants au niveau des services :

- Systèmes multiservices : ils peuvent assurer de façon générale la transmission des applications multimédia à des débits relativement élevés (jusqu'à 2 Mb/s pour le cas de l'UMTS). Il devient possible d'offrir à l'utilisateur mobile des différentes classes de services de contenu (voix, données, etc.).
- Techniques d'accès adaptées à la transmission à débit variable : les systèmes IMT2000 utilisent des techniques d'accès multiple dérivées de la technique d'étalement de spectre à large bande. Ces techniques basées sur l'utilisation des codes pour la définition des canaux permettent plus de flexibilité dans la gestion des débits ainsi que la possibilité de la répartition d'un canal physique entre plusieurs applications.
- Gestion avancée de la mobilité : des mécanismes de macro diversités et de soft handover sont considérés. Ils permettent d'améliorer la qualité de communication.

Avec l'ensemble de ces caractéristiques, les systèmes IMT2000 [IMC02] permettent d'offrir une capacité nettement plus meilleure que les systèmes mobiles de deuxième génération, et une procédure de gestion de ressources qui sera nécessaire.

1.2 Problématique de la thèse

La disponibilité des ressources radio est tributaire de la fluctuation de la charge du réseau, de la mobilité des utilisateurs et de la variation de qualité des canaux radio. Ce problème est particulièrement délicat pour les services multimédia car les facteurs qui le conditionnent deviennent un problème majeur pour l'opérateur réseau radio pour la gestion de l'allocation de ses ressources conformément à ses engagements sur la qualité de service.

Aussi bien, l'augmentation des utilisateurs et de leurs besoins de services multimédia mobiles de qualité pose des problèmes de limitation de la bande passante aux opérateurs radio mobiles. En effet, une mauvaise distribution des ressources augmente les rejets d'appels handovers et une mauvaise répartition de la bande passante entre utilisateurs qui ne tient pas compte des facteurs liés aux attentes en terme de qualité de service.

Les utilisateurs souhaitent que les opérateurs s'engagent sur la qualité de service. Or d'après la définition de la qualité de service selon la recommandation de l'IUT-T N°E-800 [TERM95], il est nécessaire que le service délivré pour un utilisateur UMTS ne présente pas des dégradations au-delà de ce qui est autorisé par le contrat de service.

Ces dégradations peuvent surgir lorsque, par exemple la bande passante sur le réseau dédiée au service diminue ou encore lorsque le retard causé par le réseau de transport augmente.

Les opérateurs sont alors confrontés à la gestion de l'allocation des ressources réseau. Les contraintes à intégrer portent notamment sur la mobilité des utilisateurs, les différentes caractéristiques du trafic et les paramètres de qualité de service. Ces contraintes engendrent une complexité d'estimation objective de la qualité de service pour des services multimédia mobiles.

Dans le contexte général de planification des réseaux mobiles de troisième génération et en tenant compte de la mobilité et de la multiplicité des services, l'étude du comportement d'un tel réseau s'avère nécessaire pour compenser une charge de trafic hétérogène du point de vue des caractéristiques des flux d'une part et la qualité de service d'autre part. Le choix d'une politique d'allocation des ressources et de gestion de la qualité de service pour un type de service donné doit faire l'objet d'étude d'impact sur les performances globales du système en terme de débit offert pour chaque type de service d'une part et en terme de respect des contraintes de qualité de service d'autre part.

En conclusion, la bonne gestion du réseau de bout en bout reste encore un problème car il est difficile de réserver en temps-réel les ressources au niveau des équipements réseaux de plusieurs domaines de façon à répondre de manière optimale à la demande des utilisateurs.

L'ensemble de ces raisons nous a amené à nous intéresser principalement dans le cadre de nos travaux de thèse, à l'allocation des ressources selon les critères de qualité de service par rapport aux paramètres des ressources disponibles soit au niveau réseau radio cellulaire ou du réseau cœur DiffServ.

1.3 Contribution de la thèse

Pour répondre à ces objectifs, ce travail de thèse porte sur la gestion de la qualité de service dans un réseau radio mobile multimédia. Un des aspects majeur que nous avons considéré lors de cette étude est la mobilité des utilisateurs. Celle-ci pose deux problèmes : tout d'abord au niveau du trafic dont la répartition fluctue au gré des mouvements des usagers dans le réseau et secondairement au niveau de la signalisation engendrée par les usagers du fait de leurs déplacements. Les études actuelles portant sur le trafic usager ou de signalisation sont basées sur les modèles de mobilité et de trafic simples et de type macroscopique (c'est-à-dire à l'échelle de quelques kilomètres). Nous nous intéressons dans le cadre de ce travail à l'élaboration d'un modèle de mobilité dans quatre modèles de trafic retenus : voix, vidéo-téléphonie, ftp et www. Le modèle obtenu est calibré grâce à des mesures effectuées sur un réseau cellulaire opérationnel.

Les modèles de mobilité et de trafic calibrés nous ont permis d'évaluer les différents paramètres de performance tels que la charge de trafic, les taux de handovers, le niveau de service (GoS : *Grade of Service*), etc. Ces paramètres de performance sont représentés par la suite à l'aide des formules génériques qui peuvent être appliquées à d'autres zones de services semblables. Il suffit donc de calibrer ces formules génériques

en fonction des paramètres suivants : la période de mesure, le rayon de zone de services, la distance entre les stations de base, etc.

Le développement d'un modèle de mobilité sera développé au sein de la troisième partie de la thèse. Ce modèle nous permettra par la suite de gérer l'allocation des ressources avec une qualité de service optimale. Il englobe des paramètres, de nature démographiques, urbains, etc. L'élaboration de ce modèle permet donc une analyse approfondie et complète qui permet, entre autre, d'estimer le trafic moyen par utilisateur pour le dimensionnement des équipements des stations de base. La connaissance de la répartition des flux de trafic permet ensuite de dimensionner les différentes connexions du système et de gérer les ressources disponibles de la façon la plus judicieuse possible.

Une méthodologie de planification cellulaire est ensuite définie qui se base sur les modèles de trafic et de mobilité. Pour cela, une zone urbaine réelle est définie comportant des sites potentiels auxquels des poids sont affectés. Différents types de trafic présentant chacun des caractéristiques de qualité de service différentes sont pris en compte car il s'avère nécessaire que la gestion de la qualité de service soit maîtrisée, prévisible et contrôlable puisqu'elle est inhérente à la fois aux services multimédias et au réseau.

Les protocoles de gestion de la qualité de service prendront en entrée les modèles de mobilité et de trafic ainsi que les paramètres d'ingénierie radio tels que le taux de handovers, la charge de trafic, le grade de service, etc., permettant par la suite d'agir sur l'allocation des ressources dans le réseau radio mobile UMTS/réseau cœur.

Pour traiter le problème d'allocation des ressources radio de bout en bout, des mécanismes de contrôle d'admission d'appel (CAC) sont considérés. Nous considérons un gestionnaire de ressources implanté dans un routeur de bordure, à l'entrée du réseau radio opérationnel, au niveau du SLA (*Service Level Agreement*). Ce gestionnaire de ressources communique en premier temps avec les stations de base pour décider de l'acceptation des demandes d'appels caractérisées par les valeurs des paramètres de mobilité et de qualité de service, et dans un deuxième temps et si l'appel est accepté, il lance alors une demande d'acceptation d'allocation de ressources aux autres réseaux participant à l'application.

1.4 Plan de la thèse

Ce travail se structure en 3 parties différentes :

La première partie concerne la mise en place de notre environnement de travail. Cette partie est composée des chapitres 2, 3, 4 et 5.

Dans le chapitre 2, nous présentons le processus de mise en place d'un réseau cellulaire qui nécessite la mise en œuvre d'une procédure de planification optimale, passant par une étude approfondie de tous les éléments d'un réseau, permettant de couvrir la demande et offrant une qualité de service spécifiée à l'avance. L'opérateur de télécommunication dispose d'un environnement à couvrir, d'un nombre estimé d'abonnés potentiels, et d'une bande de fréquence. Il s'agit alors de définir la taille de cellules, le type de trafic et la mobilité des utilisateurs.

Dans le chapitre 3, nous nous sommes intéressés à l'identification des modèles de trafic pour les quatre classes de service (*Conversational*, *Streaming*, *Interactive* et *Background class*) [NAJ01][AJI00][W3C][DAW99] en considérant une application type pour chaque classe. Afin de définir les scénarios à étudier dans les chapitres 7 et 8, nous avons défini les caractéristiques des quatre modèles de trafic : voix, vidéo-téléphonie, ftp et www dans la deuxième partie de ce chapitre.

Dans le chapitre 4, nous présentons les paramètres de qualité de service considérés dans le processus de dimensionnement qui sont la probabilité d'échecs de handover (la probabilité de coupure de communication en cours) et la probabilité de refus d'établissement d'une nouvelle phase d'appel. Une amélioration des valeurs de ces deux indices de performances témoigne d'un trafic écoulé plus important. Le respect d'un niveau de qualité de service pour un volume de trafic donné suppose un dimensionnement optimal du réseau.

Le chapitre 5, illustre les principales techniques proposées pour intégrer la qualité de service dans les réseaux UMTS/IP.

Ce chapitre présente aussi l'architecture générale d'un réseau à base des règles adoptées pour un contrôle de qualité de service de bout en bout. La satisfaction de l'utilisateur ne dépend que des paramètres objectifs (i.e. la QoS résultante de bout en bout). Afin de répondre à ses attentes, des recommandations ont été formulés par l'IETF dans le cadre du groupe IntServ/DiffServ pour le contrôle et la gestion de l'allocation des ressources dans le réseau.

La deuxième partie de notre travail, débute par un état de l'art de la gestion de la qualité de service dans un réseau radio mobile aussi bien du réseau cellulaire que du réseau cœur. Cette partie est composée du chapitre 6.

Dans ce chapitre, nous donnons un aperçu sur les travaux en cours sur la mobilité et l'allocation des ressources pour les services de télécommunication.

Dans la troisième et dernière partie de cette étude, nous proposons une architecture de pré-réservation des ressources cellulaire/DiffServ basée sur des politiques que nous développons. Cette partie est consacrée à l'allocation des ressources de bout en bout pour les services télécom en UMTS. C'est dans cette partie que nous représentons également nos politiques de bout en bout niveau réseau cellulaire/cœur et les améliorations obtenues par rapport aux systèmes classiques.

Le chapitre 7 présente le développement d'un modèle de mobilité pour les réseaux cellulaires multimédia. Pour cela, une zone urbaine est définie, elle sera considérée comme étant la zone de service, comportant des sites potentiels auxquels des pondérations sont affectées. Cette zone est décrite par des paramètres avec lesquels le modèle de mobilité est défini. Ce modèle de mobilité se base principalement sur l'étude des flux d'utilisateurs dans les axes routiers qui nous permet de déterminer en premier lieu, le taux de transition/utilisateur, et ce entre les différentes cellules du système. Ces taux donnent par la suite la probabilité qu'un utilisateur dans une cellule transite vers une autre cellule, ainsi que les taux de handovers inter-cellulaire.

Dans ce chapitre, nous avons considéré la mobilité des usagers pour le dimensionnement du réseau pour les quatre modèles de trafic étudiés dans le chapitre 3: voix, vidéo-téléphonie, www et ftp. Les modèles de trafic considérés ne distinguent pas les différents mobiles et leurs mouvements mais abordent la mobilité dans le système d'un point de vue macroscopique. Ce modèle de mobilité nous permet de déterminer au niveau d'une cellule aussi bien qu'au niveau du système entier, les paramètres seuils, au-delà desquels un appel peut être refusé : le taux de handover, la probabilité de rejet d'appel, le trafic offert, les niveaux de service, etc.

Nous nous intéressons dans le chapitre 8 à l'allocation locale des ressources en considérant essentiellement la mobilité des utilisateurs. Dans une cellule, l'allocation des ressources aux différents types de trafic se fait uniquement lors de l'occurrence d'un nouvel appel ou d'un appel handover.

Nous proposons dans le chapitre 8 une approche de gestion de l'allocation locale des ressources basée principalement sur des paramètres de qualité de service et un mécanisme de contrôle d'admission d'appel. L'objectif est alors d'augmenter le niveau d'acceptation des appels au sein du domaine tout en respectant les contraintes de qualité de service. Ce mécanisme permet d'adapter le comportement du trafic aux caractéristiques du réseau vis à vis des paramètres de qualité de service tels que : le taux de handovers, la charge de trafic, la probabilité de rejet d'appel, le grade de service (GoS), etc.

Dans notre approche, le mécanisme de contrôle d'admission d'appel se fait à deux niveaux : premièrement au niveau de CANAC (Capacité de Nouveaux Appels par Cellule) qui peut être enregistrée au niveau du SLA radio mobile, là où la mobilité d'utilisateur est considérée comme étant une contrainte principale pour l'acceptation d'un appel. La seconde étape a lieu au niveau du réseau cœur de l'opérateur (cf. chapitres 9 et 10). L'acceptation de la qualité de service définie par le contrat SLA spécifiant les paramètres des différents niveaux d'acceptation se base sur une décision collégiale et consensuelle entre les différents réseaux participants.

Dans les chapitres 9 et 10, le mécanisme de contrôle d'admission d'appel est basé sur un algorithme de consensus. Ce consensus s'applique au niveau de la couche contrôle/application pour permettre la communication entre domaines. Il est coordonné par un coordinateur tournant. Ce coordinateur connaît l'état de disponibilité des ressources de la zone où il est implanté (cf. chapitres 7 et 8). Pour mettre en place un service reposant sur une classe donnée (cf. chapitre 3), une requête est envoyée vers cet agent qui peut soit l'accepter soit la refuser en fonction de l'authentification de la requête et des ressources réseau disponibles.

En guise de conclusion, nous récapitulons les principales contributions de cette thèse et les résultats obtenus. Nous résumons notre travail qui a consisté à élaborer un modèle de mobilité des utilisateurs dans un réseau radio mobile multimédia qui nous a permis de concevoir une politique ainsi qu'un protocole de consensus pour l'allocation des ressources de bout en bout des services télécoms pour l'UMTS. Nous exposons également nos futurs axes de recherche et les perspectives de notre travail.

Références Bibliographiques

- [AHM03]Ahmed, T. and al., “*QoS dynamic management of video flow in IP DiffServ architector*,” Proc. GRES’03, 2003, pp.53-65.
- [AJI00] Ajib, WW., “*Transmission management of data temporary flow in a radio mobile access TDMA*,” PhD, ENST Paris, 2000.
- [CHE02]Cheng, Y. and Zhuang, W., “*DiffServ resources allocation for fast handoff in wireless mobile Internet*,” IEEE Communications Magazine May 2002, pp.130-136.
- [CHA98]Chahdra, P. and al. “*Network support for application oriented QoS*,” Proc. of Sixth International Workshop on Quality of Service (IWQoS’98), pp.187-195.
- [CEN95]Cengiz, C.;*GSM to UMTS/ A perspective on the future*; ICT’95, April 1995, pp. 351-358.
- [DAW99]Dawood, A.M. and Ghanbari, M., “*Content-based MPEG video traffic modelling*,” IEEE Transactions on multimedia, Vol.1, N°.1, March 1999.
- [IETF] <http://www.ietf.org>.
- [IMC02]3rd Generation Partnership Project, “*Digital cellular telecommunication system (Phase 2+) (GSM)*,” Universal Mobile Telecommunication System (UMTS); IP Multimedia (IM) session handling; IM call model; Stage 2 (3GPP TS 23.218 version 5.0.0 Release 5-2002-03).
- [NAP98]Napolitano, A. and Pannioli, F., “*Evolution of the GSM platform*,” IEEE International Conference on Universal Communications (ICUPC’98), Florence, Italy, October 1998.
- [NAJ01]Naja, R. and Thome, S., “*Multi-services call admission control and handover issues in wireless multimedia network*,” MS3G’2001, Lyon, France, December 2001.
- [QoS99]3rd Generation Partnership Project, “*Technical Specification Group Services and System Aspects*,” *QoS Concept and Architecture*; Release 1999.
- [SHE97]Shenker, S. and al., “*Specification of guaranteed quality of service*,” IETF-RFC 2212, September 1997.
- [TRA02]Trang Nguyen, T.M. and al., “*COPS-SLS: a service level negotiation protocol for the Internet*,” IEEE Communications Magazine, May 2002.
- [TER95]“*Terms and definitions related to quality of service and network performance including dependability*,” IUT-T Recommendation E-800.
- [W3C] World Wide Web Consortium W3C, <http://www.w3.org/>.
- [WRO97]Wroclawski, J., “*Specification of guaranteed load network element service*,” IETF-RFC 2211, September 1997.

Partie 1 : Contexte

Chapitre 2

Planification d'un réseau radio mobile multimédia**2.1 Introduction**

L'objectif de ce chapitre consiste à définir un réseau cellulaire qui minimise le coût de l'infrastructure radio et réseau en fonction de la couverture radio, de la taille de cellules et qui respecte des contraintes de qualité de service (taux de blocage, taux d'interruption, etc.) [TAB97]. Pour ce faire, il s'agit de considérer les caractéristiques de l'environnement à couvrir (une ville est considérée comme la zone de service constituée des sous-zones de services, réseau routier, etc.), les caractéristiques des utilisateurs à desservir (densité, comportement, etc.) et une bande de fréquence.

Dans le cas des systèmes radio mobile de 3^{ème} génération (3G), le nombre de paramètres mesurés sur le réseau en exploitation (caractéristiques géographiques, qualité de couvertures, etc.), ou le nombre des paramètres envisagés pour les services multimédia disponibles à ajuster sont plus importants que ceux du cas des systèmes de 2^{ème} génération (2G) [TAB97], en raison de l'évolution de la nature des services et la croissance du nombre d'abonnés et de services. Par conséquent, l'opérateur doit être constamment à l'écoute du changement pour déterminer les valeurs optimales des paramètres de dimensionnement qui doivent être ajustés pour optimiser l'utilisation des ressources radio. Ainsi, un système, dont les paramètres sont optimisés convenablement, offrira des communications de bonne qualité et aura une capacité importante.

Nous allons commencer dans ce chapitre par présenter les différents processus de planification des réseaux cellulaires de 3^{ème} génération. En effet, l'installation d'un réseau cellulaire nécessite la mise en œuvre d'une procédure de planification efficace, passant par une étude approfondie de tous les éléments entrants en vigueur pour la réalisation d'un réseau, permettant de couvrir la demande et offrant une qualité de service déterminée. Nous consacrons la dernière partie de ce chapitre pour la description de la mise en place d'un réseau UMTS du point de vue topologie, dimensionnement en trafic, environnements considérés, etc.

Le but de ce chapitre est de présenter le réseau radio cellulaire UMTS, ses spécificités, ses caractéristiques et son infrastructure. Ceci est important pour pouvoir mettre en œuvre notre modèle de mobilité développé au niveau du chapitre 7. Ce modèle de mobilité nous permet de déterminer au niveau d'une cellule ainsi qu'au niveau du système entier, les paramètres seuils, au-delà desquels un appel peut être refusé ou non. Ainsi, nous pouvons déterminer les limites de réservation de ressources tels que par exemple, le taux de handover, la probabilité de rejet d'appel, le trafic offert, les niveaux de service, etc. Ces paramètres sont utilisés par la suite pour paramétrer notre modèle d'allocation de ressources de bout en bout développé principalement au niveau des chapitres 8, 9 et 10.

2.2 Processus de planification des systèmes radio mobile de 3^{ème} génération

Dans le cycle de vie de la mise en place d'un réseau radio mobile de 3^{ème} génération ((1), (2), (3), (4), (5), (1), etc.) (voir figure 2.1), il s'agit en première étape de réaliser la partie analyse, dimensionnement, planification et paramétrage qui donne lieu aux modèles de trafic et de mobilité. Dans une deuxième étape, il s'agit de concevoir un modèle de gestion de la qualité de service qui se base à la fois sur un module de contrôle des paramètres estimés pour les services multimédia envisagés et sur un module qui se renseigne sur les ressources disponibles dans le réseau. En effet, ce modèle de gestion de QoS permet, en ligne, de gérer la QoS et, hors ligne, de constituer une base d'information pour re-paramétrer et redimensionner les modèles de trafic et de mobilité.

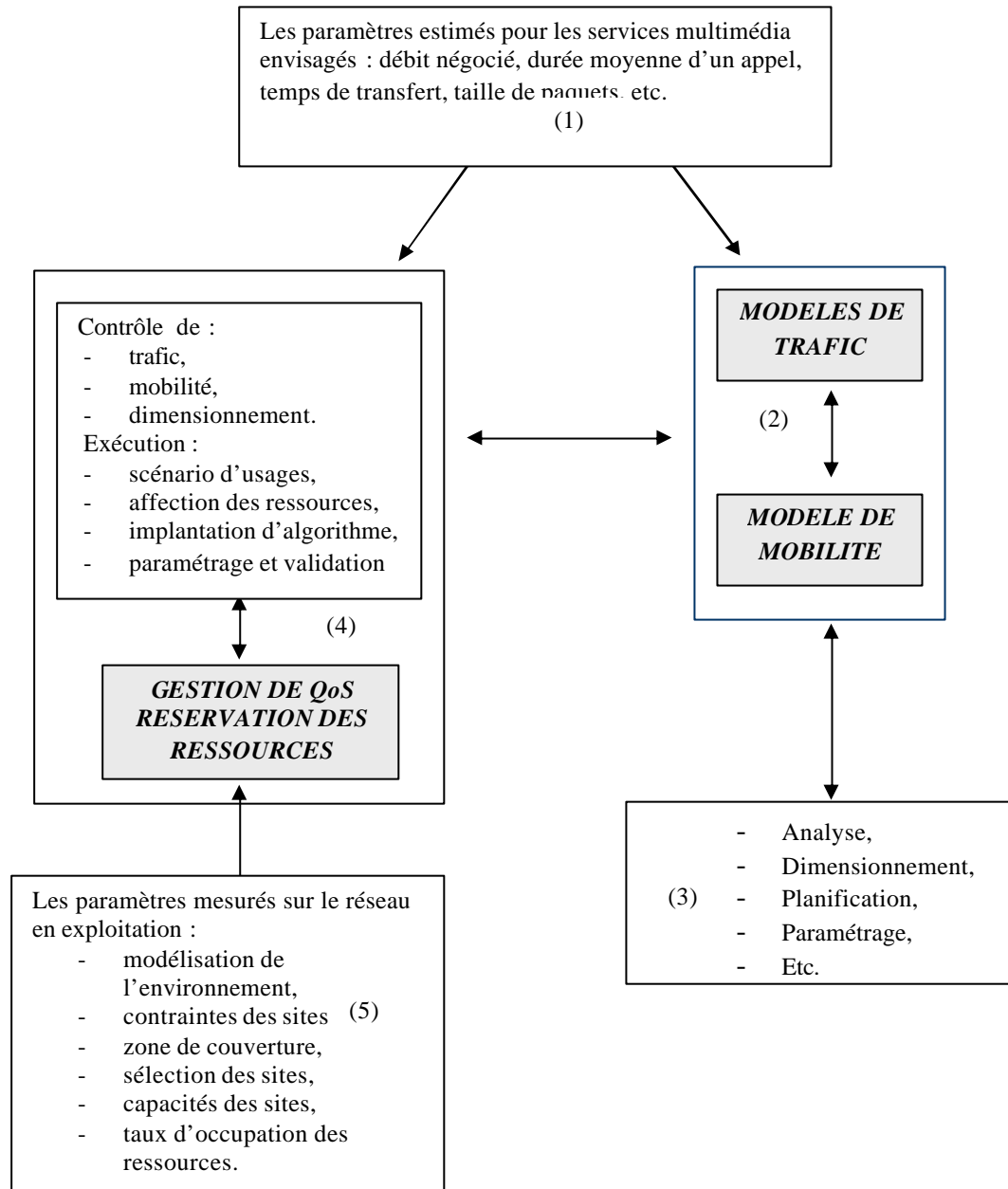


Figure 2.1 Cycle de vie d'un réseau cellulaire radio mobile

La mise en œuvre d'un réseau cellulaire nécessite la mise en œuvre d'une procédure de planification réseau efficace.

Le processus de planification peut être résumé par la figure suivante (voir figure 2.2) :

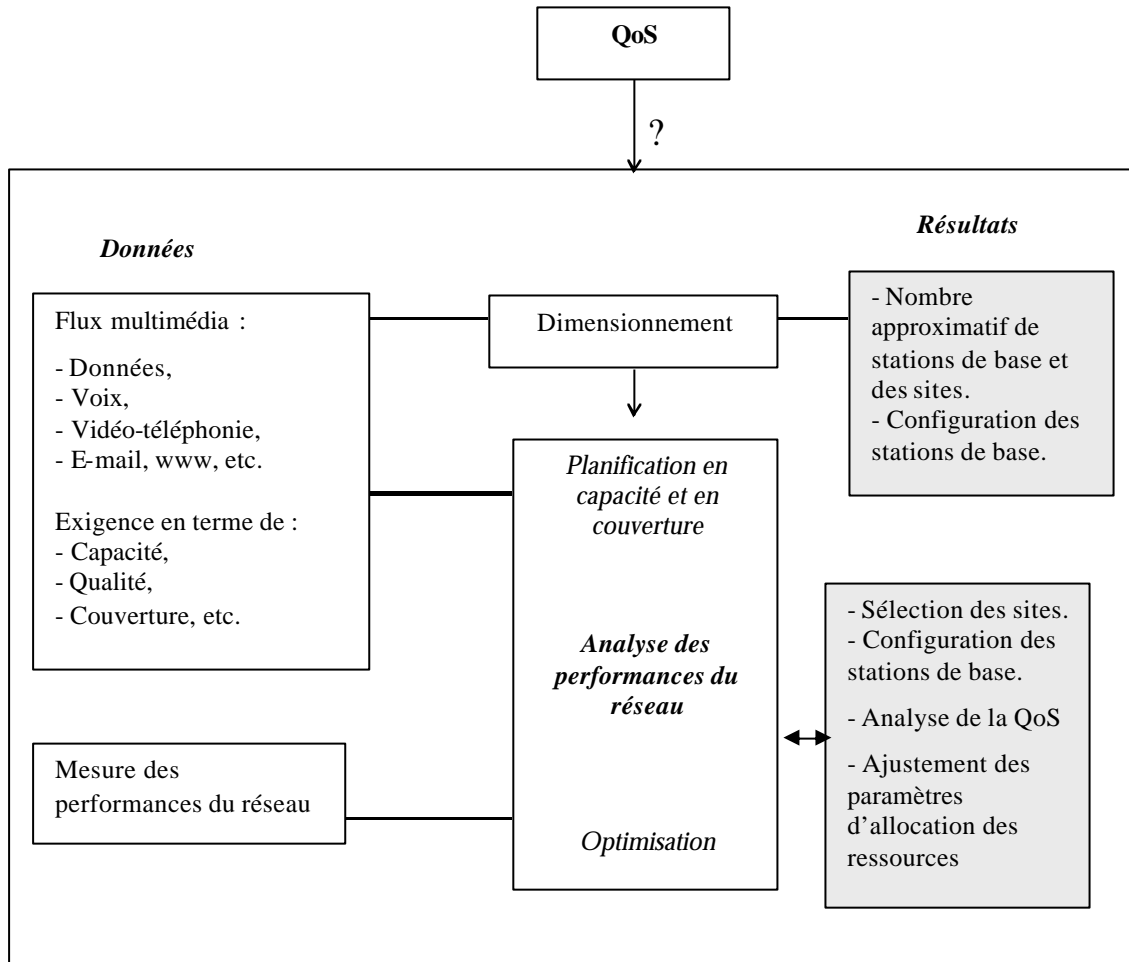


Figure 2.2 Processus de planification

Il s'agit d'un processus de planification interactif permettant le réajustement des différents paramètres de dimensionnement au cours de la planification. En effet, les différentes étapes présentées ci-dessus sont étroitement liées. Elles doivent conduire en dernier lieu à la détermination des paramètres suivants :

- Plan et nombre de stations de base ;
- Plan des équipements du réseau fixe ;
- Réseau de connexion entre les différentes entités de système ;
- Charge du réseau en trafic ;
- Taux de handovers entre cellules ;
- Valeurs maximales tolérées (les seuils) pour les critères de la QoS (probabilités de blocage des appels nouveaux, probabilités de rejet des appels handovers, etc.)

2.2.1 Objectifs dans la conception d'un réseau cellulaire

- *Grande capacité d'utilisateurs* : le système doit être capable de supporter les communications multimédia de plusieurs milliers d'utilisateurs de téléphone mobile dans la zone de service avec un nombre bien déterminé de canaux.

- *Utilisation du spectre* : l'utilisation multiple de canaux identiques dans des cellules géographiquement séparées garantissant une optimisation du spectre radio électrique.
- *Compatibilité sur l'ensemble du réseau pour un même système* : les utilisateurs doivent pouvoir utiliser leurs équipements hors de leur zone d'habitation et également dans des réseaux cellulaires différents. Un mobile est identifié alors comme en situation de *roaming*.
- *Services vers les véhicules et autres services spécialisés* : un réseau cellulaire doit offrir aux utilisateurs de terminaux cellulaires, les mêmes services lors de leur utilisation dans des véhicules. Des services spécialisés du type appel de flotte, envoi de messages ou numérisation abrégée seront disponibles.
- *Adaptation à la densité de trafic* : le trafic est différent d'un point de la zone de couverture cellulaire à un autre. La prise en charge de l'évolution du trafic géographique doit être prise en compte lors de la conception du réseau.
- *Qualité de service* : le service multimédia doit être de qualité égale à celle des réseaux fixes. En terme économique, l'accès à ce type de service doit être considéré pour un large public.

2.2.2 Types de cellules considérées

Les premiers réseaux radio mobiles se composaient de quelques émetteurs qui couvraient des zones importantes et qui n'étaient pas interconnectés entre eux. Le nombre des canaux disponibles dans ces systèmes était très limité et des communications étaient interrompus dès que l'utilisateur quittait la zone de service d'un émetteur. Un concept cellulaire doit intervenir afin d'offrir une capacité importante et fournir des services divers à une échelle nationale ou même internationale.

Pour réaliser une couverture adaptée à chaque population et à chaque environnement (voir figure 2.3), trois types de cellules sont définis :

- Les macro-cellules : leur rayon est supérieur à 1 km. Elles sont utilisées pour desservir les usagers rapides disposant des mobiles de puissance comprise entre 1 et 10 watts.
- Les micro-cellules : elles sont couvertes par des petites stations de base situées dans des rues. Le rayon des micro-cellules est compris entre 100 m et 1 km. Elles desservent principalement les usagers piétons et les véhicules lents.
- Les pico-cellules : elles sont utilisées pour couvrir l'intérieur des bâtiments. Le rayon de ces cellules est inférieur à 100 m.

La taille d'une cellule dépend de l'environnement physique [VAL99] (influant sur les conditions de propagation) et de la puissance.

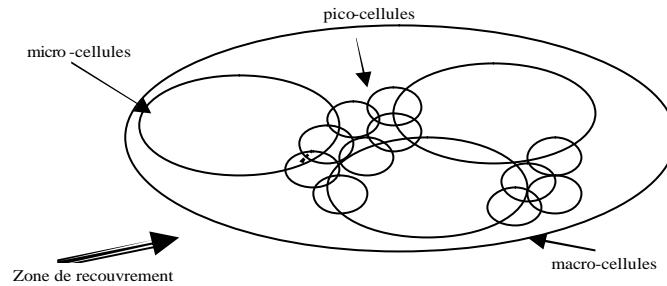


Figure 2.3 Diversité de taille de cellule

La raison d'être des cellules d'un réseau mobile est la définition d'une zone de couverture dans laquelle des canaux dédiés ou des implantations de stations de base peuvent être choisis sous une forme non exclusive. Dès que l'on a réalisé le motif cellulaire adéquat, lequel prend en compte l'espace minimum entre cellule pour éviter les interférences co-canal, un bon déploiement sera réalisé en choisissant les sites d'implantation des stations de base et les équipements nécessaires pour faire écouler le trafic estimé (voir figure 2.4).

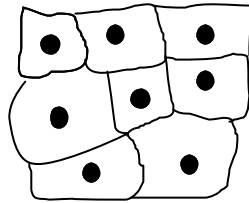


Figure 2.4 Motif cellulaire

Dès que le trafic augmente au point de perturber la QoS (atteint le taux de blocage), les cellules sont alors implantées dans des cellules de taille inférieure (voir figure 2.5).

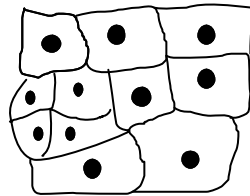


Figure 2.5 Motif cellulaire subdivisé

La cellule d'origine a été remplacée par exemple par quatre cellules de dimension inférieure, permettant un accroissement du trafic dans et aux frontières de la cellule. Cette opération peut être renouvelée sur les autres cellules par la méthode de réutilisation des canaux.

2.2.3 Architecture d'un réseau radio mobile cellulaire

Dans une architecture de réseau radio mobile cellulaire, on peut noter l'existence de deux sous-ensembles interconnectés (voir figure 2.6) [VAL99][TABB97] :

- Le sous-système radio (*Base Station System :BSS*), en charge des transmissions et de la gestion de la ressource radio. Ce sous-réseau comprend les BTS (*Base Transceiver Station*) et les BSC (*Base Station Controller*).

- Le sous-système d'acheminement ou partie fixe (*Network Sub-System : NSS*), prend en charge l'établissement des appels et les fonctions de gestion de la mobilité. Il est constitué par des équipements : MSC, VLR, HLR, etc.

2.2.3.1 Le sous système radio

La zone de couverture d'une station de base définit la notion de cellules qui peuvent être regroupées en *cluster*, recouvrant alors la région à desservir. Chaque cellule est desservie par une station de base (BTS) prenant en charge la signalisation et le bon déroulement des transmissions radio. La qualité de la couverture cellulaire et la capacité du réseau dépendent alors du placement de ces stations de base. Cela nécessite de tirer profit de la connaissance de l'environnement physique immédiat (rues, bâtiments, forêts, plaines, etc.). Il apparaît donc que cette phase de planification est fondamentale et intéresse en tout premier lieu les opérateurs qui concentrent leurs efforts sur le placement des émetteurs. Les BTS sont alors raccordées à des contrôleurs de station de base (BSC) ayant pour fonction principale de gérer la ressource radio (allocation des canaux, prise en charge du handover, etc.). C'est à ce niveau que s'arrête le sous-système radio.

2.2.3.2 Le sous système réseau

Trois éléments interviennent ensuite dans la partie fixe du réseau (MSC, VLR et HLR) (voir figure 2.6). Les centres de commutation du service mobile (*Mobile service Switching Center ; MSC*) assurent l'interconnexion des stations de base et l'interconnexion avec le réseau de télécommunication fixe classique. Il participe également à la transmission des communications de et vers un mobile. De plus, c'est à ce niveau que l'exécution de handovers est gérée.

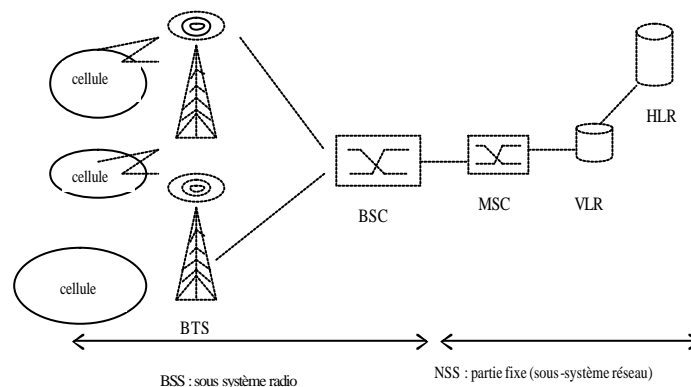


Figure 2.6 Architecture détaillée d'un réseau cellulaire radio mobile

2.2.4 Les mécanismes spécifiques des réseaux mobiles

Après avoir défini l'architecture et le concept cellulaire, nous nous intéressons maintenant à l'étude des mécanismes propres aux réseaux mobiles : la gestion du transfert inter-cellulaire (*handover*) et l'itinérance (*roaming*) [VOI99][LAG97].

2.2.4.1 Transfert inter-cellulaire (handover)

Le premier objectif des réseaux cellulaires est de fournir aux utilisateurs la possibilité de se déplacer librement au sein d'une zone de couverture. Au cours de la communication, le mobile peut être amené à changer de cellule : le système prend en charge, et de la façon la plus transparente possible pour l'utilisateur, le transfert intercellulaire de l'appel en cours : c'est le mécanisme de handover. Il est clair que la procédure de handover est une fonction critique d'un réseau mobile puisqu'elle doit assurer la continuité d'un appel en cours et fournir de la QoS aux usagers. Deux types de handovers existent : d'une part, le handover intra-cellulaire pour un changement de ressources radio au sein d'une même cellule et d'autre part, le handover inter-cellulaire pour les changements des ressources radio entre deux cellules voisines. Notons que, plus les cellules sont petites, pour répondre à une demande de fort trafic, plus le nombre de cellules traversées augmente, et donc la probabilité d'avoir un échec de handover (*Dropping Handover*) augmente également. Ainsi, plus la vitesse des mobiles est importante, plus la probabilité d'échec du handover est forte. Cette remarque justifie l'utilisation de grandes cellules (macro-cellules) pour couvrir des routes.

Techniquement, un mobile mesure périodiquement la puissance du signal reçu. Lorsqu'une forte diminution de la puissance est constatée, un mécanisme de handover est déclenché vers une des cellules d'où proviennent le signal le plus puissant dans le système GSM : un changement de fréquence est alors effectué. Ainsi, un handover est déclenché lorsqu'un utilisateur change de cellule au cours de son déplacement. Cependant un handover peut être exécuté pour deux autres raisons : forte densité de trafic et mauvaise qualité de lien radio. Enfin, il existe trois types de handovers qui dépendent du moment où le lien avec la cellule d'origine est libérée : avant la mise en place du lien avec la cellule cible (*Hard Handover*), pendant cette mise en place (*Seamless Handover*), ou encore, après cette mise en place (*Soft Handover*).

2.2.4.2 Gestion de l'itinérance (roaming)

Le rôle principal de ce mécanisme est de connaître à tout instant la position des mobiles dans le réseau. Cette fonction est fondamentale car sans sa mise en œuvre, il ne serait pas possible pour le système d'établir une communication. Nous rencontrons principalement deux méthodes de gestion de l'itinérance dont l'un des objectifs à atteindre est de minimiser le coût de ces méthodes de recherche car elles sont « gourmandes » en consommation de ressources.

2.2.4.2.1 Systèmes sans localisation

Ce système fut utilisé dans quelques réseaux mobiles de première génération. En fait, il n'existait pas de gestion de l'itinérance. Lorsqu'un abonné était recherché, des messages d'avis de recherches étaient diffusés sur toute la zone de couverture. Ce système était donc très simple mais ne supportait pas de forte charge d'appel.

2.2.4.2.2 Systèmes utilisant des zones de localisation

La zone de couverture est divisée en zone de localisation regroupant plusieurs cellules. Le système a la connaissance précise de la dernière zone de localisation visitée par le mobile. Ainsi, lors d'une recherche d'un abonné, le système va émettre dans toutes les cellules de la zone de localisation des messages d'avis de recherche. Le coût

de recherche et de consommation des ressources est alors réduit. Ce système est utilisé notamment dans le GSM, où les informations sont stockées dans des bases de données de type VLR/HLR. Enfin, puisque le réseau doit connaître en performance la zone de localisation du mobile, des fonctions de mise à jour sont définies. Nous rencontrons tout d'abord les mécanismes de mise à jour manuelle où les utilisateurs se déclarent au réseau, et la mise à jour automatique et périodique où le terminal de l'utilisateur envoie régulièrement au réseau un signal permettant d'identifier sa zone de localisation.

2.2.5 La mobilité des utilisateurs

L'évolution des systèmes cellulaires vers les systèmes de 3^{ème} est marquée par une croissance du nombre d'abonnés et de services. Cette croissance entraîne une augmentation du trafic dans les réseaux qui rencontrent de plus en plus souvent des problèmes de congestion de trafic. Une solution pour résoudre ces problèmes est la réduction de la taille des cellules. Les réseaux évoluent dans cette direction et migrent d'architectures macro-cellulaires vers des architectures micro et pico-cellulaires. Cette réduction de la taille des cellules entraîne une augmentation importante du nombre de *handover/roaming* inter et intracellulaire dans les systèmes. Le flux global de mobilité est donc égal à la somme des appels entrants avec les handovers, les appels en cours et les demandes d'allocation. La figure 2.7 résume schématiquement le modèle :

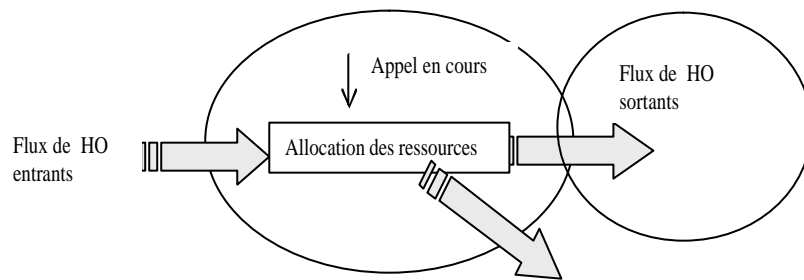


Figure 2.7 Flux global de mobilité

L'étude de la mobilité des utilisateurs permet en premier lieu, d'évaluer les taux de handovers entre cellules et en deuxième lieu, de les considérer pour une analyse de trafic multimédia plus précise. Il s'agit alors de faciliter la tâche d'allocation des ressources radio pour l'opérateur réseau qui aura besoin de la connaissance du critère de mobilité. Ce critère permet de scinder la population des utilisateurs en un ensemble d'utilisateurs rapides et un autre ensemble disjoint d'utilisateurs lents. Cette mobilité est déterminée par la connaissance de la vitesse moyenne des utilisateurs.

Plusieurs solutions ont été proposées dans la littérature [MON98] [LEU98][TABB97][NAN93] pour déterminer la mobilité d'un utilisateur. Les solutions sont basées sur l'évaluation du temps de séjour résiduel, sur le signal reçu et enfin sur la mesure des temps moyens de séjour d'un mobile dans les cellules traversées.

2.2.6 Le trafic

La finalité essentielle d'un réseau cellulaire est de permettre l'écoulement de communications entre une diversité d'origine, de mode d'accès et de destinations de manière à satisfaire au mieux les besoins des utilisateurs du réseau. Dans le cadre de la conception d'un tel réseau, des économies peuvent être réalisées en partageant différents éléments qui le composent. Tous les utilisateurs n'ont pas besoin de communiquer en

même temps ou de manière permanente. La demande de communication ne porte pas forcément sur les mêmes destinations. Il s'agit donc de définir le niveau optimal de partage des ressources du réseau et plus précisément le dimensionnement de la capacité du réseau et garantir la mobilité en évaluant le taux de handover et de *roaming*. Ces éléments définissent des objectifs de qualité pour l'écoulement des communications qui sont définies à priori. Soit le modèle de dimensionnement suivant (voir figure 2.8) :

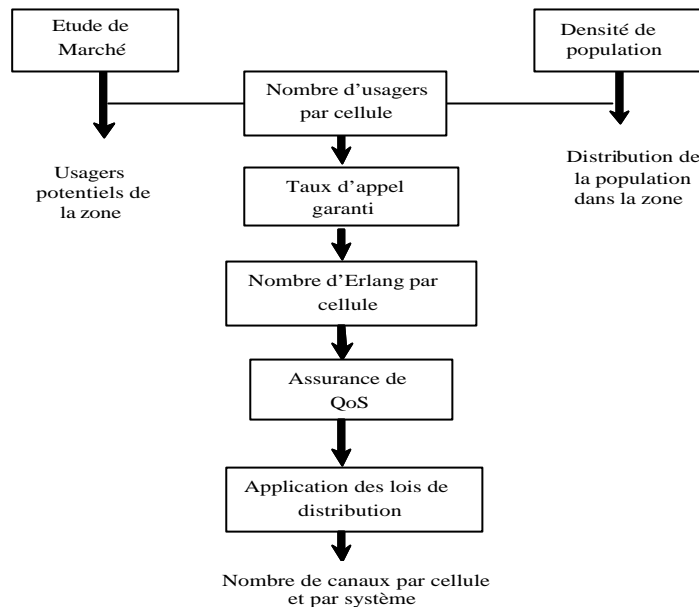


Figure 2.8 Synthétique de dimensionnement de réseau

2.2.6.1 Modélisation de la demande

La demande en trafic pour un service dépend essentiellement du débit requis par son utilisation et du nombre des utilisateurs qui l'utilise dans la zone considérée (taux de pénétration). Le volume de trafic par utilisateur dépend aussi du nombre moyen d'appels par heure, de la durée d'un appel et du débit nécessaire. Il faut noter que les deux premiers paramètres dépendent fortement de l'activité de l'utilisateur.

2.2.6.2 Techniques de prédiction

La prédiction du trafic à long terme est une activité importante des activités de planification de réseaux. La connaissance de la demande à long terme permet de définir l'infrastructure future du réseau. Le calcul de la demande de trafic pour un service dépend du nombre de personnes utilisant ce service et donc de son taux de pénétration parmi les différentes catégories d'utilisateurs. La croissance du taux de pénétration passe généralement par trois étapes successives :

- Une croissance linéaire lente : phase de démarrage ;
- Une croissance exponentielle : phase d'expansion ;
- Une croissance ralentie : phase de saturation.

2.2.6.3 Distribution de la population

Afin d'affiner la modélisation de la demande, différents types d'environnement géographiques sont segmentés en zones de trafic homogènes. Les types de zones considérés sont classés en deux catégories :

- Milieux intérieurs : les gares, aéroport, centres commerciaux, etc.
- Milieux extérieurs : centre ville (zones piétonnières, bureaux, etc.), zones urbaines (zones résidentielles, bureaux), zones rurales (peu peuplées), routes (les routes principales génèrent un trafic très différent de leur environnement). Ce type d'environnement vient donc se superposer à l'un des types de milieu extérieur. Les routes (et les rues) secondaires sont également prises en compte.

2.2.6.4 Distribution horaire

La demande en trafic dépend également du moment de la journée et pendant l'heure chargée pour pouvoir dimensionner le réseau. La connaissance des variations de trafic pour chaque environnement nous permet de définir les zones de congestion, ainsi que les configurations de réseau appropriées.

2.3 La mise en place d'un réseau UMTS

Lors de la mise en place d'une couverture UMTS régionale ou nationale, la première opération à effectuer est la planification du réseau et la deuxième est le dimensionnement de l'infrastructure considérée. Ces opérations visent à déterminer le nombre de sites nécessaires pour obtenir une couverture déterminée.

Etant donné les spécifications de l'UMTS par rapport au GSM, ces opérations requièrent la connaissance d'un certain nombre de données afin de déterminer la demande de trafic :

- Densité de population ;
- Taux de pénétration estimé (appels entrants/sortants) ;
- Mobilité des utilisateurs ;
- Zones d'activités ;
- Services supports ;
- Trafic par service, etc.

Lors de telles opérations, un certain nombre d'hypothèses ont été considérées, concernant le profil des abonnés (voix et/ou donnée), le type de mobilité (piéton, voiture, etc.), la charge relative de chaque cellule ou encore les services envisagés. A partir de ces données, nous pouvons estimer le nombre de sites nécessaires par zone géographique.

L'objectif est de préciser l'impact des ces différents paramètres sur le dimensionnement d'un réseau UMTS, afin d'offrir aux usagers une qualité de service optimale. La qualité de service revêt de multiples aspects. Nous nous intéressons à ce niveau à deux aspects importants : le taux de blocage des appels et le taux de coupure des communications.

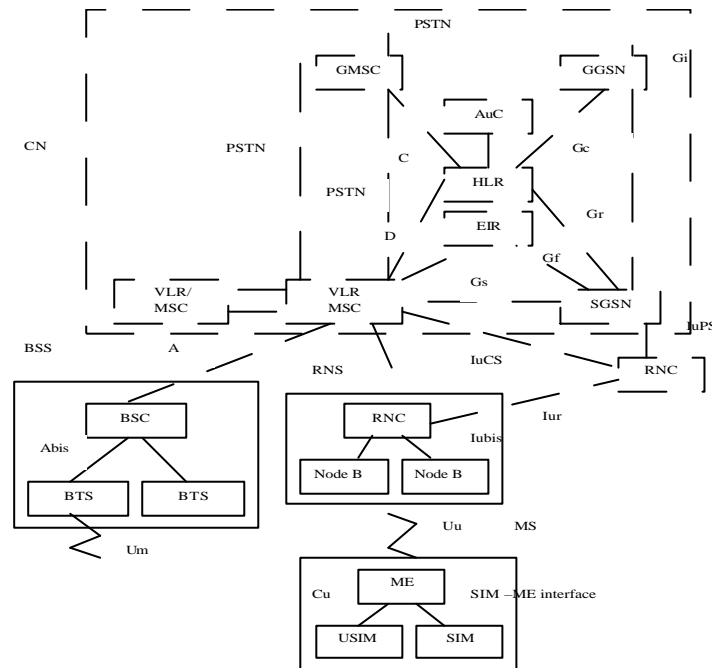
2.3.1 La topologie d'un réseau UMTS

L'architecture du réseau de base UMTS et son évolution reposent sur deux principes (voir figure 2.9) :

- La nécessité de préserver les investissements GSM et de conserver l'architecture du sous système réseau du GSM ;
- L'utilisation extensive de technologies de commutation et de transmission de paquets, notamment IP, pour bénéficier de coûts d'exploitation réduits.

Le réseau cœur présente les fonctions suivantes :

- Gestion de l'information de localisation de l'utilisateur ;
- Contrôle des caractéristiques du réseau et des services ;
- Commutation et transmission des informations de signalisation et des informations générées par l'utilisateur ;
- Communication avec le réseau d'accès (UTRAN) et les autres réseaux cœurs.



- HLR : Home Location Register
- VLR : Visitor Location Register
- SGSN : Serving GPRS Support Node
- GGSN : Gateway GPRS Support Node
- Auc : Authentication Center
- EIR : Equipment Identity Register
- MSC : Mobile Switching Center
- GMSC : Gateway Mobile Switching Center
- PLMN : Public Land Mobile Network
- BSS : Base Station System
- BSC : Base Station Controller
- BTS : Base Transceiver Station
- RNC : Radio Network Controller
- RNS : Radio Network System
- CN : Core Network
- ME : Mobile Equipment
- SIM : Subscriber Identity Module
- USIM : UMTS Subscriber Identity Module
- PSTN : Public Switched Telephone Network/ RTC

Figure 2.9 Topologie d'un réseau UMTS [NETW02]

2.3.2 La planification d'un réseau UMTS

La connaissance de l'environnement est indispensable à une bonne planification radio : elle permet en effet d'anticiper sur la demande future et les moyens d'y répondre au mieux. En outre, elle garantit à l'opérateur une rentabilité correcte de son réseau. Notons que plus la connaissance de l'environnement est fine, plus le *design* du réseau deviendra pertinent.

Une configuration typique est par exemple celle retenue par [LOU01] :

- Zone A : centre ville de type urbain dense (centre d'affaire par exemple) ;
- Zone B : principaux lieux de trafic (centres commerciaux, pôles industriels, zones urbaines, etc.) ;
- Zone C : zones à plus faible trafic (îlots résidentiels, zones suburbaines éloignées, etc.).

Signalons que le principe de base d'une planification de réseau UMTS est identique à celui d'une planification GSM. En effet la planification cellulaire consiste à positionner les sites suivant une maille de type hexagonal, comme indiqué sur la figure ci dessous. Bien que, en réalité, il ne soit que très peu probable de respecter rigoureusement une maille régulière, notamment en raison de l'hétérogénéité des milieux de propagation rencontrés (cluster, relief, etc.). Il sera nécessaire d'évaluer une distance inter-sites moyenne sur laquelle s'appuiera pour réaliser le positionnement des sites [BAR00][LOU01].

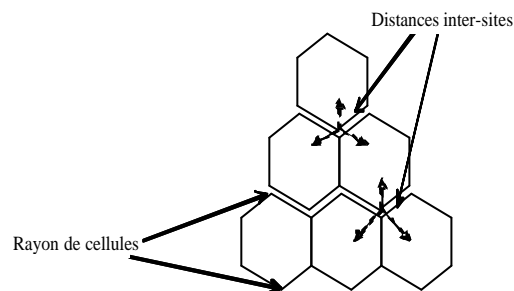


Figure 2.10 Positionner les sites suivant une maille de type hexagonale

On remarquera que la distance inter-sites se déduit du rayon de cellules à l'aide de la relation suivante :

$$D_{\text{inter-sites}} = \frac{3}{2} \text{Rayon} \quad (2.1)$$

Différents environnements sont définis dans (UMTS Forum Report 6 [3GPP]) le tableau 2.1 :

Environnement	Densité théorique (util/km ²)	Type de cellule
Urbain dense	180 000	micro/pico
Urbain	100 000	macro/micro
Sub-urbain	10 000	macro
Rural	100	macro

Tableau 2.1 Différents environnements sont définis

2.3.2.1 Dimensionnement en trafic

Le principe est le suivant : on cherche à évaluer le nombre d'utilisateurs par site en fonction de données *Marketing*. Le nombre de sites nécessaires est alors le rapport entre le nombre d'utilisateurs par site et le nombre total d'utilisateurs dans la zone. Le dimensionnement d'un réseau UMTS en fonction du trafic se décompose en trois étapes :

- Estimation de la demande globale exprimée en kbps ;
- Définition de la capacité des sites, ou encore la bande passante du site (kbps) ;
- Décompte du nombre de sites.

2.3.2.2 Estimation de la demande

La première étape du processus de dimensionnement consiste en l'estimation de la consommation des ressources pour une distribution d'utilisateurs d'un profil donné et au moment le plus chargé. La demande des utilisateurs s'exprime en kbps afin d'apprécier sur une base commune la consommation de services de type circuit (classiquement exprimée en Erlang), et la consommation de services de type paquets (exprimée en kbps).

Services i	Débits (kbps)	Mode
S	8	circuit
LCD 32	32	circuit
LCD 64	54	circuit
LCD 144	144	circuit
LCD 384	384	circuit
UDD 32	32	paquet
UDD 64	64	paquet
UDD 144	144	paquet
UDD 384	384	paquet
UDD 2048	2048	paquet
S/Speech, LCD : Long Constrained Delay, UDD : Unconstrained Delay Data		

Tableau 2.2 Segmentation des services proposée par l'ETSI

Les données d'entrée du processus de dimensionnement sont délivrées par les services envisagés par les opérateurs [LOU01]. Ces données sont :

- Les profils utilisateurs qui donnent la consommation moyenne, à l'heure la plus chargée, attachés à chaque support (voir tableau 2.2, segmentation proposée par l'ETSI [UMTS-FORUM]).
- La distribution des utilisateurs.
- La consommation relative en fonction des supports par profil utilisateur.

2.3.2.2.1 Mode circuit

Le trafic total demandé par un utilisateur utilisant le service i s'exprime par la formule suivante [LOU01] :

$$T_i = \sum_{k=1}^{N_{profil}} \mathbf{a}_k \cdot \mathbf{e}_{k,j} \cdot \bar{t}_{k,i} \text{ (Erl)} \quad (2.2)$$

avec :

- N_{profil} : le nombre total de profil d'utilisateurs dans la zone considérée ;
- \mathbf{a}_k : la proportion d'utilisateurs de profil k dans la zone considérée (pourcentage du nombre total d'utilisateurs) ;
- $\mathbf{e}_{k,i}$: la consommation relative du service i par utilisateur de profil k (%), (i.e., le ratio du temps passé à consommer du service i sur le temps moyen total d'une consommation de l'utilisateur de profil k ;
- $\bar{t}_{k,i}$: le trafic moyen (Erl) généré par l'utilisateur de profil k sur le service i . Le trafic est défini pour absorber les pics de trafic en heure de pointe.

On peut déduire la demande globale pour des QoS données. Pour N_{users} utilisateurs, la demande exprimée en kbps est :

$$Demande_{circuit}(N_{users}) = \sum_{i=1}^{N_{sc}} Erlang_B^{-1}(N_{users}, T_i, QoS_i) \cdot R_i \quad (2.3)$$

avec :

- N_{sc} : le nombre de services en mode circuit
- R_i : le débit utile du service i (kbps)
- QoS_i : la qualité de service paquet i , (i.e., la probabilité de blocage P_b (%))
- $Erlang_B^{-1}(t, P_b)$ donne le nombre de canaux nécessaire pour écouler le trafic t avec une probabilité de blocage P_b . Cette fonction est l'inverse de la loi d'Erlang B qui s'exprime comme suit :

$$P_b(t, N) = \frac{\frac{t^N}{N!}}{\sum_{i=0}^N \frac{t^i}{i!}} \quad (2.4)$$

2.3.2.2.2 Mode paquet

Le volume total de données utiles demandé par un utilisateur utilisant le service i s'exprime par :

$$V_i = \sum_{k=1}^{N_{profil}} \mathbf{a}_k \cdot \mathbf{e}_{k,j} \cdot \bar{v}_{k,i} \quad (2.5)$$

avec :

- N_{profil} : le nombre total de profil d'utilisateur dans la zone considérée ;
- \mathbf{a}_k : la proportion d'utilisateurs de profil k dans la zone considérée (pourcentage du nombre total d'utilisateurs) ;
- $\mathbf{e}_{k,i}$: la consommation relative du service i par utilisateur de profil k (%) (i.e., le rapport entre le temps passé à consommer du service i , et le temps moyen total d'une consommation de l'utilisateur de profil k) ;

- $\overline{v_{k,i}}$: le volume moyen (kbps) généré par l'utilisateur de profil k sur le service i sur le temps moyen total d'une communication de l'utilisateur de profil k . Le trafic est défini pour absorber les pics de trafic à l'heure de pointe.

2.3.3 Le modèle de l'environnement considéré

Nous nous plaçons dans l'environnement ci-dessous, (voir figure 2.11, cf. chapitre 7, Glossaire 1) l'infrastructure sur laquelle le modèle de trafic et de mobilité va être appliqué.

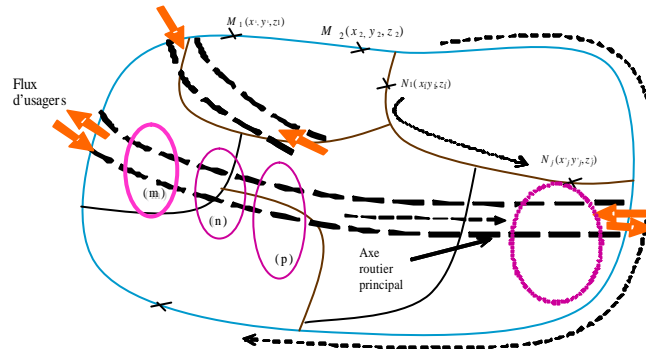


Figure 2.11 Modèle d'environnement considéré

L'environnement géographique choisi peut être une ville où il s'agit d'un ensemble de sous-zones.

Lors de la phase de la planification de modélisation du trafic et de la mobilité des utilisateurs, le modèle d'environnement peut être décrit par les paramètres indiqués dans la figure 2.11. Ces paramètres seront utilisés par la suite pour aboutir à un modèle optimal de mobilité des utilisateurs (cf. chapitre 7).

2.4 Conclusion

Dans ce chapitre nous avons présenté le processus de planification d'un réseau radio mobile de 3^{ème} génération. Ce processus englobe un nombre important de paramètres réseau envisagés pour les services multimédia tels que l'environnement à couvrir, un nombre estimé d'abonnés potentiels, et une bande de fréquence allouée pour la transmission des communications. Il s'agit alors de définir la taille des cellules, le type de trafic, la mobilité des utilisateurs et une stratégie fiable d'allocation des ressources (nombre des canaux, réservation des routes, etc.) qui satisfasse la demande des utilisateurs en services multimédia. Ces paramètres réseau sont utilisés par la suite pour gérer la qualité de service de l'allocation de ressources. Ils constituent une base d'information pour re-paramétrer et re-dimensionner le réseau.

Références Bibliographiques

[BAR00]Barnaba, J.M., “*Ingeniery of cellular networks*,” Cours MST, TDF-C2R -AGR-JMB-407-00.

[BES99]Besn-Othman, J. and Valois, F., “*Multiservices allocation in hierchical cellular networks*,” the 4th IEEE Symposium on Computer and Communications (ISCC’99), July 1999, Sharm el Sheikh, Red Sea, Egypt, pp. 80-86.

[CEN95]Cengiz, C., “*GSM to UMTS/A perspective on the future*,” ICT’95, April 1995, pp. 351-358.

[ETSI][http : // www.etsi.org/](http://www.etsi.org/).

[FOS93]Foschini, G.J., Gopinath, B. and Miljanic, Z., “*Channel cost of mobility*,” IEEE Transactions on vehicular technology, Vol. 42, No.4, November 1993.

[LEU98]Leung, K.K., “*Traffic models of wireless communication network*,” IEEE Journal on selected areas of communications, Vol 12, No.8, October 1998.

[LAG97]Lagrange, X., “*Analyse of hierarchical cellular network*,” First International Workshop on Mobile and Wireless Communications Networks, April 1997, pp. 161-169.

[LOU01]Laussert, A. and al., “*Radio planification of UMTS network : 3G networks study*,” UMTS group, C2R- SER - SC- 0233- 2001.

[MON98]Montenegro. G., and al., “*Time dependent analysis of mobile communication trafic in a ring shaped service area with non uniform vehicule distribution*,” IEEE Transaction on vehicular technology, Vol.41, No.3, August 1992.

[NAP98]Napolitano, A. and Pannioli, F., “*Evolution of the GSM platform*,” IEEE International Conference on Universal Communications (ICUPC’98), Florence, Italy, October 1998.

[NAN93]Nanda, S., “*Teletraffic models for urbain and suburbain microcells; cell sizes and handoff rates*,” IEEE Transactions on vehicular technology, Vol.42, No.4, Nov 93.

[OTH99]Othman, J.B. and Valoi, F., “*Multiservices allocation in hierarchical cellular networks*,” the 4th IEEE Internaional Symposium on Wireless Communications (ISWC’99), pp.66-78, Victoria (Canada), June 1999.

[QoS99]3rd Generation Partnership Project, “*Technical Specification Group Services and System Aspects; QoS Concept and Architecture*,” Release 1999.

[SES92]Seskar, I. and al., “*Rate location area updates in cellular environnement*,” 42 th IEEE Vehicular technology conference 1992.

[STE95]Steel, R. and al., “*Teletraffic performance of GSM 900/DCS 1800 in street microcells*,” IEEE Communications magazine, march 1995.

[THO98]Thomas, R. and al., “*Influence of the moving of the mobile stations on the performance of a radio mobile cellular network*,” Proceeding of the Nordic Seminar on Digital Land Radio Mobile Communications, 1998.

[TAB97]Tabbane, S., “*Mobile network*,” Edition Hermes 1997.

[Rauhala98] Rauhala, K, “*Wireless ATM base line text*,” Technical report, ATM forum, San Diego, CA, April 1998.

[TER95] “*Terms and definitions related to quality of service and network performance including dependability*,” IUT-T Recommendation E-800.

[TABB97] Tabbane, S., “*Location management methods for third-generation mobile systems*,” IEEE Communication Magazine; August 1997.

[UMTS-FORUM] [http : // www. umts-forum.org/](http://www.umts-forum.org/).

[VAL99] Valois, F. and al., “*Performance modelling of hierarchical networks using pepa*,” Process Algebra and Performance Modeling Workshop 1999 (PAPM’99), Zaragosa, Spain, Sept. 1999, pp. 139-154.

[W3C] World Wide Web Consortium W3C, [http : //www.w3.org/](http://www.w3.org/).

[ZHO98] Zhou, Y. and Jabbari, B., “*Performance modelling and analyse of hierachical wireless communications network with overflow and take back traffic*,” the Ninth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’98), Boston, USA, September 1998.

[3GPP] [http: // www.3gpp.org/](http://www.3gpp.org/).

Chapitre 3

Analyse des modèles de trafic

3.1 Introduction

La diversité des services, rend complexe l'étude des performances d'un réseau radio mobile multimédia, particulièrement lorsqu'il s'agit du contrôle d'admission et de la gestion de la qualité de service. Une analyse préalable de l'ensemble des services potentiels permet d'établir la correspondance entre les différents services potentiels et les classes de services prédéfinies pour l'UMTS par référence au domaine d'application et aux exigences de qualité de service.

Dans ce chapitre, nous allons présenter les différentes classes de service telles que définies dans la norme 3GPP. Ces modèles seront utiles pour la modélisation de la mobilité des utilisateurs au niveau du chapitre 7, notamment nous nous sommes limité au choix des modèles de trafic pour les trois premières classes de services, à savoir voix, vidéo-téléphonie et web pour les différents types d'usage vu que ces modèles sont caractérisés par de faibles débits. Ils ne présentent pas de contraintes temporelles sévères et sont faiblement sensibles aux délais de transmission et aux contraintes de synchronisation (cf. annexe D). Ces modèles sont nécessaires à l'évaluation de la charge du réseau (évaluation quantitative du trafic).

3.2 Classes de service

Selon les spécifications 3GPP [3G23.107][KOO01], il est possible de partitionner, sur la base de la qualité de service, l'ensemble des services potentiels en quatre classes (voir tableau 3.1) :

- Services de type conversationnel (*Conversational class*) ;
- Services à flux de données continu (*Streaming class*) ;
- Services interactifs (*Interactive class*) ;
- Services en mode téléchargement (*Background class*).

Le critère de classification prépondérant est la sensibilité aux délais de transmission. Les deux premières classes sont prévues pour les services du type temps réel alors que les deux autres classes concernent le reste des applications caractérisées par de faibles contraintes de temps ou qui sont insensibles aux délais de transmission.

Classe de trafic	Exemple d'application	Type de capacité de réservation
<i>Conversational class</i>	<i>Voice, Video telephony</i>	<i>Static</i>
<i>Streaming class</i>	<i>Real time streaming video</i>	<i>Static</i>
<i>Interactive class</i>	<i>Web browsing. Real time control channel</i>	<i>Dynamic</i>
<i>Background class</i>	<i>Download of mails and files</i>	<i>Dynamic</i>

Tableau 3.1 Classes de services

3.2.1 Classe des services conversationnels

Sous cette classe, on peut classer toutes les applications qui permettent la conversation directe entre deux ou plusieurs utilisateurs. On trouve, en particulier, les communications téléphoniques, les vidéo-conférences, les audio-conférences et de façon plus générale, les services interactifs temps réel.

Pour ce genre de service, la qualité de service est spécifiée par référence à la perception de l'utilisateur qui permet de dégager des seuils d'appréciation du service tels que la distorsion maximale tolérée du signal audio/vidéo reçu et le retard maximal à la réception (à l'origine des problèmes d'écho). Les caractéristiques suivantes sont typiques à cette classe de service :

- Délai de transmission très réduit ;
- Taux de distorsion du signal limité ;
- Conservation des relations temporelles du flux multimédia (voir le paragraphe suivant).

3.2.2 Classe des services à flux continu

Cette classe de service inclut toutes les applications temps-réel qui ont lieu dans un seul sens de transmission, généralement en mode diffusion (*Distributional Service*) et

qui sont généralement caractérisées par un flux de données quasiment continu dans le temps. Parmi les exemples typiques de telles applications figurent les séquences vidéo (clips et extraits de films) et audio (extraits de musicaux).

Les applications relatives à cette classe sont en général de type multimédia en mode diffusion. La présence de plusieurs types d'entités dans la composition de ce genre d'applications impose des contraintes de qualité de service relatives à la sensibilité aux erreurs (à cause du taux de compression élevé) et à la synchronisation entre les entités.

La synchronisation s'établit à deux niveaux [3G23.107] :

- Synchronisation entre composants différents (*Intermedia Synchronisation*) : il s'agit de conserver les relations temporelles entre les composants dépendants (*Time dependant media components*). Un exemple de composants média dépendants est le couple voix et séquences vidéo. Pour cet exemple, la relation temporelle est définie par la dépendance sémantique, par référence à l'utilisateur, qui impose la réception simultanée de séquence vidéo et de la trame audio associée.
- Synchronisation entre les unités consécutives d'information au sein du même composant (*Intramedia Synchronisation*) : il s'agit de la conservation du délai de transmission des différentes trames transmises. Cela signifie que la durée d'inter-arrivée de deux trames consécutives à la réception doit être égale à la durée séparant les instants de génération de ces deux trames au niveau de la source d'émission.

Il est utile de noter que deux modes de transmission de données multimédia sont possibles :

- Le premier mode, appelé transmission multi-canaux (norme H.623 [3G25.107], par exemple), assure la transmission de chaque composant média sur un canal spécifique. Pour ce genre de transmission, les deux niveaux de synchronisation doivent être assurés.
- Le deuxième mode (norme MPEG-4 [3G25.107], par exemple) a lieu sur un seul canal, le flux des données transmises est considéré comme étant le résultat d'un multiplexage entre les différents composants média. Pour ce mode, seule la synchronisation intra-composants doit être assurée puisque la synchronisation inter-composants est résolue par la procédure de multiplexage.

3.2.3 Classe des services interactifs

Il s'agit ici de l'ensemble des applications interactives classées non temps réel, c'est à dire qui ne présentent pas de contraintes temporelles sévères ou qui sont insensibles aux délais de transmission et aux contraintes de synchronisation. Dans cette classe de service, on trouve toutes les applications faisant intervenir la transmission de données en mode interactif telles que les consultations de bases de données distants, l'accès aux LAN, Internet et Intranet.

Pour ce genre d'applications, la contrainte prépondérante est la reconstitution sans erreurs du message global à partir du flux de données transmises. Pour un taux d'erreur binaire (BER) spécifié à l'avance, il est impératif de mettre en œuvre toutes les

procédures de protection contre les erreurs, en particulier les procédures de retransmission. Cependant, la chaîne de transmission ne doit pas introduire des délais aberrants pour la reconstitution fidèle d'un message à la réception, ce délai étant conditionné par un seul paramètre *Time Out*.

3.2.4 Classe des services en mode téléchargement

Cette classe de services inclut l'ensemble des applications qui ne sont pas interactifs. Cela signifie que la transmission a lieu dans un seul sens en mode téléchargement (vers l'utilisateur du service). Au niveau de la qualité de service, la seule contrainte à respecter est de pouvoir reconstituer, à la réception, le message transmis sans erreur (en respectant un BER seuil). A l'inverse des applications interactives de la troisième classe, aucune contrainte temporelle n'est imposée.

Les paramètres de la QoS les plus importants tels que définis dans les spécifications 3GPP [3G23.107][KOO01] sont :

- Débit maximal : la connaissance de ce paramètre sert à la réservation des codes pour la transmission ;
- Délai de transfert : c'est le délai maximal durant la durée de service. Ce paramètre est important lors du choix du mode de transfert du flux de données ;
- Priorité d'accès et allocation de ressources : ce paramètre permet de superviser l'allocation de ressources aux différents services ainsi que la procédure d'accès aux ressources.

3.3 Modèles de trafic

Dans ce chapitre, nous nous intéressons au choix des modèles de trafic pour les trois premières classes de services, à savoir voix (pour *Conversational class*), vidéo (pour *Streaming class*) et Web (pour *Interactive class*). Ces modèles sont nécessaires à l'identification de la charge du réseau (évaluation quantitative du trafic).

La spécification de la charge du réseau suppose une connaissance préalable des caractéristiques des services de point de vue trafic. La modélisation classique des services par des processus de Poisson n'est plus valide dès lors qu'il s'agit de la transmission de données. Par conséquent, l'analyse du comportement du réseau nécessite la connaissance des statistiques décrivant l'activité de chaque type d'application pour ne pas s'écarter de la réalité.

3.3.1 Modèle de communications téléphoniques

Les communications téléphoniques constituent le service le plus classique dont le comportement statistique est maîtrisé depuis longtemps. Le comportement d'un utilisateur exploitant ce service au cours du temps est modélisé par un processus Markovien du type *ON-OFF* [NAJ01][AJI00]. Les statistiques de ce modèle sont :

- L'occurrence des appels téléphoniques est un processus de Poisson caractérisé par un taux moyen d'appel de valeur typique $I_{appel} = 5.5 \times 10^{-5}$ appels/s soit l'équivalent de 0.2 appels par heure ;

- La durée d'un appel est un processus exponentiel de moyenne m_{appel} telle que $1/m_{appel}=150s$;
- La durée de l'appel est une alternance de périodes d'activité (*talkpurt*) et de périodes de silence qui sont distribués exponentiellement avec des moyennes respectives $m_{activité}$ et $m_{silence}$ telles que $1/m_{activité}=1.00s$ et $1/m_{silence}=1.35s$, correspondant à un taux moyen de communication effective de l'ordre de 40%.

3.3.2 Modèle de flux multimédia

Le codage MPEG prend une importance croissante au niveau du codage des flux de données multimédia, c'est la raison pour laquelle nous avons choisi un modèle de trafic basé sur ce type de codage. Il est également important de noter que le volume de séquences vidéo constitue la partie prépondérante dans le trafic multimédia. Par conséquent, il est suffisant d'approximer le trafic multimédia par un modèle de la composante vidéo, surtout que la transmission ait lieu via des procédures de multiplexage. Le modèle de trafic que nous avons retenu dans ce contexte [DAW99] s'appuie sur le codage MPEG-4 et prend en considération la texture et le degré d'animation des séquences vidéo.

Selon ce modèle, le flux des séquences vidéo correspond à une série de trames de données de même durée à raison de 25 trames par seconde. De point de vue codage, ces trames sont classées en trois classes I, P et B correspondant respectivement aux formes intrinsèques (*Intra*), prédite (*Predictive*) et doublement prédite (*Bidirectionally-predictive*). La différence se fait au niveau du taux de compression. L'occurrence de ces trois classes de trames est structurée sous forme de groupement d'image (GOP, *Group of Pictures*) de 12 trames incluant une trame I, trois trames P et huit trames B.

Outre le classement en trame I, P ou B, les trames sont également classées par référence au contenu de la séquence qui se traduit par la texture et par le degré d'animation (*motion*). Dans le cadre de ce modèle, on utilise trois niveaux de texture H (pour *High*), M (pour *Medium*) et L (pour *Low*) et trois niveaux d'animation H, M et L. Ce qui donne au total neuf types de trames. Le tableau 3.2 récapitule les caractéristiques de ces différents types de trame.

Type de trame	Texture	Animation	TMC (bit/pixel)
1	L	I	0.348
2	L	M	0.537
3	L	H	0.601
4	M	L	0.568
5	M	M	0.786
6	M	H	0.971
7	H	L	0.956
8	H	M	1.214
9	H	H	1.869

Tableau 3.2 Classification des trames par contenu

3.3.3 Modèle de flux Internet

Pour modéliser le trafic sur Internet, nous avons retenu le service de consultation des sites web (*Web Browsing*), l'application la plus exploitée sur Internet. La modélisation à ce niveau fait référence d'une part au comportement de l'utilisateur et d'autre part au mode de transmission par paquets via le réseau Internet. La figure 3.1 illustre le flux de données relatif à une session de consultation du web.

Le flux de données, selon ce modèle, peut être décomposé en plusieurs sessions de consultation du Web. Pendant chaque session, l'utilisateur consulte un ensemble de sites Web se ramenant à un appel des pages HTML correspondantes. La composition des pages HTML sous forme d'hypertexte fait que leur téléchargement est matérialisé par la transmission de plusieurs *datagrammes* de taille variable. Avant de décider d'amorcer la consultation d'un autre site, l'utilisateur a besoin d'un temps de lecture pour explorer le site courant [AJI00].

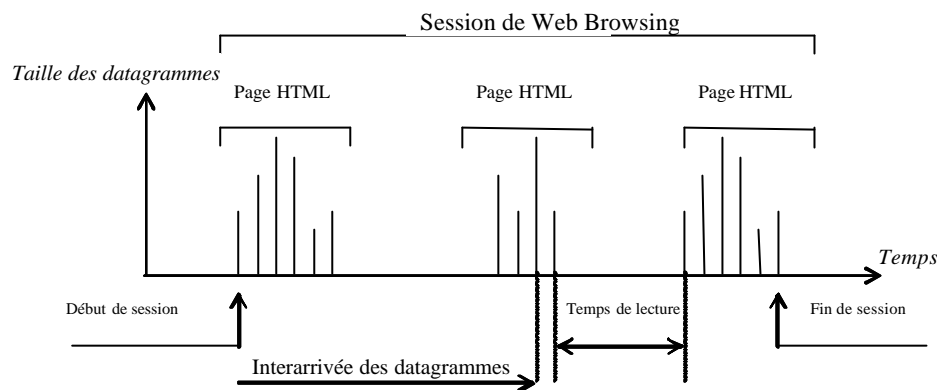


Figure 3.1 Flux de données pour une session de *Web browsing*

Les caractéristiques statistiques de ce modèle sont les suivantes :

1. L'occurrence de session est un processus Poissonien ;
2. Pour chacune des sessions :
 - Le nombre d'appels de pages HTML suit une distribution géométrique de moyenne typique $m_{na} = 5$ appels/session ;
 - Le temps de lecture des pages HTML suit une distribution exponentielle de moyenne m_{t_1} telle que $1/m_{t_1} = 4 \text{ à } 12 \text{ s}$;
 - Le nombre de *datagrammes* par appel suit une distribution géométrique de moyenne $m_d = 10$ *datagrammes*/appel ;
 - La durée d'inter-arrivée de *datagrammes* suit une distribution exponentielle dont la moyenne est en fonction du débit de transmission sur le réseau (voir tableau 3.3) ;

Débit (kbps)	Temps d'inter-arrivée (ms)
8	500
32	125
64	62.5
144	27.7
384	10.4
2048	1.95

Tableau 3.3 Temps d'inter-arrivée des *datagrammes*

- La taille des *datagrammes* suit une distribution de Pareto dont la densité de probabilité est donnée par la formule (3.1) ($x_{\max} = 66\,666$ octets) :

$$f(x) = \begin{cases} 0, & x < k \\ \frac{ak^a}{x^{a+1}}, & k \leq x \leq x_{\max} \\ \left(\frac{k}{x}\right)^a, & x = x_{\max} \end{cases} \quad (3.1)$$

3.3.4 Modèle de Transfert de fichiers

3.3.4.1 Le modèle FUNET

Ce modèle considère des statistiques sur l'utilisation de courrier électronique. Il présente un modèle «satisfaisant» pour modéliser l'application de transfert des fichiers (ftp). La fonction de distribution de probabilité est une distribution de Cauchy à peu près : Cauchy (0.8,1). La taille maximale d'un paquet est de 10 koctets. L'expression mathématique de la fonction de Cauchy [BEL00] est :

$$\text{Cauchy} (0.8,1) = f(x) = \frac{1}{(\pi(1 + (x - 0.8)^2))} \quad (3.2)$$

3.3.4.2 Le modèle RAILWAY

Il peut présenter le trafic des e-mails [BEL00]. La taille moyenne des paquets est de 170 octets ; la taille maximale est de 1000 octets. La fonction de répartition est donnée par l'expression suivante :

$$F(x) = 1 - e^{\left(-\frac{x}{170}\right)} \quad (3.3)$$

3.3.5 Modèle de flux Web sur Internet

Le modèle de flux de données Web a été décrit dans paragraphe 3.3.3 de ce chapitre. Durant le téléchargement des pages Web pendant une session donnée, l'utilisateur passe par des phases de silence de deux types :

- Temps de lecture : durant ce temps l'utilisateur entre dans une phase de lecture ou de réflexion qui se traduit par l'arrêt d'envoi de requêtes d'appels de page Web.
- Temps d'attente entre deux *datagrammes* successifs : bien que ce temps est transparent pour l'utilisateur, il ne demeure pas ainsi, si on le compare par rapport à la taille temporelle d'un *datagramme* (par référence à son débit). Par exemple, pour un *datagramme* de taille moyenne de 480 octets qui est transmis à un débit de l'ordre de 144 kbps, la taille temporelle de ce *datagramme* sera de l'ordre de 26.6 ms alors que le temps d'écart entre deux *datagrammes* successifs à ce débit sera de l'ordre de 27.7 ms (cf. tableau 3.2).

Ces deux phases de silence peuvent constituer des paramètres déterminants pour le choix du bon scénario d'accès aux ressources.

Théoriquement, le flux de données web peut être modélisé par un processus du type MMPP [RUE96][HLAS98.], comme illustré par la figure 3.2. Les taux de transition I_s , m_s , I_p , m_p , I_d et m_d sont définis comme suit :

- I_s : taux d'arrivées de session.
- m_s : inverse de la durée moyenne d'une session.
- I_p : taux d'appels de pages Web à l'intérieur d'une session donnée.
- I_d : taux d'arrivées de *datagrammes* relativement à une page web donnée.
- m_d : inverse de la durée moyenne d'un *datagramme*.

Ces taux de transition peuvent être déduits à partir des paramètres du modèle comme indiqué dans la figure 3.1. Les paramètres m_{td} , m_{nd} , m_{t1} et m_{na} ont été définis dans le cadre du modèle de flux Internet (cf. paragraphe 3.3.3).

Dans les équations (3.4), \bar{x} désigne la taille moyenne d'un *datagramme* calculée sur la base de la loi de Pareto et vaut 480 octets lorsqu'on ne tient pas compte des coefficients de retransmission et de codage canal introduits par la chaîne de transmission. D_{web} désigne le débit (en kbps) de transmission du flux Web.

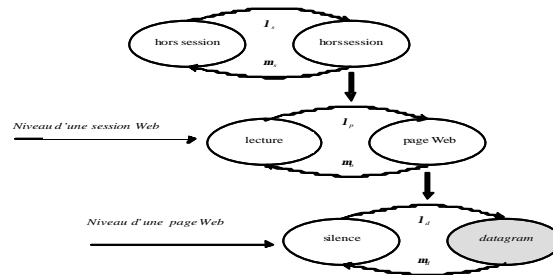


Figure 3.2 Représentation du flux Web par des processus MMPP

$$\left\{ \begin{array}{l} I_d = m_{td} \\ m_D = \frac{D_{web} \times 1000}{8 \times \bar{x}} \\ m_p = \frac{1}{m_{nd} \times (\frac{1}{m_d} + \frac{1}{I_d})} \\ m_s = \frac{1}{m_{na} \times (\frac{1}{m_p} + \frac{1}{I_p})} \\ I_s = \text{fonction (profil - usager)} \end{array} \right. \quad (3.4)$$

A partir de la décomposition du processus présentée ci-dessus, nous pouvons dégager les caractéristiques de flux en trames d'occupation du canal dédié à ce service. La réservation de canaux de communication peut avoir lieu à l'un des trois niveaux du processus à savoir : au niveau session, au niveau page Web ou au niveau de datagrammes :

- Dans le cas où les ressources sont réservées durant toute une session à un débit constant D_{web} : ce type de connexion peut se justifier par le fait que la définition d'une session correspond à l'accès à un seul serveur Web ; elle permet donc de réduire considérablement la signalisation spécifique à l'établissement des connexions radio (au niveau de l'UTRA) ou bien au niveau du serveur. En contre partie, cette option constitue un gaspillage de ressources. En effet, la durée moyenne d'une session est égale à $\Delta t_s = \frac{1}{m_s}$ alors que la durée globale de communication utile est égale en moyenne à $\Delta t_d = \frac{m_{na} \times m_{nd}}{m_d}$ ce qui donne un taux d'occupation du canal $t_s = \frac{\Delta t_d}{\Delta t_s}$ dont la valeur peut être calculée sur la base des formules (3.4), D_{web} étant exprimé en kbps.

$$t_s = \frac{m_{nd} \times \bar{x}}{m_{nd} + (\frac{m_{nd}}{m_{td}} + \frac{1}{m_{tl}}) \times D_{web} \times 125} \quad (3.5)$$

- La deuxième alternative consiste à allouer un canal de transmission à débit constant durant le téléchargement de la page Web courante, ce qui permet d'éviter l'occupation du canal de transmission durant la phase de lecture (le processus n'est pas en état d'activité). Ce choix a pour conséquences l'augmentation des procédures d'allocations/libérations des ressources (en moyenne, augmentation de l'ordre de 5 fois) ce qui signifie une charge de signalisation plus importante mais, en contre partie, cela permet une augmentation nette de taux d'occupation du canal de transmission.

- La dernière alternative consiste à allouer le canal durant le transfert d'un datagramme à débit variable. A ce niveau, on ne s'intéresse plus au taux d'occupation du canal du moment qu'il s'agit de la transmission d'information utile durant toute la connexion. Cependant, le nombre d'allocation/libération de canal s'amplifie en moyenne de 10 fois par rapport à la deuxième alternative.

3.4 Conclusion

Dans ce chapitre nous avons présenté les caractéristiques de chaque modèle de trafic. L'expression de la demande en trafic fait intervenir deux paramètres fondamentaux qui sont la densité des arrivées de demande de service et la durée de service. Ces deux grandeurs sont des paramètres personnalisés, considérés pour chaque type de service multimédia demandé dans les réseaux de 3G (accès à des bases de données et serveurs, transfert de fichiers, courrier électroniques, accès au réseau Internet, transmission de données à débit plus élevé, etc.).

Pour bien dimensionner le réseau radio mobile multimédia et analyser ses performances du point de vue charge de trafic, grade de service, réservation des ressources, etc. (cf. chapitres 7 et 8), la mobilité des utilisateurs doit être considérée dans les modèles de trafic choisis. Notons que ces modèles de trafic ne distinguent pas les différents mobiles ni leurs mouvements, mais modélisent la mobilité macroscopique dans le système. Ces modèles seront utilisés pour gérer la qualité de service et optimiser la conception du futur réseau pour l'allocation des ressources radio (cf. chapitres 8 et 10).

Références Bibliographiques

[AJI00]Ajib, W.W., “*Transmission management of data temporary flow in a radio mobile access TDMA*,” PhD thesis, ENST Paris, 2000.

[BEL00]Bellili, H., “*Radio resource allocation methodology of multimedia GSM network*,” PhD thesis, U.PARIS 6, 15 September 2000.

[DAW99]Dawood, A.M. and Ghanbari, M., “*Content-based MPEG video traffic modelling*,” IEEE Transactions on multimedia, Vol.1, No.1, March 1999.

[HLAS98]Hlavacs, H. and al., “*Traffic source modelling*,” Technical Report N° TR-99101, Institute of Applied Computer Science and Information Systems, University of Vienna, 1998.

[KOO01]Koodli, R. and Puskari, M., “*Supporting packet data QoS in next cellular networks*,” IEEE Communication Magazine, February 2001.

[NAJ01]Naja, R. and Thome, S., “*Multi-services call admission control and handover issues in wireless multimedia networks*,” MS3G’2001, Lyon, France, December 2001.

[RUE96]Rueda, A. and Kinsner, W., “*A survey of traffic characterization techniques in telecommunication networks*,” Proceeding of the IEEE Canadian Conference on Electrical and Computer Engineering, Calgary, Canada, Vol. 2, May 1996.

[3G23.107]3G TS 23.107, “*Quality of service concept and architecture*,”

Caractéristiques de qualité de service d'un réseau radio mobile cellulaire

4.1 Introduction

Le dimensionnement d'un réseau mobile de 3^{ème} génération doit permettre d'écouler une certaine quantité de trafic en respectant un niveau de qualité de service minimal pour chaque type de service demandé. Les paramètres de qualité de service les plus importants à considérer dans le processus de dimensionnement et planification sont la probabilité d'échec de handover (la probabilité de coupure de communication en cours) et la probabilité de refus d'établissement d'un nouvel appel. Une amélioration de ces deux indices de performances témoigne d'un trafic supporté plus important. Le respect d'un niveau de qualité de service pour un volume de trafic donné suppose un dimensionnement correct du réseau.

Nous allons commencer dans ce chapitre par la description des critères de performance réseau tels que la probabilité de refus d'établissement de nouvelle connexion et la probabilité de coupure d'appel en cours ainsi que les autres critères tels que le taux de handover et la capacité de réseau qui peuvent être pris en considération dans l'évaluation de la qualité de service de l'allocation de ressources. Puis, nous présentons les caractéristiques principales de la qualité de service du point de vue délai, gigue, etc. Enfin, nous consacrons la dernière partie de ce chapitre pour la classification, le suivi et la supervision de la qualité de service lors de la phase du re-paramétrage et re-dimensionnement du réseau.

4.2 Critères de performance réseau

Les politiques d'allocation des ressources dans un réseau radio mobile multimédia visent à optimiser les critères de performances les plus importants. Rappelons qu'ils sont, au moins, au nombre de deux [LAG97] :

- La probabilité de refus d'établissement de nouvelle connexion (*blocking connection*) : P_b . Elle correspond à un nouvel appel entrant sur la zone de couverture du réseau cellulaire qui ne peut être établi par manque de ressources fréquentielles ;
- La probabilité de coupure d'appel en cours, encore appelée probabilité d'échec de handover (*dropping handover*) : P_d . Un tel évènement se produit lorsqu'un utilisateur en cours de communication ne trouve pas de fréquences disponibles dans la cellule destination : sa communication est alors coupée. Il est clair que l'ensemble des politiques d'allocation des ressources tendent essentiellement à minimiser cette valeur car il est plus gênant pour un utilisateur d'être coupé en cours de communication plutôt que de se voir refuser l'accès au réseau par manque de ressources.

Ces mesures de probabilité peuvent être traduites simultanément par la mesure du niveau de service offert (*Grade of Service* - GoS). Cette variable est une mesure pondérée de la probabilité de refus d'établissement d'un nouvel appel et de la probabilité de coupure d'appel en cours. Nous avons alors :

$$GoS = gP_b + (1 - g)P_d \text{ où } g \in]0,1[\quad (4.1)$$

Plus la valeur de g est faible, plus la probabilité de coupure d'appel en cours est prédominante dans le niveau de service GoS. Nous constatons généralement que g prend la valeur de 0.1 [LAG98].

Bien entendu, d'autres critères peuvent être pris en considération dans l'évaluation d'une politique d'allocation de ressources. Ainsi nous rencontrons :

- Le taux de handovers. Ce paramètre traduit la charge de signalisation dans le réseau due aux handovers générés. Il correspond au nombre moyen de handovers exécutés par un utilisateur pendant une communication. Il peut être intéressant d'évaluer ce paramètre car il représente la charge de signalisation générée suivant les différentes politiques d'allocation de ressources,
- la capacité du réseau, exprimée en Erlang. Une politique d'allocation performante offrira, à charge équivalente, des probabilités de blocage plus faibles. C'est ce que cherche à traduire ce paramètre.

L'ATM Forum [RAU98] propose de définir des valeurs cibles pour ces paramètres de QoS :

- La probabilité de refus d'établissement d'une nouvelle phase d'appel doit être inférieure à 10% (valeur maximale tolérée) ;
- La probabilité de coupure d'appel en cours (due à un échec de handover) doit être inférieure à 1% (valeur maximale tolérée).

Les valeurs seuils proposées par [TABB97] sont :

- La probabilité de refus d'établissement d'une nouvelle phase d'appel est de l'ordre de 1% (valeur maximale tolérée) ;
- La probabilité de coupure d'appel en cours (due à un échec de handover) varie entre 1 et 2% (la valeur maximale tolérée).

4.3 Critères de classification des paramètres de qualité de service

4.3.1 Caractéristiques principales de qualité de service

Associer une classe de service et garantir des ressources à un certain trafic implique de pouvoir prédire le comportement du réseau (contrôle d'admission d'appel, pré-réservation et réservation des ressources, etc.).

La qualité de service se base sur les critères qui caractérisent le service dans le réseau permettant de délivrer un service de bonne qualité.

Ainsi, le délai de transit par paquet est la caractéristique principale retenue pour définir les besoins de qualité de service de bout en bout [DAH01]. Par rapport à ce paramètre, on distingue :

- Les services dits élastiques ou non temps réel ; par exemple, connexion à distance, ftp, www. L'application attend de recevoir les paquets avant de traiter les données, le délai de transit peut largement varier, le service est de type *best-effort*.
- Les services temps réel sensibles au délai et surtout à la gigue. Parmi ces services, on trouve par exemple : la voix, la vidéo, etc.
- La caractérisation de la qualité de service vidéo-téléphonie temps réel est généralement exprimée par les critères suivants :
 - Délai : les applications temps réel telles que la téléphonie, la vidéo, etc, présentent, par nature, des contraintes de temps. Par exemple dans une conversation humaine l'interactivité (*full duplex*) est primordiale, il faut pouvoir se couper la parole, interrompre l'autre partie, etc. Pour cela, le délai de transmission doit être borné. Le délai total de bout en bout ne doit pas excéder 200 ms (pour les applications interactives c'est souvent le temps d'aller - retour A-R qui est considéré). Le délai A-R ne doit pas donc dépasser 400 ms.

Le délai de bout en bout t sur UMTS/DiffServ [MAN00], se décompose en trois portions :

- délai d'émetteur t_1 : compression et codage, compression, et mise en paquet,
- délai de réseau t_2 : propagation, commutation et files d'attentes,
- délai de récepteur t_3 : *buffer* de gigue, dépaquetisation, décompression et décodage.

$$t=t_1+t_2+t_3 \quad (4.2)$$

- **Gigue** : ce n'est autre que la variation de délai qui pourrait être constant, ce qui conserverait la synchronisation du signal entre l'émetteur et le récepteur. Cette variation de délai détruit la base de temps du signal reçu. La principale cause de gigue réside dans les files d'attente des routeurs. Elle résulte de la variation de remplissage des files d'attente, ainsi le temps de traverser le réseau par des paquets appartenant à un même flot. La gigue peut faire varier sa valeur moyenne pour n paquets envoyés. Elle est donnée comme suit :

$$\Delta t = \frac{1}{n} \sum_{i=1}^n (t_i' - t_i) \quad (\text{variance moyenne}) \quad (4.3)$$

Le problème de gigue doit compenser au niveau d'application du terminal du destinataire en maintenant une mémoire tampon qui sert à collecter les paquets arrivant en retard. Si sa taille est la même que la valeur maximale de la gigue, tous les paquets seront récupérés par l'application. Cette mémoire est appelée *buffer* de gigue.

- **La perte** : signifie qu'une partie de l'information sera absente du signal lors de sa réception par le terminal de destination. Ceci peut avoir des conséquences graves sur la qualité subjective de la communication. En mode paquet, la perte est un problème qui demeure critique pour certaines applications telles que la vidéo-téléphonie, et qu'il faut savoir le résoudre avec les politiques de qualité de service. Quatre causes peuvent conduire à la perte de paquets :
 - la durée de vie TTL (*Time To Live*) épuisée,
 - le retard à l'arrivée supérieur à la taille du *buffer* de gigue (*Drop out*),
 - la congestion (*Queue overflow*) : rejet au niveau des routeurs,
 - un paquet erroné dû aux erreurs de transmission (vérifications de *Checksum*).

La transmission de la voix et de la vidéo doit composer avec ce défaut majeur du mode paquet. Lorsque le taux de perte se limite à un pourcentage réduit, le terminal est capable de reconstruire le signal. Au delà de 10%, la qualité se dégrade rapidement.

- **Bande passante ou débit maximum** : c'est le taux de transfert maximum.

L'ITU [DAH01] a défini un barème de mesure de la qualité de service pour les applications temps réel (MOS : *Mean Opinion Score*), sur une échelle à cinq graduations (voir tableau 4.1) :

	Excellent	Bon	Moyen	<i>Best effort</i> IP
MOS	4-5	3.8-4.2	2.9-3.8	2-2.9
Réseau	Télécoms	IP privé	IP bien géré	Internet
Délai pour établir l'appel	< 1.5	1.5-3s	> 3s	> 3s
Délai A-R (ms)	< 150	150-250	250-450	> 450
Gigue (ms)	< 1	< 20	< 18	> 80
Pertes de paquets (%)	Faible	< 5	< 10	< 10

Tableau 4.1 Mesure effectuée par l'ITU

4.3.2 Classes de qualité de service

Quatre classes de qualité de service peuvent être définies dans [RAU98] en fonction de leur sensibilité au retard et de leur tolérance aux erreurs. Le tableau 4.2 récapitule les caractéristiques de ces différents types de classes.

<i>Traffic class</i>	<i>Conversational class</i>	<i>Streaming class</i>	<i>Interactive class</i>	<i>Background class</i>
<i>Maximum bit rate</i>	×	×	×	×
<i>Delivery order</i>	×	×	×	×
<i>Reliability</i>	×	×	×	×
<i>Transfer delay</i>	×	×		
<i>Guaranteed bit rate</i>	×	×		
<i>SDUsize information</i>	×	-	-	-
<i>Traffic handling priority</i>	-	-	×	-
<i>Allocation /Retention priority</i>	×	×	×	×

Tableau 4.2 Classes de trafic pour l'UMTS

Les flux audio et vidéo diffèrent des flux de données sur plusieurs points. Un flux de données, par exemple HTTP ou FTP, requiert une transmission fiable. Aucune contrainte ne pèse sur le délai de transit ni sur la gigue. TCP offre un contrôle de congestion qui réagit aux pertes observées. Les capacités de TCP sont difficilement exploitables pour le transport de flux multimédia. Dans le cas particulier des flux audio et vidéo, la validité, et par conséquent l'utilité, d'un paquet dépendent de la stabilité du temps de transfert. Il est inutile de retransmettre un paquet perdu si ce dernier doit arriver après que la séquence à laquelle il appartient ait été jouée. Non seulement la retransmission est inutile mais elle consomme des ressources ce qui risque de retarder des nouveaux paquets. Dans ce cas, il est préférable d'accepter quelques pertes et de maintenir un flux continu et régulier d'information au niveau du récepteur. En fonction de ces classes, des paramètres sont définis et doivent être respectés selon le type de leur classe.

4.3.3 Profil d'usage

Ce profil intègre différentes composantes orthogonales constituées des informations suivantes [CHE94] :

- Profil de service individuel ou profil d'utilisateur ;
- Type de contrat ;
- Niveau de service demandé ;
- Gestion des priorités entre flux de média, etc.

4.3.4 Les besoins en qualité de service

- L'utilisateur peut choisir le scénario d'usage le mieux approprié à ses besoins ;

- La stratégie utilisée doit être basée sur un algorithme qui gère l'allocation des ressources disponibles, et ce pour répondre aux différents besoins de qualité de service;
- La continuité d'un service est nécessaire lors d'un handover indépendamment des zones géographiques (immeubles, zones urbaines, zones rurales, etc.) ;
- Les scénarios d'usages sont choisis suivant un jeu de débit allant de 8 Kbit/s à 2 Mbit/s.

4.3.5 Le champ d'étude de qualité de service

- Les scénarios d'usages qui permettent de classifier grossièrement les différents types de contraintes en matière de qualité de service ;
- Les dimensions de la qualité de service, qui identifie pour chaque scénario, les grandeurs à mesurer, à contrôler et à garantir, pour satisfaire les exigences souhaitées ;
- La modélisation de la qualité de service afin de construire des systèmes de gestion de qualité de service ;
- La définition d'un modèle de gestion de la qualité de service qui permet de surveiller le respect des contrats de qualité de service et de prendre des mesures d'adaptation en cas de non respect.

4.3.6 Les défis posés par qualité de service

L'utilisation croissante d'applications multimédia (*Conversational*, *Streaming*, *Interactive* et *Background class*) et l'utilisation massive de réseau partagé par un nombre important de trafic mobile, nécessite des systèmes pouvant gérer la qualité de service de manière dynamique. De tels systèmes sont sensés réagir à des exigences nouvelles de qualité de service au cours de leur vie. Ils doivent être capables de négocier des contrats de qualité de service et de gérer la qualité de service dynamiquement selon, tout d'abord un profil d'usage bien déterminé et ensuite suivant la disponibilité de réseau de point de vue ressources et capacités [ERI99] (voir figure 4.1) :

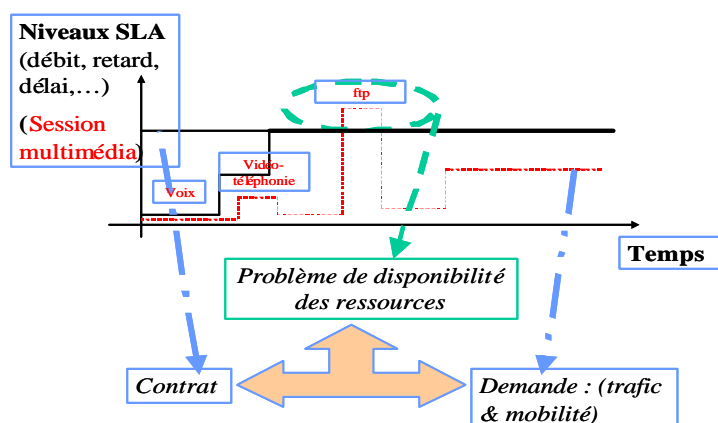


Figure 4.1 Satisfaction entre la demande, le contrat et la disponibilité des ressources

4.3.7 Gestion de qualité de service

La gestion de la qualité de service est réalisée par un certain nombre de processus au moment de l'établissement du service demandé. Deux types de gestion sont définis :

- Gestion statique : intervient au moment de l'établissement d'un service demandé par une application multimédia par exemple pour négocier la qualité de service de bout en bout.
- Gestion dynamique : intervient pendant l'exécution du service demandé, par exemple pendant la transmission d'un flux multimédia ou lors de la renégociation des ressources suite à la demande d'un nouveau service.

La négociation de la qualité de service est requise pour obtenir un accord sur le contrat de la qualité de service entre tous les processus considérés du système de télécom. Sa difficulté consiste à trouver un compromis qui satisfait les dits processus.

La qualité de service fournie par les processus du système varie au cours du temps, par exemple suite à :

- Une négociation de nouveaux services ;
- La disponibilité de nouvelles ressources qui avaient fait défaut ;
- La non disponibilité de ressources (cas d'encombrement) ;
- Un mauvais dimensionnement des modèles de trafic et de mobilité ;
- L'échec de handover ;
- Une défaillance au niveau du protocole de communication ;
- L'échec du protocole de négociation des ressources (complexité, divergence, etc.).

Les utilisateurs ont des exigences en ce qui concerne la vitesse d'exécution de certains services, et la disponibilité du système pour exécuter ces différentes applications. Satisfaire les exigences de qualité de service de bout en bout implique la coopération de tous les processus du système qui sont concernés par cette application/session pour fournir et garantir le service de bout en bout [ERI99]. La garantie de la qualité de service doit être effectuée de bout en bout tout au long de l'exécution de l'application. Deux scénarios de communication majeurs (voir figures 4.2 et 4.3) sont définis :

- Utilisateur/utilisateur : par opposition à une qualité de service au niveau d'une même cellule.
- Utilisateur/serveur de données : par opposition aux applications collaboratives.

Une politique de gestion de la qualité de service est nécessaire pour que le système puisse déterminer s'il pourra satisfaire le niveau de qualité de service demandé.

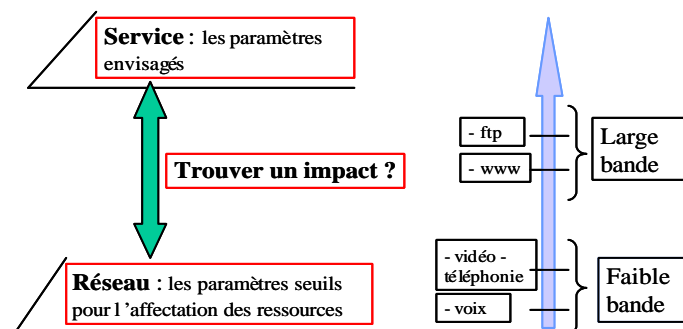


Figure 4.2 Impact entre les paramètres mesurés au niveau de service et réseau

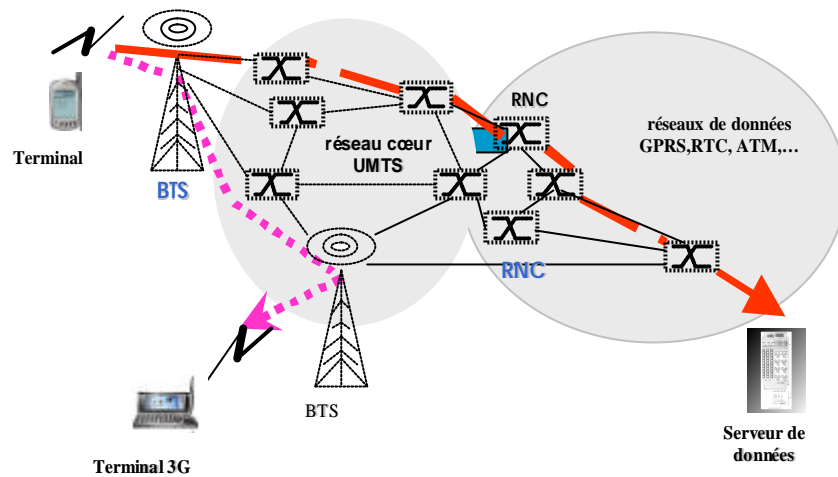


Figure 4.3 Qualité de service de bout en bout

4.4 Suivi et supervision de la qualité de service

Une fois le réseau cellulaire déployé et mis en service, intervient la phase de l'exploitation et de la maintenance. Cette phase permet à l'opérateur de débiter un nouveau cycle qui comprend le suivi de la qualité de service, ainsi que l'optimisation du réseau. La procédure de handover influe directement sur la qualité de service. C'est dans ce cadre que la 3^{ème} partie de notre travail s'intéresse à l'optimisation de cette procédure afin d'améliorer et d'assurer la qualité de service de l'allocation des ressources réseau.

4.4.1 Indicateurs de qualité de service

L'analyse des indicateurs permet le suivi de la qualité de service. En effet, ces indicateurs permettent la localisation des anomalies dans le réseau et par la suite, l'identification des causes de ces problèmes afin de faire les actions correctives nécessaires.

On peut citer deux types d'indicateurs :

- Les indicateurs globaux : qui résument l'efficacité de tout le réseau. Ils sont employés pour la quantification globale du réseau et pour l'estimation de l'impact d'une mauvaise qualité sur le client. Ils permettent aussi la comparaison entre les réseaux (concurrence, etc.).
- Les indicateurs intermédiaires : qui nous renseignent sur l'efficacité des services intermédiaires du réseau. Ce sont des indicateurs globaux. Ils permettent la détection, l'identification et la localisation des problèmes dans le réseau, ainsi que l'identification des causes.

4.4.2 Concept de qualité de service

La qualité de service dans les réseaux 3G est l'effet global produit par la qualité de fonctionnement de ses services qui détermine le degré de satisfaction de l'utilisateur [LAR01]. Pour permettre une qualité de service acceptable, il y a plusieurs critères à ajuster, dont les plus importants sont les suivants :

- La couverture : la détection de ce problème s'explique par l'inexistence des barrettes sur l'écran du terminal, de ce fait, il ne peut pas être détecté par le système mais peut être évalué par les plaintes des abonnés et par les mesures radio. Les actions peuvent être, une diminution dans le nombre des sites, une mauvaise configuration du réseau (position des sites, types d'antennes, direction et hauteur), un problème d'installation (pertes des puissances dans les câbles) ou des problèmes de maintenance.
- Le taux d'appel réussi : la diminution de cette valeur implique que les utilisateurs ne peuvent pas établir une communication, ce problème est évalué par l'opérateur grâce aux mesures radio.

Les actions de l'échec d'établissement d'appels s'expliquent par le niveau d'accès minimum dans la cellule, l'incohérence entre dimensionnement des ressources et trafic réel, l'interférence, et la mauvaise couverture radio.

4.4.3 Paramètres de réseau

L'ajustement des paramètres de travail est une tâche essentielle lors de la mise en exploitation du réseau. Il permet l'activation ou la désactivation de certaines fonctionnalités pour le maintien de la qualité et l'optimisation du réseau.

Il existe deux types de paramétrage :

- Les paramètres constructeurs (ou fournisseurs d'équipements) : ce sont des paramètres système (activation de certaines fonctionnalités telles que le chiffrement, le contrôle de puissance, etc.) préconisés par le constructeur et sont, aussi, relatifs à l'équipement (version de logiciel, etc.).
- Les paramètres ingénierie : ces paramètres sont à l'initiative des opérateurs. Ils sont définis et/ou modifiés au niveau de l'OMC (*Operating Maintenance Center*).

L'optimisation de ces paramètres est un processus délicat mais constitue une tâche essentielle pour le maintien d'une qualité de service acceptable surtout suite à des modifications de certaines fonctionnalités ou services.

4.4.4 L'OMC (*Operating and Maintenance Center*)

Le système d'exploitation et de maintenance OMC est connecté aux MSC et BSC et possède les fonctions suivantes :

- L'accès à distance à tous les éléments qui composent le réseau (BSS, MSC, VLR, etc.) ;
- La gestion des alertes et de l'état du système avec la possibilité d'effectuer différentes sortes de test permettant l'analyse des prestations et la surveillance de la qualité de fonctionnement de ce dernier ;
- Le stockage de toutes les données relatives au trafic des utilisateurs, nécessaires à la facturation ;
- La supervision du flux du trafic dans les centrales et l'introduction de changements éventuels dans le même flux ;

- La gestion des utilisateurs et la possibilité de localiser leur positions à l'intérieur de la zone de couverture.

4.5 Conclusion

Au cours de ce chapitre, nous avons défini les critères de performance réseau tels que la probabilité de refus d'établissement de nouvelle connexion, la probabilité de coupure d'appel, le taux de handover, la capacité de réseau, etc. Pour bien dimensionner le réseau radio mobile multimédia et analyser ses performances, ces critères seront calculés et évalués dans une zone de service opérationnelle (cf. chapitre 7 et 8). Le calcul sera basé sur un modèle de mobilité développé dans les modèles de trafic choisis au niveau du chapitre 3.

Les critères de classification des paramètres de la qualité de service sont présentés au niveau du paragraphe 4.3. Dans certains cas, le réseau ne pourra pas fournir la qualité de service demandée par l'utilisateur. Il s'agit alors de voir s'il vaut mieux refuser à l'utilisateur l'accès au réseau ou lui délivrer un service mais avec une qualité de service dégradée.

Pour satisfaire l'utilisateur avec un niveau de service acceptable, nous allons utiliser dans les chapitres 8 et 10 des mécanismes de contrôle d'admission d'appel pour la gestion de la qualité de service de l'allocation de ressources. Ces mécanismes consistent à utiliser les modèles de trafic retenus au niveau du chapitre 3 tels que la voix, vidéo, ftp, etc. et à regarder l'impact sur la satisfaction objectif de l'utilisateur d'une qualité de service dégradée. Un des résultats de ce contrôle pourra être le relâchement des contraintes nécessaires en terme de qualité de service, héritées des réseaux fixes, dans lesquels les ressources n'étaient pas aussi limitées que dans les réseaux mobiles.

Ces mécanismes de contrôle d'admission d'appel seront alimentés par un protocole qui informe le service d'une possible dégradation (cf. chapitres 8, 9 et 10). Le service peut alors passer en mode dégradé si possible ou bien changer ses paramètres de qualité de service. Ce protocole peut être vu en deux phases :

- les ressources nécessaires à la qualité de service sont-elles disponibles ?,
- une phase de négociation avec le service.

Références Bibliographiques

- [BEL00]Bellili, H., *"Radio resource allocation methodology for GSM multimedia networks,"* PhD thesis, U.Paris, 6-15 September 2000.
- [CHE94]Cheung, J.C.S. and al., *"Network planning for third generation mobile radio systems,"* IEEE Communication Magazine, November 1994, pp. 54-59.
- [CEN95]Cengiz, C., *"GSM to UMTS / a perspective on the future,"* ICT'95, April 1995, pp. 351-358.
- [ELI96] Elias-Bejjani, J., *"Supervision of QoS of real time distributed systems,"* PhD thesis, ENST Paris, 24 october 1996.
- [ERI99]Eric, G.S., *"QoS and networks,"* CNAM - CEDERIC, December 1999.
- [LAG01]Lagrange, X., *"Services and architecture of GSM,"* Dept. RSM, V.1.0, ENSTB 02 february 2001.
- [LOU01]Laussert, A. and al., *"Radio planification of UMTS network : 3G networks study,"* UMTS team, C2R/ASER/SC/0233/2001.
- [MAN00]Maniatis, S. and al., *"End to End quality of Service Issue Over Next Generation Mobile Internet,"* IEEE 2000.
- [MIH98]Mihailescu, C. and al., *"Resource allocation for integrated voice/www traffic in UMTS/TDD System,"* 1998.
- [OTH99]Othman, J.B. and Valois, F., *"Multiservices allocation in hierarchical cellular networks,"* (MAHCN); 4th IEEE Internaional Symposium On Wireless Communications (ISWC'99), Victoria (Canada), June 1999, pp. 66-78.
- [RAU98]Rauhala, K., *"Wireless ATM base line text,"* Technical report, ATM Forum, San Diego, CA, April 1999.
- [TDF95]TDF C2R, *"Storm-Software tools for the optimisation of ressources in mobile systems,"* Technical report, ACTS-EUROPEAN Project, 1995.
- [TABB97]Tabbane, S., *"Location management methods for third-generation mobile systems,"* IEEE Communication Magazine ; August 1997
- [VAL99]Valois, F. and al., *"Performance modelling of hierarchical networks using pepa,"* Process algebra and performance modelling workshop 1999 (PAPM'99), Zaragosa, Spain, September 1999, pp. 139-154.

Chapitre 5

**Configuration d'un réseau radio mobile
de bout en bout****5.1 Introduction**

Le monde des télécommunications est en pleine mutation du fait de nouveaux besoins, notamment dans le domaine du haut débit, de la téléphonie mobile et de l'internet. Pour répondre à ces besoins, les offres de services attendues par les utilisateurs vont induire de profonds bouleversements techniques sur les architectures de réseaux.

Les opérateurs de télécommunication exploitent une mosaïque de réseaux hétérogènes, dont la gestion globale est complexe. Les différents modes de gestion seront organisés par service. L'architecture idéale de ce type de réseaux vers laquelle les opérateurs souhaitent s'orienter, se traduit par le passage d'un réseau par service à un réseau unique à haut débit pour tous les services. Dès lors, l'enjeu, pour ces opérateurs, est d'offrir de nouveaux services à faible coût en comparaison aux autres opérateurs.

Ainsi, les réseaux NGN (*Next Generation Networks*), réseaux de prochaine génération, doivent permettre de raccorder tout client, quel que soit son mode d'accès et d'assurer tout type de service. Les utilisations et les besoins des utilisateurs se résument par une transmission dans d'excellentes conditions de qualité de service de la voix, des données et des images animées. Les particularités de l'architecture de réseau NGN reposent sur un transfert par paquets associés à l'utilisation de fibres optiques, une intelligence dans la signalisation séparée du transfert et des interfaces ouvertes

standardisées. Le NGN va aussi permettre de généraliser la mobilité du terminal, des services et de la personne. Ainsi, pour préparer cette évolution, un certain niveau de développement stratégique doit être envisagé comme la mise en œuvre d'un réseau paquet à haut débit et d'un système d'information économique et performant.

Dans ce chapitre, nous nous intéressons en particulier à la description générale des réseaux UMTS-NGN tels qu'ils ont été spécifiés au niveau des recommandations du 3GPP. Cette description a pour but de se familiariser avec les spécificités de ce type de réseau pour lequel nous avons développé un protocole ainsi qu'un mécanisme de la gestion de la qualité de service de l'allocation des ressources de bout en bout de services multimédia (cf. chapitre 9 et 10).

5.2 Evolution des réseaux radio mobiles vers le multimédia tout IP

Le *release 5* des spécifications élaborées par le 3GPP, définit une architecture de référence divisée en trois domaines pour le réseau UMTS [IMC02] :

- Le domaine de commutation de circuit (CS) ;
- Le domaine de commutation de paquet (PS) ;
- Le domaine multimédia IP (IM).

La nouvelle architecture UMTS introduite par la *Release 5* correspond à une architecture «tout IP». Ceci veut dire que le protocole IP est utilisé non seulement dans le réseau mais filaire aussi au niveau du réseau radio. Cette version présente en effet, un cas particulier du NGN qui se base sur la mobilité et le multimédia tout IP. Cette nouvelle version de réseau est basée sur une séparation entre les fonctions de contrôle et de transport pour le service. Ainsi, on distingue deux plans au niveau de l'architecture UMTS-NGN : le plan de contrôle et le plan de transport.

A notre avis, en se basant sur les recommandations du 3GPP, Il existe globalement trois étapes pour l'évolution de l'UMTS vers le multimédia tout IP :

- Première étape : lancement de l'UMTS et prise en charge des terminaux 3G, sur la base de spécifications stabilisées du 3GPP. Le déploiement de la troisième génération doit être rapide avec une large couverture réseau et un réseau de base multi-fonctionnel (domaines circuit et paquet). Cette architecture doit assurer la continuité entre les services de deuxième et de troisième génération, afin d'attirer des nouveaux utilisateurs.
- Deuxième étape : extension du réseau de base 3G vers le tout IP pour accroître le marché de terminaux et permettre l'introduction des nouveaux services multimédias.
- Troisième étape : achèvement du tout IP et remplacement du parc installé par les nouveaux terminaux IP. A ce stade, l'infrastructure en mode paquet et la solution multimédia introduite au cours de l'étape précédente sont étendu progressivement pour tous les utilisateurs. Dans cette étape, les terminaux IP utiliseront, au niveau des applications, des protocoles axés sur l'IP, comme le protocole d'établissement de session SIP. Ces terminaux sont compatibles voix et données.

5.3 Approvisionnement en ressources en UMTS-NGN

Différents stratégies d'approvisionnement en ressources sont utilisés dans les réseaux de télécommunications pour les services multimédias. Ces stratégies dépendent de la proportion de trafic prévu en temps réel et sa répartition pour les services multimédias. Les stratégies d'approvisionnement couramment utilisées sont : le sur-approvisionnement, le contrôle souple et l'admission stricte.

5.3.1 *Le sur-approvisionnement*

En l'absence de technologie de qualité de service pour différencier les paquets multimédia des paquets concernant d'autres types de trafic, les applications temps-réel sont prises en charge en fournissant une capacité de réseau dépasse largement la charge de trafic totale proposée. Cette solution peut notamment convenir à des fournisseurs de réseaux de transport dont l'activité consiste à vendre de la bande passante. La gestion des ressources consiste en des mesures de contrôle et de surveillance. Ainsi, l'opérateur peut augmenter la capacité avec le temps ou réduire le sur-approvisionnement nécessaire au respect des garanties promises.

5.3.2 *Le contrôle souple*

Cette deuxième stratégie consiste à sur-approvisionnement le réseau uniquement pour le trafic multimédia temps réel. Cette stratégie nécessite un mécanisme de hiérarchisation des paquets dans les éléments de réseau. Par exemple, l'hiérarchisation des paquets dans les éléments de réseau peut être effectuée à l'aide du DSCP (le code d'accès aux services différenciés). Cette stratégie part de l'hypothèse que le trafic interactif ne présente qu'une proportion faible du trafic global et que le contrôle d'admission par flux en temps réel n'est pas indispensable. Toutefois, elle impose une surveillance des flux de trafic en temps réel en périphérie de réseau afin de protéger les précieuses ressources du réseau. Un ensemble de mesures de trafic et une surveillance peuvent améliorer encore cette solution.

5.3.3 *L'admission strict*

Cette troisième stratégie utilisée au niveau du système UMTS-NGN repose sur l'approvisionnement strict en ressources et sur le contrôle d'admission par flux multimédia en temps réel. Cette solution s'impose dans les scénarios où la charge multimédia en temps réel représente une part importante de la capacité totale. Ceci est le cas par exemple dans les réseaux d'accès UMTS-NGN. Comme pour la stratégie de contrôle souple, il faut différencier les paquets en terme de service dans les routeurs. Il faut aussi surveiller le trafic temps réel à la périphérie. Pour réaliser le contrôle d'admission par flux, cette stratégie nécessite également l'inter fonctionnement de la signalisation d'appel et de l'approvisionnement en ressources. Deux mécanismes permettant la mise en œuvre de l'admission stricte : l'administration de la bande passante et la signalisation dans la bande.

5.3.3.1 *Gestion de la bande passante*

L'administration de la bande passante est assurée grâce à la signalisation inter-domaines. Elle permet :

- Le contrôle d'admission de bout en bout ;
- La localisation du terminal correspondant ;
- L'analyse du profil et des ressources du destinataire ;
- La négociation du type de média et des paramètres de communication ;
- La vérification de la disponibilité du correspondant ;
- L'établissement et le suivi de l'appel,...

5.3.3.2 Signalisation dans la bande

Le contrôle strict d'admission repose sur l'utilisation d'un protocole de signalisation dans la bande pour la réservation des ressources. L'activation du contexte de protocole de données en mode paquet PDP pour l'UMTS-NGN est un exemple de signalisation de réservation de ressources dans la bande. Le protocole de signalisation dans la bande réserve des ressources dans tous les éléments du réseau se trouvant sur les chemins des données. Chaque élément du réseau affecte le contrôle d'admission de ressources par flux. Cette méthode impose la synchronisation et l'entrelacement de la signalisation d'appel et de réservation des ressources dans la bande. Le réseau UMTS-NGN a besoin de cette fonctionnalité car le trafic multimédia temps-réel est appelé à représenter une part importante du trafic global. De plus les ressources radio en UMTS-NGN sont rares (c'est d'ailleurs le cas de tous les systèmes de téléphonie radio mobile), ce qui rend la signalisation de réservation des ressources dans la bande, indispensable et insuffisante [SS7][IPM02].

Le système UMTS-NGN est basé sur une architecture «tout IP». La réservation des ressources dans ce système se fait donc de bout en bout. Par conséquent, le domaine DiffServ est étendu entre l'accès au premier réseau de transport et le terminal de l'utilisateur.

Le conditionnement de trafic à la frontière d'un domaine DiffServ avec le système UMTS-NGN suppose le rejet éventuel de l'excès en bit ou en paquet du flux du service tel que spécifié dans la *release 5* du 3GPP. Cet excès dépend des ressources physiques disponibles sur les réseaux de transport. L'abandon des bits et des paquets en excès, par rapport à la bande passante allouée au service à délivrer, engendre une dégradation de la qualité de service. La gestion de l'allocation des ressources dans la bande est donc nécessaire pour résoudre ce problème.

5.4 Correspondance entre qualité de service IP et qualité de service UMTS

Les applications et les services utilisant les protocoles basés sur IP connaissent une grande popularité car l'usage d'un réseau IP et d'un système de communication mobiles offre l'opportunité d'utiliser des services spécifiques à la mobilité avec une meilleure adaptation à la transmission de données (vidéo-téléphonie, www, ftp, etc.).

Cependant, la gestion de la qualité de service reste un problème majeur pour les opérateurs réseaux mobiles. Ces opérateurs souhaitent aujourd'hui proposer à leurs clients des services dont la qualité de service soit connue à l'avance et contractualisée. Il n'est probablement pas possible d'utiliser dans ces grands réseaux une technique de réservation des ressources qui impose à chacun des éléments du réseau de mémoriser des informations sur tous les flux applicatifs qui le traversent et de ne considérer que ses propres contraintes. De cette constatation est née l'approche dite de services différenciés

(DiffServ) dont l'une des avantages est d'offrir des niveaux d'abonnement différents aux usagers.

DiffServ est un mécanisme de qualité de service spécifié par l'IETF dans les RFCs 2474 et 2475 [3GPP.853] afin de contrôler les flux de trafic dans le réseau. Le PHB (*Per Hop Behavior*) spécifie l'utilisation d'informations dans l'en-tête IP pour indiquer la classe de qualité de service désirée. Sa complexité est repoussée aux extrémités routeurs du réseau où le trafic est moins important. Il permet l'établissement d'une qualité de service personnalisée selon les besoins du trafic, et utilise moins de signalisation ce qui le rend plus souple.

D'un point de vue global, l'architecture de la qualité de service de bout en bout comprend deux entités : le réseau de transport IP, qui est un réseau mondial, et le réseau d'accès UMTS. La correspondance de qualité de service des deux mondes est définie par l'opérateur [3GPP.853][QoS99] (voir figure 5.1) :

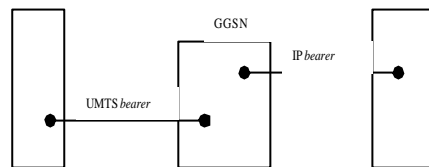


Figure 5.1 Mapping de QoS IP et QoS UMTS

D'une manière plus détaillée, l'architecture de QoS de bout en bout est donnée par la figure 5.2. Un service support représente une capacité de transmission entre deux interfaces avec, des caractéristiques techniques de débit, de taux d'erreur, etc. La mise à disposition d'une liaison point à point numérique à 9,6 kbps constitue un exemple simple de service support. Un service support fourni peut être vu comme un canal permettant d'échanger des informations. Il faut en préciser les caractéristiques et l'interface qui permet d'invoquer ses opérations. La qualité de service est ainsi également liée au service support utilisé dans la transmission.

Il en résulte, qu'il existe différents niveaux pour la définition et la gestion de la qualité de service. Ces niveaux dépendent essentiellement de l'architecture du réseau et sont illustrés par la figure 5.2. Ainsi, par exemple, la qualité de service dans l'environnement personnel de service d'un utilisateur n'est pas la même qu'au niveau application, car le plus important à ce niveau est de connaître les paramètres qui gèrent la qualité de service de bout en bout.

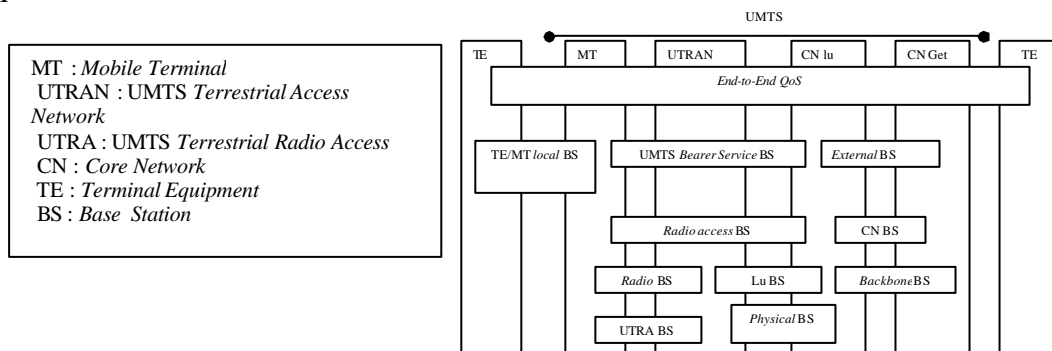


Figure 5.2 Architecture de bout en bout

La gestion de la qualité de service pour le TE sera réalisé en utilisant le service de TE/MT *local Bearer Service*, le service de UMTS BS, et le service de *External BS*. Signalons que les services de TE/MT BS et *External BS* ne sont pas au niveau du réseau UMTS. L'opérateur UMTS offre seulement les divers services offerts par UMTS BS. C'est ce service qui fournit la qualité de service UMTS.

5.4.1 Paramètres de QoS bout en bout

La qualité de service est généralement assimilée à la discrimination des services, autrement dit à la définition de classes différenciées de service. Mais cela signifie aussi une garantie de service et pour cela une réservation des ressources. Pour implanter ces classes ou garantir une disponibilité des ressources, il faut définir une ou plusieurs politiques sur les nœuds du réseau considéré permettant d'implanter la qualité de service demandée à l'échelle d'un réseau en utilisant par exemple le mécanisme de contrôle d'admission dynamique.

5.4.2 QoS de bout en bout : plan de transmission

L'objectif de la gestion de la qualité de service de bout en bout est de maintenir les caractéristiques et les engagements du flux de données vis-à-vis du profil de qualité de service négocié. La gestion de la qualité de service lors du passage d'un plan à un autre se décompose en quatre entités fonctionnelles principales : *Mapping*, *Classification*, *Resource Manager* et *Traffic conditionner* (voir figure 5.3) :

- *Mapping Function* : assure le marquage de chaque paquet d'information avant la transmission vers une autre partie du réseau pour que le paquet reçoive un traitement destiné au transfert par le service transport.
- *Classification Function* : oriente les paquets vers les services dans le cas où il y aurait plusieurs services de transport UMTS établis. Cette fonction se base sur l'en-tête IP.
- *Resource Manager* : gère et distribue les ressources disponibles entre les services partageant la même ressource. Ordonnancement, allocation, contrôle de puissance, etc. Le gestionnaire de ressource distribue les ressources selon la qualité de service.
- *Traffic Conditionner* : mesure, marque et met en forme le trafic suivant la qualité de service négociée.

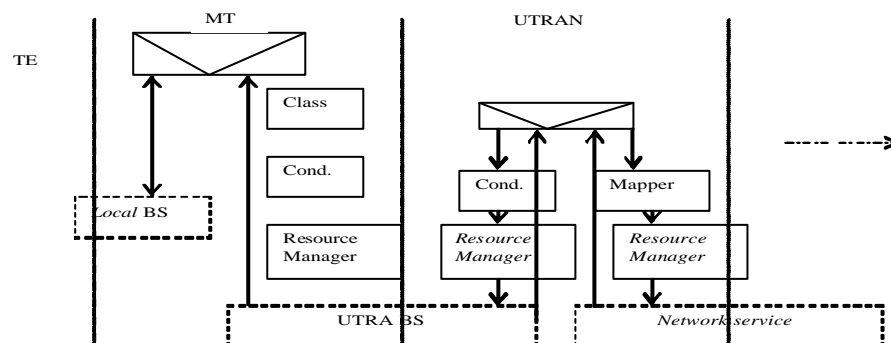


Figure 5.3 Gestion bout en bout de QoS

5.4.3 Délai bout en bout dans l'UMTS/IP

Le délai de bout en bout est le paramètre le plus intéressant pour les applications temps- réel à cause de l'interactivité. Pour que la convention soit acceptable, il faut que la durée de transfert ne dépasse pas 300 ms. Et pour optimiser ce temps, les travaux se concentrent sur le réseau IP qui doit offrir le meilleur temps possible avec garantie de service (*Qbone*) (voir figure 5.4).

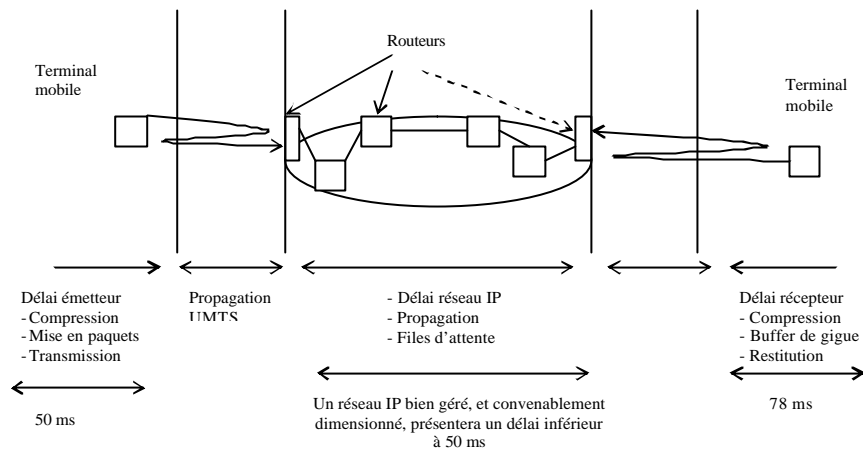


Figure 5.4 Délai de transfert de bout en bout sur UMTS/IP

5.5 Architecture générale d'un réseau UMTS-NGN à base de règles

La figure 5.5 illustre l'architecture générale d'un réseau [ELL02] adoptant des règles pour un contrôle de qualité de service de bout en bout. Le serveur de règle intégrant un PDP (*Policy Decision Point*) aux sources d'information dont il dispose (serveur de temps, serveur d'allocation des ressources BB (*Bandwidth broker*)) pour prendre une décision qu'il achemine ensuite vers le nœud du réseau intégrant le PEP (*Policy Enforcer Point*). Il existe deux modèles de gestion dans l'architecture de réseau à base de règles :

- Le modèle *Outsourcing* dans lequel le PEP fait appel au PDP pour répondre à un événement qui doit être résolu sur la base des règles dont le PEP ne dispose pas. Ce genre de situations est typiquement associé aux protocoles de signalisation de bout en bout (RSVP : *Resource Reservation Protocol*).
- Le modèle *Provisionning* suppose que le PDP prépare la configuration d'un équipement à partir de nouvelles règles d'utilisation du réseau (changement des règles de gestion, etc.). Le PDP charge ensuite les informations de configuration dans le PEP. Ce modèle est utilisé pour contrôler l'utilisation du réseau et configurer l'architecture présentée dans la figure 5.5. Il fait apparaître les fonctions suivantes :
 - le stockage des données et l'interrogation des bases de données,
 - l'interprétation des règles et la prise de décision,
 - la mise en application de la décision.

Ces fonctions sont supportées par les deux composantes principales de l'architecture : le serveur de règles (PDP) et l'équipement de réseau (PEP : les routeurs)

relié au serveur par un protocole COPS (*Common Open Policy Services*) (voir paragraphe 5.3.3). C'est un protocole retenu par l'IETF (*Internet Engineering Task Force*) pour la liaison entre le serveur de règle et ses clients.

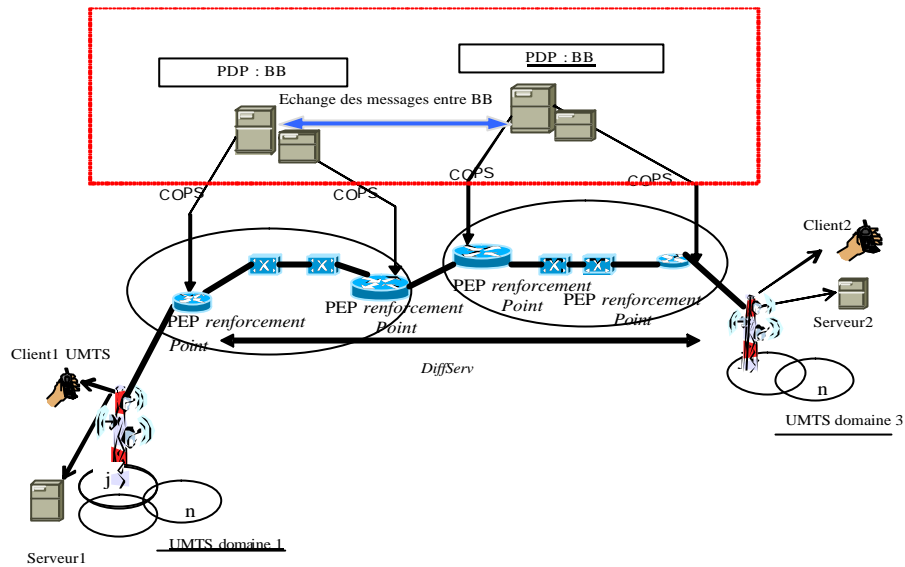


Figure 5.5 Architecture générale d'un réseau à base de règles

5.5.1 Le serveur d'allocation des ressources

Pour traiter le problème de gestion des ressources, un agent (un serveur) BB [ELL02] communique avec le serveur de règle PDP pour prendre une décision. Il possède les fonctions suivantes (voir figure 5.5) :

- Implantation des règles d'accès au réseau ;
- Surveillance de l'allocation des ressources pour un trafic marqué ;
- Interprétation et authentification des requêtes pour le marquage de trafic et l'allocation de ressources de bout en bout ;
- Information des routeurs de périphériques, de la présence de flux marqués et leurs caractéristiques ;
- Communiquer avec les BBs d'autres domaines.

Cet agent connaît l'état des ressources de la zone où il est implanté. Pour mettre en place un service reposant sur une classe donnée, une requête est envoyée vers cet agent qui peut soit l'accepter soit la refuser en fonction de l'authentification de la requête et des ressources disponibles. Les allocations fournies par le BB peuvent être statiques, dynamiques ou mixtes. Les différentes solutions sont les suivantes :

- Allocation statique, où les profils de trafic marqués sont prédéfinis et pré-alloués statiquement entre tous les domaines adjacents par des accords bilatéraux, sans nécessité de négociation entre BB adjacents quand les utilisateurs désirent émettre du trafic.
- Allocation statique avec interaction entre BB, solution plus souple que la précédente, permettant à chaque BB d'informer son voisin du taux d'utilisation de cette allocation statique.

- Allocation dynamique, très souple mais difficile à implanter, car l'allocation se fait au moment de la demande (facturation appel par appel).
- Solutions mixtes.

Le PDP devient une partie du BB (voir figures 5.6), et fournit les informations nécessaires aux PEP (routeur d'entrée ou de sortie d'un domaine) pour supporter les flux authentifiés et autorisés.

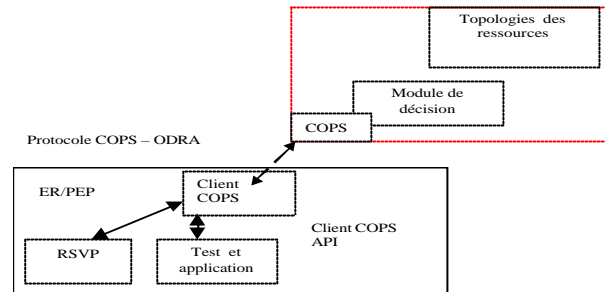


Figure 5.6 Les modules software

Les décisions prises par le serveur de règles, dans le cadre du modèle *provisioning*, sont véhiculées vers les équipements du réseau chargés de les mettre en application. Ces équipements, qualifiés de PEP ou client, mettent en œuvre les outils dont ils disposent pour faire respecter les décisions : la réservation de bande passante, la priorité de trafic, etc.

5.5.2 Le serveur de règle (Policy Decision Point)

Le serveur de règle de la figure 5.7 tient le rôle de PDP. Il interagit avec les équipements du réseau dont il gère le comportement : routeurs, *switches*, équipements VPN, et autres PEP.

Pour prendre sa décision, le serveur de règles interprète les données qu'il consulte dans les différentes bases BBs. A partir de cet ensemble de données, l'interpréteur du serveur de règle formule une décision en réponse à la requête d'un PEP. Le PEP doit en demander le renouvellement ou attendre qu'un nouvel événement se présente et émettre une autre requête vers le serveur.

Dans le cadre du module *provisioning*, le serveur de règle doit spontanément adapter le comportement des équipements du réseau aux évolutions qu'il peut constater, soit dans les règles à appliquer, soit dans les données provenant du réseau.

Le serveur de règles doit faire face à différentes classes de règles: CoS/QoS (*Class of Service/Qualité of Service*), discrimination d'applications, authentification des usagers, etc.

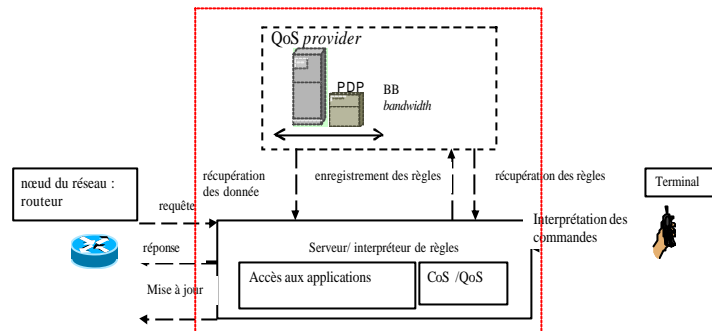
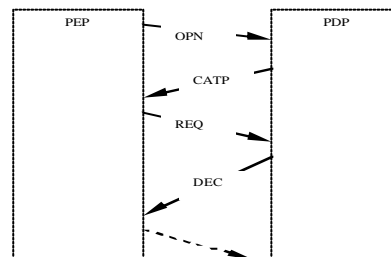


Figure 5.7 Fonctionnalité du serveur de règles

5.5.3 Le protocole COPS

COPS est un protocole extensible et simple de question/réponse spécialement conçu pour l'échange d'informations entre le serveur de règle et son client. Chaque message COPS renferme des objets d'informations dont le contenu est interprété en fonction du type de client COPS (par exemple le client RSVP).

Avant tout échange de données, le PEP doit ouvrir une connexion vers son PDP responsable et s'identifier. Il peut alors générer sa requête à laquelle le PDP répondra par un message de décision. Lorsque le PEP a installé la nouvelle configuration, il doit en avertir le PDP. Le serveur de règles peut ensuite modifier ou mettre à jour les informations de configuration par de nouveaux messages de décision. A chaque décision, le PEP efface la configuration nommée et renvoie une confirmation au PDP (voir figure 5.8).



OPN : client OpenN, le PEP ouvre une connexion TCP avec le PDP ;
 CAT : Client Accept, le PDP accepte la connexion ;
 REQ : REQuest, le PEP envoie une requête au PDP. La demande contient un identificateur ainsi que des objets politiques nécessaires pour la prise de décision politique du PDP ;
 DEC : DECision, le PDP envoie la décision politique par un message DEC.

Figure 5.8 Exemple de communication entre le PEP et le PDP

Le contexte de chaque requête dépend de ce qui l'a déclenchée. COPS identifie plusieurs types d'événements susceptibles de déclencher une requête :

- L'arrivée d'un message (contrôle d'admission) ;
- L'allocation des ressources ;
- Une demande d'informations de configuration donnée, etc.

Dans l'exemple représenté par la figure 5.9, le serveur de règle a établi la règle suivante : à l'instant t_1 le mobile M ne peut utiliser son application vidéo qu'avec un

débit bas D_1 , vu l'indisponibilité des ressources dans le réseau parce que la charge de trafic est importante.

Lorsque le mobile M tente de lancer son application vidéo avec un débit $D_3 > D_1$, il déclenche le mécanisme suivant :

- Le routeur réceptionne les paquets et s'aperçoit en examinant l'adresse IP et le numéro de port qu'il s'agit d'un flux vidéo émanant du terminal M ;
- Le client COPS intégré au routeur transmet une requête au serveur de règles pour vérifier l'autorisation d'admission ;
- Sur réception de la requête, le serveur de règle interroge le serveur BB pour réserver une ressource D_3 ;
- Une fois que ces données obtenues, le serveur de règle les combine et formule une décision ;
- L'interpréteur émet le résultat vers le routeur par l'intermédiaire du protocole COPS ;
- Sur réception de la décision, le routeur refuse l'application de mobile M et essaye de négocier pour la satisfaire avec un débit supérieur.

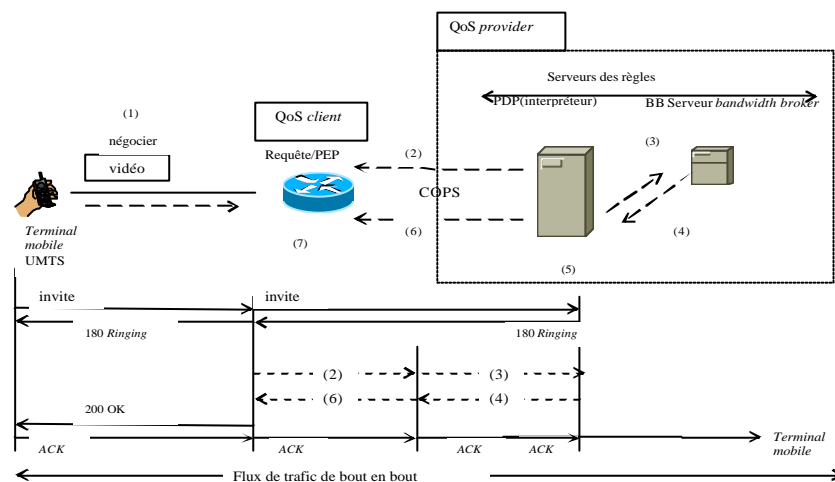


Figure 5.9 Exemple d'un PEP muni du client COPS

La structure des messages COPS

Un message COPS encapsule une série d'objets [TRA02]. L'en-tête du message définit principalement le type de client, le type de l'opération concernée par le message et la longueur totale de celui-ci. La structure du message est représentée sur le tableau suivant :

<i>Request (REC)</i>	<i>Client- Open (OPN)</i>
<i>Decision</i>	<i>Client- Accept (CAT)</i>
<i>Report State</i>	<i>Client- Close (CC)</i>
<i>Delete Request state (DRQ)</i>	<i>Keep – Alive (KA)</i>
<i>Synchronize State Req (SSQ)</i>	<i>Synchronize Complete (SSC)</i>

Tableau 5.1 Structure d'un message COPS

Comme nous l'avons remarqué, les serveurs de règles doivent pouvoir entrer en contact avec les serveurs de règles voisins BBs, également avec ceux qui ne supportent pas de protocole de transaction ou utilisent d'autres protocoles. Dans ce cas, l'utilisation d'un module de consensus doit faciliter leur intégration dans l'architecture du réseau (cf. figure 5.5) à base de règles.

La tolérance aux pannes de l'un des serveurs BB est une préoccupation des concepteurs du module de consensus. Il est en effet susceptible de véhiculer des informations de surveillance des équipements du réseau. Pour répondre à cette contrainte, le protocole COPS [TRA02] assure la surveillance de la connexion par des échanges périodiques de messages. En cas d'interruption de la connexion, le module de consensus doit remplir les tâches suivantes :

- Maintenir à jour la liste des équipements dont il est responsable ;
- Véhiculer les décisions du serveur de règles vers les PEP ;
- Surveiller l'état des équipements qui lui sont rattachés.

Lorsque le module de consensus est informé des changements de règles par l'interpréteur, il est responsable de la mise en état des autres serveurs (BBs voisins) en conformité avec les évolutions prescrites (voir figure 5.10).

Régulièrement, le module de consensus interroge le serveur de règles pour connaître les éventuelles modifications liées aux équipements dont il est responsable. L'interpréteur formule une décision et informe le module de consensus via COPS [TRA02]. Le module de consensus traduit la décision en une série de commandes spécifiques aux nœuds et les applique au routeur pour le mettre dans un état conforme à la décision du serveur de règles.

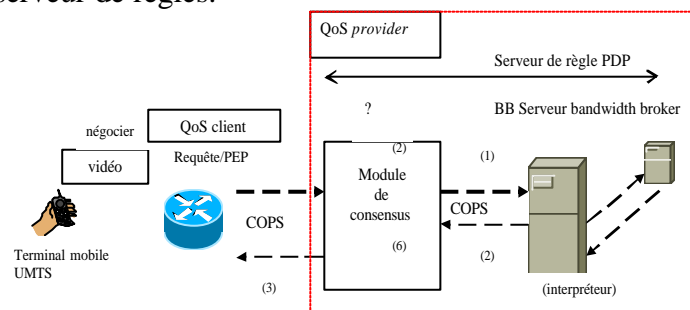


Figure 5.10 Exemple d'un PEP accessible par un consensus

5.5.4 IntServ, le protocole RSVP

L'IETF [WRO97][SHE97] a étudié l'enrichissement du modèle de service *Best-effort* sur les réseaux IP. Le groupe de travail IntServ a proposé le protocole de signalisation RSVP (*Resource Reservation Protocol*), RSVP repose sur une allocation explicite des ressources. Il est utilisé pour assurer une discrimination de service pour les applications sensibles aux délais.

Un hôte utilise RSVP pour demander une quantité de service spécifique au réseau pour une application à flux de données. RSVP transmet la requête à travers le réseau en passant par chaque lien utilisé pour transmettre le trafic. A chaque nœud, RSVP tente de faire une réservation de ressources pour le flux de données. RSVP demande donc une

collaboration étroite entre hôtes et réseau. Tous les nœuds entre le récepteur et l'hôte source doivent donc supporter RSVP.

Quand un routeur RSVP reçoit un message RESV, il utilise un algorithme de contrôle d'admission pour authentifier la demande et allouer les ressources. Si la demande ne peut être satisfaite, le routeur renvoie un message d'erreur au récepteur. Si les ressources sont disponibles, le routeur envoie le message RESV au routeur suivant. Quand le dernier routeur reçoit le message RESV et accepte la demande, il envoie un message de confirmation au récepteur. Quand l'émetteur ou le récepteur ferme la session RSVP, la réservation est libérée explicitement de proche en proche.

La réservation dans chaque routeur se fait d'une manière logicielle d'où la nécessité d'une mise à jour périodique par le routeur. D'autre part, le trafic RSVP peut traverser des routeurs non RSVP. Ceci introduit une faiblesse dans la chaîne qualité de service se traduisant par un routeur au mode *Best-effort* puisqu'il n'y a pas d'allocation de ressources.

Schématiquement, le réseau se décrit comme des sources-destinations reliées entre elles par des routeurs. De façon macroscopique, c'est l'interconnexion de plusieurs types de réseaux : des sites clients (terminaux individuels, réseaux métropolitains ou plaques régionales ADSL, etc.) interconnectés par des réseaux d'opérateurs (voir figure 5.11).

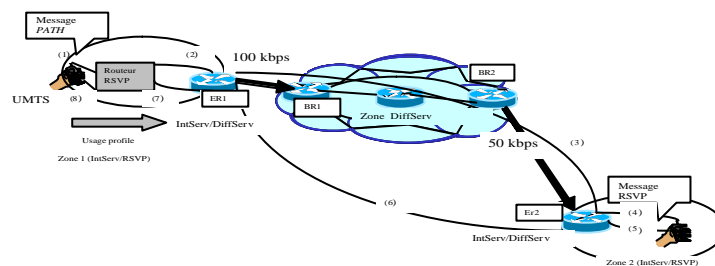


Figure 5.11 Exemple de réseau IntServ

La figure 5.11 illustre un scénario d'acheminement d'un message *PATH* dans le réseau où la visibilité de l'utilisateur est IntServ, alors que le système de communication utilise DiffServ, ce qui rend ainsi IntServ client de DiffServ.

Le déroulement de scénario suit alors les étapes suivantes :

- L'émetteur UMTS A génère un message RSVP-*PATH* contenant les caractéristiques du trafic de l'application ;
- Le message *PATH* permet d'établir les informations d'état et les réservations dans les routeurs de la zone 1 où la garantie de qualité de service est fournie par des méthodes de réservation ;
- Le message *PATH* est acheminé d'une manière transparente à travers le réseau de transit DiffServ pour arriver à la zone 2 ;
- Le message *PATH* arrive au récepteur B en mettant en place les informations de réservation dans les routeurs de la zone 2 ;
- La station B génère un message RSVP-RESV, indiquant un intérêt dans ce trafic offert avec un certain niveau de service. En retournant vers l'émetteur, ce message peut être rejeté dans la zone 2 à cause d'un manque de

ressources. Sinon, les réservations appropriées sont établies dans les routeurs de la zone 2 ;

- Les messages RESV traversent la zone DiffServ en tunnel jusqu'au routeur ER1. A ce niveau, le message est traité par un module DAC (*DiffServ Admission Control*) qui compare les ressources demandées aux ressources disponibles au niveau de service DiffServ correspondant. Si le message RESV est accepté, le DAC (*DiffServ Admission Control*) soustrait la capacité demandée de la capacité disponible ;
- Etablissement des réservations dans les routeurs IntServ de la zone 1 où le message RESV peut également échouer par manque de ressources ;
- L'émetteur UMTS A reçoit le message RESV ;
- L'émetteur UMTS A commence à établir l'appel. Dans ce cas, deux solutions peuvent être envisagées :
 - soit c'est l'émetteur UMTS A qui fixe lui même les bits du champ DSCP en fonction du type de service demandé par RSVP (*controlled load* ou *guaranteed*),
 - soit c'est le routeur ER1 qui classe et fixe le champ DSCP, ce qui est le plus probable. Cette classification se fait en fonction de règles de correspondance établies de façon statique ou via la diffusion d'une règle *ad hoc* par le système de gestion des règles.

Les fonctions nécessaires pour le support des classes DiffServ peuvent être introduites au niveau des routeurs (voir tableau 5.2) [AJM01][BHA01][KHA02], sans ajouter des fonctionnalités nouvelles aux hôtes. C'est là un des points forts de cette approche par rapport à d'autres qui imposent une modification des systèmes d'extrémité.

La figure 5.11 donne un exemple d'architecture de réseau connectée à un autre domaine. Chaque routeur possède des fonctions différentes suivant sa position dans le domaine et applique un traitement particulier aux paquets : à l'entrée du domaine (BR1), à l'intérieur du domaine, à la sortie du domaine en (BR2).

Un routeur (voir tableau 5.2) se voit donc chargé de fonctions nouvelles. Ces fonctions vont de la gestion des configurations et des paramétrages nécessaires au traitement des différentes classes de service, à la cohabitation entre les services DiffServ et IntServ. La signalisation RSVP devra également interagir avec les différentes autres fonctions.

Fonctionnalités des routeurs

- Le routeur d'accès (ER1) : son rôle est d'identifier les flux des données et positionner les codes DSCP, principalement en analysant les informations adresse source, adresse destination, port source et port destination et les répartir dans les classes de service.
- Le routeur interne : son rôle est plus simple. Il n'a plus à classer les paquets, mais seulement à les router et à gérer les files d'attente en sortie.
- Les routeurs frontières (BR1) et (BR2) : doivent agréger les différentes classes de trafic vérifier la conformité de chacun d'entre eux avec les contrats négociés.

	Routeurs frontières (1) E2 et (3) E'2
Fonctions principales	Contrôle de la QoS routeurs <i>Cisco</i>
Débit	20 mbps
Délai	40 ms
Retard	E1-E (2)=5ms ; E(2)-E'2=7ms ((1) – (3) = 15ms)

Tableau 5.2 Les routeurs - paramètres du QoS

5.5.5 DiffServ au sein des réseaux UMTS

En se référant aux spécifications du 3GPP [QoS99], on s'aperçoit que le réseau cœur d'un UMTS est un domaine DiffServ (services différenciés) [NWG98]. DiffServ intervient au niveau de la couche 3 du modèle OSI (couche réseau IP). Le principe de DiffServ est qu'un octet des services différenciés (DS) est contenu dans l'en-tête de chaque paquet IP. Cet octet permet de déterminer et de sélectionner le comportement adéquat (PHB-*Per Hop Behavior*) que chaque nœud traversé du réseau doit adopter vis-à-vis du paquet reçu pour le transmettre au nœud suivant. Ainsi les ressources de réseau sont allouées dans le réseau dorsal pour chaque flux concentré. Toute la complexité est donc déportée sur les routeurs de périphérie. Par exemple, pour le domaine DiffServ qui représente le réseau cœur de l'UMTS Release 1999, le RNC (contrôleur de réseau radio) et le GGSN (*Gateway GPRS Support Node*) font office de routeurs de périphérie.

Routing par étiquetage en UMTS

Le routage multi-protocole par étiquetage MPLS (*Multiprotocol Label Switching*) [RFC01] se situe au niveau des couches 2 et 3 du modèle OSI et intervient au réseau cœur UMTS [SIEM]. Le point fort de l'utilisation de MPLS au niveau du réseau cœur est le découpage entre la technologie de transmission et les systèmes statistiques qui gèrent la QoS. En effet, MPLS offre la possibilité d'utiliser n'importe quel réseau de transport (ATM, Ethernet pour LAN, etc.). MPLS permet de :

- Contourner la dictature de la table de routage ;
- Réserver les ressources sur des chemins choisis ;
- Mettre en œuvre une ingénierie de trafic.

MPLS est basé sur la permutation d'étiquette. La permutation est réalisée en analysant une étiquette entrante qui est ensuite permutée avec l'étiquette sortante. Les étiquettes ne sont imposées sur les paquets qu'en périphérie du réseau MPLS et supprimées à l'autre extrémité.

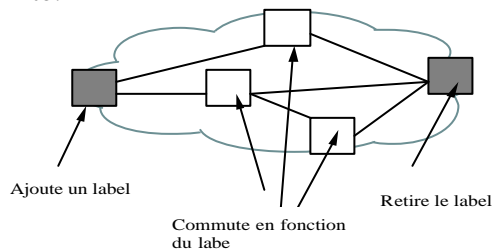


Figure 5.12 Principe de base du MPLS

5.6 Conclusion

L'UMTS est plus généralement les systèmes de 3^{ème} génération, feront sans doute place à des systèmes qui offrent des services plus diversifiés et universels grâce à des interfaces radio qui implémentent la modulation par étalement de spectre. Le réseau cœur de l'UMTS est une évolution du sous système réseau du GSM et du GPRS.

Le système UMTS est basé sur une architecture « tout IP ». La réservation de ressources dans ce système se fait donc de bout en bout. Par conséquent, le domaine DiffServ est étendu entre l'accès au premier réseau de transport et le terminal de l'utilisateur.

Références Bibliographiques

[AJM01]Ajmin, S. and al., “*Weighted fairness guarantee for scalable DiffServ assured forwarding*,” Communications, 2001.ICC 2001.IEEE International Conference on, Vol.8, 2001, pp. 2365-2369.

[BAK02]Bakiras, S. and Victor O.K. Li., “*Efficient resource management for end to end QoS guarantees in DiffServ network*,” Communications, 2002. ICC 2002. IEEE International Conference on, Vol.2, 2002, pp. 1220-1224.

[BHA01]Bhatnagar, S. and al., “*Providing quality of service guarantees using only edge routers*,” Global Telecommunications Conference, 2001. Globecom'01. IEEE, Vol.4, 2001, pp. 2542-2548.

[CHE02]Cheng, Y. and Zhuang, W., “*DiffServ resources allocation for fast handoff in wireless mobile Internet*,” IEEE Communications Magazine May 2002, pp.130-136.

[DAH01]Dahmouni, H., “*Conception and planification of mobile networks: video-telephony application on UMTS/IP*,” work elaborated for Philips Philips Consumer communications society - October 2001.

[ELL02]Ellen, W. Zegura and al., “*QoS routing for anycast communications: motivation and an architecture for DiffServ networks*,” IEEE Communication Magazine, June 2002.

[GUB99]Gubbi, R., “*Multimedia streams and quality of service in the next generation wireless home network*,” Mobile multimedia Communications, 1999. (MoMuC'99) 1999 IEEE International workshop on, 1999, pp. 232-235.

[IPM02]3rd Generation Partnership Project; Technical Specification Group Services and System Aspects, “*IP Multimedia Subsystem (IMS)*,” Stage 2 (Release 5) ; 2002-01.

[IETF][http : //www.ietf.org](http://www.ietf.org).

[IET][http : //www.ietf.org/html.charters/intserv-charter.html](http://www.ietf.org/html.charters/intserv-charter.html)

[IMC02]3rd Generation Partnership Project, “*Digital cellular telecommunication system (Phase 2+) (GSM)*,” Universal Mobile Telecommunication System (UMTS); IP Multimedia (IM) session handling; IM call model; Stage 2 (3GPP TS 23.218 version 5.0.0 Release 5-2002-03.

[KHA02]Khalil, I. and Braun, T., “*Edge provisioning and fairness in VPN-Diffserv networks*,” the 9th International Conference on computer communication and network (ICCCN 2000), 2002, Vol.2, pp. 1220-1224.

[MAN00]Maniatis, S. and al., “*End to End quality of Service Issue Over Next Generation Mobile Internet*,” IEEE 2000.

[NW98]Network Working Group; Request for Comments: 2474, “*Definition of the Differentiated Services Field (DS Field) in the IPV4 and IPV6 Headers*,” December 1998.

[QoS99]3rd Generation Partnership Project; Technical Specification Group Services and System Aspects, “*QoS Concept and Architecture*,” Release 1999.

[RFC01] IETF RFC 3031 (1/2001), “*Multiprotocol Label Switching Architecture*,”

[SIEM]*The siemens UMTS core solution for UMTS*.

[SS7]SS7; [http : //www.techfest.com/networking/wan/ss7.htm](http://www.techfest.com/networking/wan/ss7.htm).

[SHE97]Shenker, S. and al., "*Specification of guaranteed quality of service*," IETF-RFC 2212, September 1997.

[TRA02]Trang Nguyen, T.M. and al., "*COPS-SLS : a service level negotiation protocol for the Internet*," IEEE Communications Magazine, May 2002.

[WRO97]Wroclawski, J., "*Specification of guaranteed load network element service*," IETF-RFC 2211, September 1997.

[3GPP.853]3GPP TR 25.853; http://www.3gpp.org/ftp/Specs/2000-12/R1999/25_series/

Partie 2 : état de l'art

Chapitre 6

**Etat de l'art de la gestion de QoS de l'allocation des
ressources dans un réseau radio mobile****6.1 Introduction**

Les groupes de travaux de l'IETF ont défini un ensemble de protocoles et de mécanismes pour l'amélioration de la qualité de service dans l'UMTS/IP. Le modèle DiffServ sera vraisemblablement supporté par les routeurs de cœur de réseau. Or un chemin peut être constitué d'une séquence de routeurs ayant chacun son propre comportement local et un contrat de service associé. Pourtant, le service fourni, ainsi que le dimensionnement de ce service dépendent de la topologie réelle du réseau et du trafic effectif. Ceci est un aspect positif de DiffServ car il permet aux fournisseurs (et aux constructeurs de routeurs) une marge de flexibilité en ce qui concerne le déploiement des services et les structures tarifaires.

Cependant, cela rend l'évaluation de la qualité de service de bout en bout plus difficile. D'autre part, le dimensionnement et la configuration du réseau sont deux problèmes particulièrement intéressants. Le dimensionnement consiste à déterminer et à allouer (physiquement ou logiquement) les ressources requises aux différents points dans le réseau. La configuration consiste à distribuer les bons paramètres aux équipements réseaux afin de réaliser les objectifs de dimensionnement.

Le dimensionnement et la configuration seront notés sous l'appellation gestion de la qualité de service. Cette gestion peut être statique ou dynamique. Dans le cas statique, la gestion de la qualité de service dans le réseau peut être effectuée manuellement par

l'administrateur de réseau en fonction de la topologie et de la matrice de trafic, ou bien automatiquement via un mécanisme d'allocation invoqué uniquement au moment de l'établissement des contrats avec les clients.

Ce chapitre s'attachera donc à développer et cerner le cadre général de notre contribution au sein de cette thèse. Notre contribution porte sur la problématique de la gestion de la qualité de service de l'allocation de ressources dans un réseau radio mobile multimédia du point de vue dimensionnement et configuration. L'allocation de ressource est effectuée, selon le contrat de trafic, en fonction des paramètres de qualité de service disponibles sur les réseaux de transport et des ressources radio disponibles au niveau du réseau radio cellulaire (cf. chapitres 8, 9 et 10). L'allocation est également effectuée par rapport aux préférences des utilisateurs.

Dans ce chapitre nous commençons par définir le dimensionnement des ressources radio. Il s'agit de déterminer la demande de ressources dans le système. Pour cela, on calcule le trafic offert en Erlang dans chaque cellule par le nombre d'utilisateurs dans la cellule et le taux d'activité par utilisateur. Le dimensionnement des ressources tient compte de la mobilité des utilisateurs (cf. chapitre 7). Ensuite nous présentons les différentes techniques d'allocation des ressources pratiquées au niveau des systèmes radio mobiles de 2^{ème} génération tels que le FCA (*fixed channel allocation*), le DCA (*Dynamic channel allocation*) et le HCA (*Hybrid channel allocation*). La dernière technique peut être appliquée au niveau des systèmes de 3^{ème} génération. Nous terminons ce chapitre par l'état de l'art des travaux de pré-réservation des ressources et de gestion de la qualité de service.

6.2 Dimensionnement des ressources du système

6.2.1 Capacité d'un système cellulaire

La capacité d'un système radio mobile est surtout limitée par manque de spectre fréquentiel. Une première estimation de la capacité d'un système cellulaire consiste à déterminer pour un système régulier le nombre de canaux par unité de surface. La capacité intrinsèque en canaux / km^2 ainsi obtenue se calcule selon [TAB97] :

$$k = \frac{nBW}{KBA} \quad (6.1)$$

où n donne le nombre de canaux par porteuse, K la taille du motif régulier, B la bande de fréquence occupée par un canal duplex, BW la bande de fréquence disponible et A la surface d'une cellule.

Pour un premier dimensionnement d'un système il faut déterminer la demande de ressources dans le système. Pour cela, on calcule le trafic offert en Erlang dans chaque cellule par le nombre d'utilisateurs dans la cellule et le taux d'activité par utilisateur. Le trafic offert dans une cellule i pour le cas d'utilisateurs immobiles est :

$$r_i = \frac{l_i}{m} = \frac{U_i l_u}{m} \quad (6.2)$$

où I_i est le taux d'arrivée d'appels dans la cellule, m le taux de service dans le système, U_i le nombre d'utilisateurs dans i et I_U le taux d'activité par utilisateur.

Le nombre de canaux c peut être déterminé pour une probabilité de blocage d'appel P_b donnée en utilisant la formule d'Erlang-B pour les systèmes à perte :

$$P_b = \frac{\frac{r^c}{c!}}{\sum_{k=0}^c \frac{r^k}{k!}} \quad (6.3)$$

où r est la charge de trafic dans la cellule.

Ce premier dimensionnement du réseau ne tient pas compte de la mobilité des utilisateurs. Un dimensionnement plus fin sera présenté dans le chapitre 7 (cf. paragraphe 7.3.7).

6.2.2 Mobilité des utilisateurs

La mobilité des utilisateurs est la caractéristique majeure des systèmes radio mobiles. Elle permet aux utilisateurs de disposer de service du réseau dans toute la zone de service (et celles d'autres visités) et de communiquer pendant qu'ils se déplacent dans le réseau. Deux fonctionnalités principales des systèmes radio mobiles assurent cette mobilité des utilisateurs, qui sont, la gestion d'itinérance [TAB97] et les handovers intercellulaires.

Il existe différentes causes pour l'initiation d'un handover intercellulaire. Dans le cadre de ce travail, nous nous concentrons sur les handovers dus aux changements de cellules pendant un appel multimédia. Une autre raison justifiant le handover, telle que le maintien de la qualité radio, ne sera pas étudié.

Chaque handover modifie la répartition du trafic dans le réseau. Dans le cadre de ce travail, on étudie l'influence de la mobilité des utilisateurs sur les performances du système.

Plusieurs modèles de mobilité de trafic téléphonique ont été présentés dans la littérature. Dans [THO98], on considère une densité constante d'utilisateurs pour des systèmes macro-cellulaires. Les limites de cette approche pour des systèmes micro ou pico-cellulaires sont démontrées dans [NAN93]. D'autres auteurs [FOS93] proposent des solutions avec des mouvements pseudo aléatoires.

Une autre approche se base sur des modèles stochastiques qui représentent de manière microscopique les mouvements de chaque utilisateur. De telles approches sont analysées par des simulations à événements discrets. Un simulateur de mobilité de véhicules est présenté dans [SES92] pour déterminer les taux de mise à jour de localisation. Dans [STE95], un tel simulateur de mobilité est utilisé pour étudier les performances des réseaux micro-cellulaires.

Dans [LEU98], un modèle déterministe se basant sur les flux de véhicules est présenté pour des réseaux d'autoroutes à une dimension. Ce modèle se limite à étudier de manière macroscopique les dynamiques globales dans le système. Les densités d'utilisateurs et leur taux d'arrivée et de départ sur l'autoroute sont décrits en fonction de leur emplacement et du temps. Dans [MON92] l'exemple d'un réseau cyclique est étudié à partir d'un modèle similaire.

Une version simplifiée du modèle de mobilité décrit dans [BAT98], basée sur le graphe de rue consiste à adapter le trafic routier disponible pour calibrer le modèle. Elle permet d'étudier des systèmes à deux dimensions. Les vitesses des différentes classes d'utilisateurs sont constantes dans l'ensemble du système et les mouvements des mobiles sont indépendants. Le manque des données statistiques sur la mobilité des utilisateurs est un problème majeur pour toute analyse de mobilité. Pour calibrer le modèle, les seuls renseignements dont la version dispose sont les flux de véhicules par jour dans les rues du réseau.

C'est au sein d'un tel contexte que se positionne le chapitre 7, focalisé par conséquent sur le développement d'un modèle de mobilité considéré dans quatre modèles de trafic (voix, vidéo-téléphonie, ftp et www). Le modèle est adapté aux zones urbaines, calibré par des mesures réelles c'est-à-dire des mesures récupérées du centre d'opération et de maintenance (le compteur OMC) prises du terrain et du système à partir des sites existants et des réseaux routiers considérés.

6.2.3 Impact d'une qualité de service dégradée

Les opérateurs doivent bâtir une architecture de réseau qui combine sous réseaux IP et infrastructure UMTS. Dans le cas général, on peut distinguer deux situations :

- Un réseau IP unique que l'on associe au réseau d'accès UMTS ;
- Plusieurs sous réseaux IP définis au sein du même domaine UMTS de l'opérateur.

Dans le premier cas (un seul réseau IP) le problème principal à traiter est le contrôle de la qualité de service au cours des handovers sur l'interface radio.

En effet, plusieurs mécanismes sont nécessaires pour garantir les ressources au terminal mobile lorsque l'utilisateur change de cellule radio. Toutes les ressources réservées à l'application dans une cellule ne sont pas forcément disponibles dans la nouvelle cellule d'accueil. Dans ce cas on peut prévoir un protocole qui informe l'application d'une dégradation possible. L'application peut alors passer en mode dégradé si possible ou alors changer ses paramètres de qualité de service. Ce protocole peut être vu en deux phases :

- Un contrôle d'admission (les ressources nécessaires à la QoS sont-elles disponibles ?) ;
- Une phase de négociation avec l'application.

Dans le deuxième cas (plusieurs sous réseaux IP) le problème du support de la Qualité de service est plus compliqué dans un environnement mobile car les ressources

réservées dans les routeurs intermédiaires le long du chemin pris par les paquets de données doivent être mises à jour. Ce scénario peut nécessiter l'apport de modifications dans les mécanismes actuels de contrôle de qualité de service au niveau IP.

6.3 Différentes techniques d'allocation des ressources radio

Considérant un spectre radio donné divisé en un ensemble de canaux radio disjoints et non-interférents, et que l'on souhaite les utiliser simultanément. Un problème qui se pose est celui de l'allocation de tels canaux et de la réutilisation du spectre fréquentiel. Différentes politiques d'allocation et de gestion ont été développées pour allouer un canal à un appel. Katzela [KAT94] propose une étude des principaux schémas d'allocation des canaux dans le cadre des réseaux mobiles cellulaires qui sont comparés en terme de complexité et de performances. Nous rencontrons deux grands schémas d'allocation de canaux : d'une part l'allocation statique (FCA : *Fixed Channel Allocation*) et d'autre part l'allocation dynamique (DCA : *Dynamic Channel Allocation*). A ces deux schémas vient se superposer la famille des schémas d'allocations hybrides (HCA : *Hybrid Channel Allocation*) visant à prendre le meilleur des schémas dynamiques et des schémas statiques.

6.3.1 Fixed channel allocation (FCA)

Les systèmes de première et deuxième génération sont basés sur une allocation fixe des canaux. Chaque station de base possède un ensemble déterminé de canaux qui lui ont été attribués lors de la phase de planification du réseau, et qui ne changeront pas sauf en cas de reconfiguration de celui-ci (re-planification des fréquences). Cette méthode a l'avantage de la simplicité, et elle permet d'optimiser la distance de réutilisation ainsi que le niveau d'interférences C/I . Toutefois, de par sa nature statique, un tel schéma ne permet pas de répondre aux brusques variations de trafic et à une répartition non uniforme des utilisateurs, ce qui la rend donc essentiellement efficace dans le cadre de trafic spatialement uniforme.

Dans un cas simple, si la distribution des utilisateurs est uniforme le même nombre de canaux est alloué pour chaque cellule. Cette distribution reste efficace si la répartition des utilisateurs ne change pas au cours du temps. Remarquons que dans ce cas, la probabilité de blocage dans le système est alors la même que la probabilité qu'un appel soit bloqué dans une cellule. Mais le trafic dans un système cellulaire n'est pas forcément uniforme et subit des variations spatiales et temporelles, par conséquent, une allocation uniforme peut entraîner des blocages alors que d'autres cellules peuvent être sous utilisées. Pour éviter une mauvaise utilisation des canaux, nous pouvons :

- Soit faire une allocation non-uniforme des ressources dans les cellules : il faut pour cela posséder un modèle fin de trafic attendu ;
- Soit utiliser une stratégie de canal : les canaux non-utilisés des cellules peu chargées sont ré-affectés aux cellules surchargées situées à une distance supérieure à la distance minimale de réutilisation. Cela correspond alors à une redistribution périodique des canaux en fonction de la charge.

Dès qu'une cellule utilise tous ses canaux, elle peut en emprunter un des canaux de ses voisines pour accepter un nouvel appel. Le canal emprunté ne doit pas créer

d'interférences et devient indisponible pour d'autres cellules (*Channel Blocking*). Plusieurs méthodes sont alors possibles :

- Simple : tout canal d'une cellule peut-être emprunté temporairement par une cellule voisine. Plusieurs variantes existent suivant l'algorithme de sélection du canal utilisé. Le plus connu est le simple *Borrowing* où un canal libre peut être prêté à un voisin dès lors qu'il ne crée pas d'interférences. D'autres solutions existent comme le simple *Borrowing Form Richest* où le canal emprunté provient de la cellule qui possède le plus de canaux libres ou encore le SBR-BA prenant en compte le blocage de canal. D'une façon générale, les meilleurs algorithmes en termes de probabilité de blocage la plus faible sont les algorithmes de sélection de canal les plus simples, les moins complexes et satisfaisant la distance minimale de réutilisation.
- Hybride : les canaux affectés à une cellule sont divisés en deux ensembles : d'un côté les canaux locaux dont l'utilisation est réservée pour une cellule et de l'autre, les canaux *empruntables* par les cellules voisines. La méthode la plus simple développée est le *Simple Hybrid Channel Borrowing* où le ratio des canaux locaux sur les canaux empruntables est déterminé par une estimation du trafic, et peut être modifié dynamiquement.

6.3.2 Dynamic channel allocation (DCA)

Notons que les variations temporelles et spatiales du trafic de systèmes cellulaires conduit à rendre les stratégies de type FCA moins efficaces, le recours aux stratégies DCA est tout à fait justifié dans le cas d'une répartition non uniforme du trafic. Ici, tous les canaux sont placés dans un ensemble commun partagé par un ensemble de station de base : il n'y a pas de lien définitif entre un canal et une cellule. Lors de l'arrivée d'un nouvel appel, un canal est alloué dynamiquement en prenant soin que le niveau d'interférences (C/I) soit maintenu à un niveau satisfaisant. L'aspect dynamique de ces schémas d'allocation donne une flexibilité permettant de s'adapter aux variations du trafic. Cependant, ces stratégies sont moins efficaces que FCA sous forte charge et la gestion des canaux au niveau du système est plus complexe.

Les stratégies d'allocation DCA sont séparées en deux familles suivant le type de contrôle employé : soient centralisées (allocation faite par un contrôleur), soient distribués (allocation faite par la station de base de la cellule d'où provient l'appel).

Méthode DCA Centralisée

Un canal d'un ensemble partagé est affecté à un appel pour une utilisation temporaire par le contrôleur centralisé. Différentes stratégies de sélection du canal éligible sont proposées. Parmi celles-ci, nous pouvons citer :

- FA (*First Available*) : Le canal sélectionné est le premier cherché répondant au problème de distance de réutilisation et d'interférences. Par rapport à FCA, FA supporte 20% de trafic supplémentaires pour des charges faibles à modérées.
- LODA (*Locally Optimized Dynamic Assignment*) : La fonction de coût est basée sur la future probabilité de blocage dans le voisinage de la cellule dans laquelle l'appel sera initié.

- CROS (*Channel Reuse Optimisation Scheme*). L'objectif est de maximiser l'utilisation des canaux du système, le canal élu sera celui dont la distance avec le co-canal est la plus courte.
- *Selection with Maximum Usage on the Reuse Ring* : Le canal sélectionné est celui qui est le plus utilisé parmi les cellules co-canaux.

Il existe également des méthodes centralisées avec ré-arrangement de canaux. Ces méthodes sont alors utilisées pour améliorer les performances de DCA. Le principe est de changer les appels de canal dès que possible vers un autre canal disponible dans le but de conserver une distance de réutilisation minimale entre les cellules utilisant le même canal. Bien que ces stratégies soient efficaces car l'allocation des canaux est proche de l'optimal, il faut bien voir que le contrôleur central doit être capable d'absorber l'ensemble des demandes bien que les algorithmes se montrent parfois complexes et long en calcul. Pour résoudre ce goulot d'étranglement, une solution consiste à distribuer la gestion des canaux à chaque station de base.

Méthodes DCA Distribuées

Le principe de ces méthodes est de laisser le choix du canal à la station de base grâce à la connaissance de canaux utilisés dans les cellules voisines ou grâce aux mesures de puissance des signaux reçus.

- Dans le premier cas, la station de base alloue un canal et l'information est alors diffusée aux autres stations. Ce système fournit de très bons résultats, toutefois les mises à jour d'informations deviennent importantes à effectuer sous de très fortes charges. Citons par exemple l'algorithme *Local Packing Dynamic* - DCA où l'affectation d'un canal se fait suivant la matrice d'occupation des canaux et toujours suivant le niveau d'interférences des co-canaux ;
- Une autre solution attrayante consiste à faire effectuer par la station de base des mesures de puissance de signaux. La station utilise alors cette information locale pour allouer un canal sans avoir besoin de communiquer avec d'autres stations de base du réseau. On peut qualifier ce système d'autogéré. Le temps de réponse est alors très court et l'optimal du nombre de canaux alloué est presque atteint, mais quelques effets indésirables peuvent survenir comme des interruptions ou encore une instabilité du système. *Sequential Channel Search* est la stratégie la plus simple : chaque station de base et chaque mobile recherche séquentiellement et dans le même cadre le premier canal disponible avec un seuil C/I acceptable.

6.3.3 Hybrid channel allocation (HCA)

Les méthodes hybrides visent à tirer des bonnes propriétés des schémas énoncés précédemment, le principe de base sera alors de partager les canaux en deux groupes. D'une côté, un ensemble fixe de canaux alloués suivant FCA lors de la planification des fréquences dans le réseau ; et de l'autre un pool de fréquences partagé par tous les utilisateurs permettant d'accroître la flexibilité. Dans tous les cas, il s'agit d'utiliser en priorité les canaux propres aux cellules (FCA), avant de passer, dans le cas où tous les canaux sont occupés, aux ressources gérées par une stratégie DCA. Un paramètre essentiel de HCA est le ratio des canaux statistiques et dynamiques. En général, ce ratio

est fonction de la charge de trafic estimé lors de la planification cellulaire mais peut être amené à évoluer à cause d'une redistribution de la charge. Lors d'études de performances [KAT94], il semblerait que pour une charge inférieure à 15% il soit préférable d'avoir en majorité des canaux DCA. Entre 15% et 30%, un système équilibré, donne de bonnes performances alors qu'entre 30% et 40%, DCA doit être minoritaire mais doit rester présent pour réagir aux brusques variations du trafic. Au-delà de 40% de charge, les meilleurs résultats sont obtenus sans utiliser de gestion dynamique, donc uniquement des stratégies d'allocation de type FCA.

Comme nous l'avons vu, un des mécanismes les plus importants dans les réseaux mobiles, est la gestion du transfert inter-cellulaires., il apparaît que la procédure de handover échoue (*dropping handover*) lorsque le système n'arrive pas à trouver un canal disponible dans la cellule que le mobile va visiter. Afin de réduire cette probabilité de coupure d'appel en cours, quelques variantes des mécanismes d'allocation proposent de réserver une partie des canaux pour l'exécution de handovers uniquement [KAT94] : ce sont les stratégies *guard channels*.

6.4 Processus de la gestion de QoS de l'allocation des ressources de bout en bout

L'étude de l'allocation des ressources du point de vue dimensionnement et configuration, présentée dans les paragraphes précédents nous permet de dégager deux étapes dans le processus d'allocation. Il s'agit en première étape de réaliser la partie analyse, dimensionnement, planification et paramétrage qui donne lieu aux modèles de trafic et de mobilité. Celle ci est effectuée dans un environnement radio cellulaire opérationnel (cf. chapitre 7). Dans une deuxième étape, il s'agit de concevoir un modèle de gestion de la qualité de service de bout en bout qui se base à la fois sur un module de contrôle des paramètres estimés pour les services multimédia envisagés et sur un module qui se renseigne sur les ressources disponibles dans le réseau (cf. chapitres 8, 9 et 10). En effet, ce modèle de gestion de la qualité de service d'allocation des ressources permet, en ligne, de gérer la qualité de service dynamiquement selon tout d'abord un profil d'usage bien déterminé et ensuite suivant la disponibilité du réseau du point de vue ressources et capacité, et hors ligne, de constituer une base d'information pour re-paramétriser et re-dimensionner les modèles de trafic et de mobilité. La figure 6.1 résume le processus que nous venons de présenter.

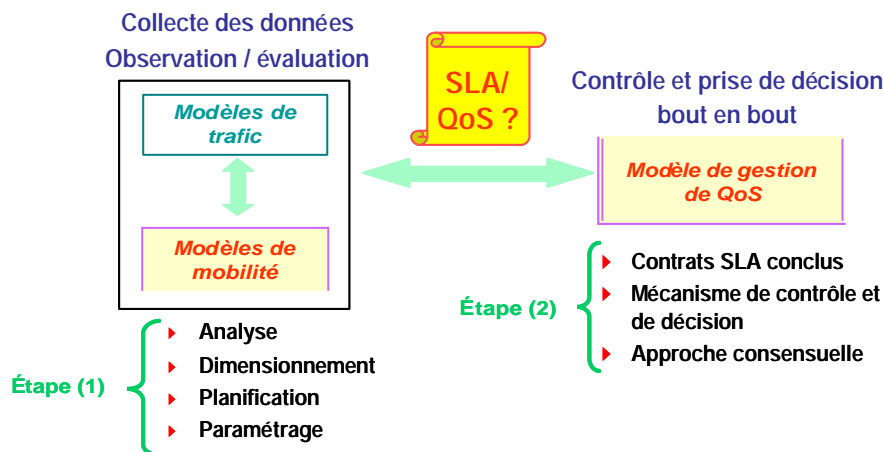


Figure 6.1 Processus d'allocation des ressources de bout en bout

Nous appliquons ce processus pour l'allocation des ressources dans les réseaux radio mobiles multimédia de bout en bout. Par conséquent, les approches exposées respectivement dans les chapitres 7, 8, 9 et 10 tiennent compte des spécificités de ce processus d'allocation.

6.5 Travaux sur l'allocation des ressources et la gestion de la qualité de service

Dans ce paragraphe nous allons présenter les principaux projets et travaux concernant la gestion de la qualité en fonction des ressources réseau disponibles.

6.5.1 Approches et travaux divers

Cependant, la réservation des ressources nécessite un protocole de pré-réservation, un mécanisme de contrôle d'admission et un mécanisme général de contrôle de l'application de la politique. Toutes ces fonctions sont complexes et nécessitent des modifications importantes au niveau des infrastructures réseaux actuelles. Leur impact est de rendre le réseau peu performant et moins facile à étendre et à densifier [AHM03].

Toutefois, d'autres approches innovantes existent et permettent d'adapter l'application aux conditions du réseau. Nous présentons, dans ce qui suit, quelques approches de réservation et d'adaptation qui tiennent compte des conditions fluctuantes du réseau. [CHAD98] présente une architecture basée sur la notion de qualité de service orientée applications. Un courtier permet aux applications et aux fournisseurs de services de réserver les ressources réseaux nécessaires. Ces courtiers peuvent être hiérarchiques, ce qui veut dire qu'un courtier peut appeler les services d'un autre courtier. Ce schéma de réservation est statique et l'état calculé du réseau par le courtier ne correspond pas exactement à son état réel. Nous voyons apparaître à ce niveau le problème de la validité de l'information sur laquelle se basent les décisions. Un concept clé de leur proposition est la boucle de rétroaction (*Feedback*) qui contrôle l'ajustement de l'application aux conditions du réseau. Selon cette proposition, l'application doit mettre en œuvre un moniteur qui effectue des mesures régulières sur la qualité de service reçue et un réacteur qui réagit en réponse à cette qualité de service reçue.

De manière similaire dans [CHA00], l'application peut s'adapter aux conditions grâce à un système d'adaptation intégré dans son environnement d'exécution. Le système contrôle continuellement les ressources disponibles et le comportement de l'application en terme de garantie de qualité de service. Il ajuste cette qualité de service automatiquement d'une manière précise à un instant précis. Plusieurs profils d'exécution sont disponibles. L'adaptation dans ce système consiste à choisir un profil et à l'appliquer lorsque les conditions d'exécution sont réunies.

Dans [AHM03], l'adaptation est laissée à la charge du routeur de bordure. Ce routeur permet d'améliorer ou de dégrader la qualité de service des flux multimédia en fonction de l'état du réseau. Un des problèmes de cette approche est sans doute le problème de la mise en échelle.

6.5.2 Projets et travaux divers

Tjelta [TJEL01] présente des directives et des recommandations basées sur la technologie des agents. Ces directives et recommandations ont pour but une gestion optimale des ressources et de la qualité de service au niveau des réseaux. Le modèle de gestion des ressources et de la qualité de service est illustré par la figure 6.2.

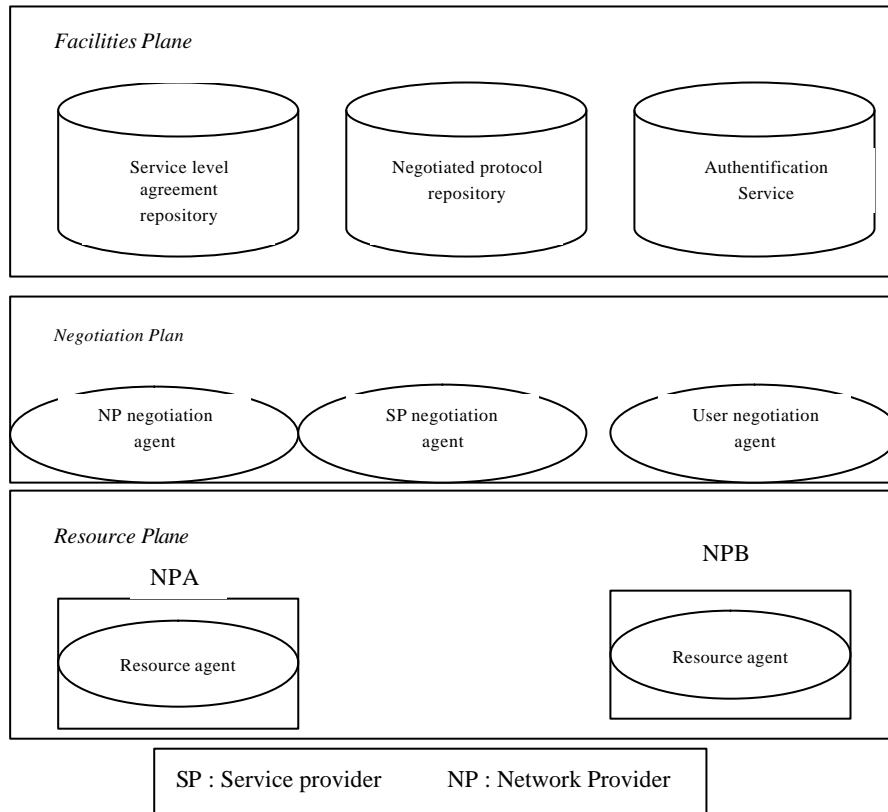


Figure 6.2 Modèle Tjelta

Ce modèle est basé sur trois couches. La couche *facilities plane* implémente les répertoires des niveaux de contrat de services, des protocoles de négociation et des algorithmes d'authentification des services. Cette couche incorpore aussi les entités nécessaires pour l'interaction avec la couche *negociation plane*. Cette dernière gère les négociations avec les utilisateurs, les fournisseurs de services et l'opérateur du réseau. Au niveau de la couche *resource plane*, les NP (A et B) gèrent les ressources disponibles sur les supports fixes et sur les interfaces radio (gestion de cellules, gestion de handover, etc.). Malgré que ce projet nous semble intéressant, l'utilisation de la technologie des agents a tendance à limiter les performances du modèle de gestion de ressources et de qualité de service proposée.

Le projet Eurescom P1008 [P1008] intitulé «*Inter-operator interfaces for ensuring end to end QoS*» a pour but de résoudre les problèmes de la gestion de la qualité de service pour des services IP. La résolution de ces problèmes se fait par l'utilisation de mécanisme de négociation entre les différentes parties du système. La négociation se fait entre des interfaces utilisées pour la gestion des composantes de la qualité de

service et l'échange de flux de données. Des procédures de mesures sont développées pour déposer d'une connaissance temps réel de l'état de services et du réseau.

Le projet Eurescom QUASIMODO [P906][QUAS] étudie la conception et l'implémentation d'une méthodologie globale avec des outils pour le contrôle de la qualité de service sur différents types de réseaux IP. Il se fixe parmi ses objectifs de bâtir une corrélation entre les paramètres de qualité de service au niveau de l'utilisateur (qualité de service perçue), les paramètres de l'application (service) et les paramètres de performance du réseau (NP).

Le but du projet ARROWS (*Advanced Radio Resource Managment for Wireless Services*) est de fournir une méthodologie avancée d'une part pour l'allocation et la gestion des ressources radio (RRM), et d'autre part la gestion de la qualité de service. Cette méthodologie mathématique est appliquée aux modes UTRA-TDD et UTRA-FDD en UMTS, afin de pouvoir supporter des services qui intègrent voix, images et données.

Jong [KUNG00] rappelle les différentes considérations sur lesquels doit se baser un modèle de gestion de qualité de service dans un réseau 3G. Ces considérations sont :

- Localisation de l'utilisateur ;
- Connexion toujours existante ;
- Contrôle de flux de données ;
- Gestion de la bande passante ;
- Contrôle du taux d'erreur ;
- Gestion efficace de la mobilité des utilisateurs.

Manvi [MANVI00] présente un modèle de gestion de la qualité de service. Ce modèle est basé sur les agents mobiles pour satisfaire les cinq principes fonctionnels d'une architecture de qualité de service. Ces principes sont : l'intégration, la séparation, la transparence, la gestion asynchrone des ressources et la performance. Le but principal de ce modèle est la gestion de la largeur de bande. La simulation de ce modèle a été réalisée sous un réseau de machines sous Linux. L'interaction entre les différents agents et la communication des flux entre les différents nœuds du réseau sont modélisés par un modèle markovien.

6.5.3 Discussion

Les travaux et projets cités ci-dessus, concernant la gestion de la qualité de service en fonction des ressources disponibles sont souvent limités à la gestion d'un aspect particulier du problème de l'allocation des ressources des services multimédia dans le cadre de l'UMTS (conception et mis en œuvre, pré-réservation par rapport aux ressources du réseau, etc.).

Les travaux et projet de gestion de qualité de services traitent des types particuliers de gestion. La gestion dynamique des services temps réel de bout en bout, en fonction des capacités des réseaux reste un problème d'actualité. En effet, à notre connaissance, aucune proposition dans ce sens n'est actuellement effectuée. Dans nos travaux nous

proposons un modèle afin de résoudre cette problématique qui nous semble intéressante aussi bien en terme scientifique que d'usage avec l'arrivée du multimédia.

La gestion globale de la qualité de service d'allocation des ressources dans un réseau UMTS (Release 1999 du 3GPP) ou UMTS-NGN (Release 5 du 3GPP) n'est pas non plus traitée dans les travaux et projets dans l'état de l'art. Nous apportons à travers nos travaux une solution globale à ce problème. Nous proposons au sein de nos travaux un modèle basé sur des configurations d'études réelles. Ce modèle assure la gestion de la qualité de service par rapport aux ressources du réseaux qui participent à l'acheminement d'une application donnée.

Dans notre modèle, l'allocation de ressources est laissée à la charge des gestionnaires de ressources de chaque domaine qui participe à l'application (cf. chapitre 10, figure 6.3). La configuration de ces éléments et leur pré-réservation reflète les niveaux SLA entre les fournisseurs de service et l'application utilisateur de service. Cette configuration de la pré-réservation est un outil d'administration capable de fournir la configuration adéquate aux routeurs en temps réel. La pré-réservation de ressources dynamique des flux multimédia consiste à trouver quand nécessaire une route avec un niveau SLA acceptable (objet du chapitres 9 et 10).

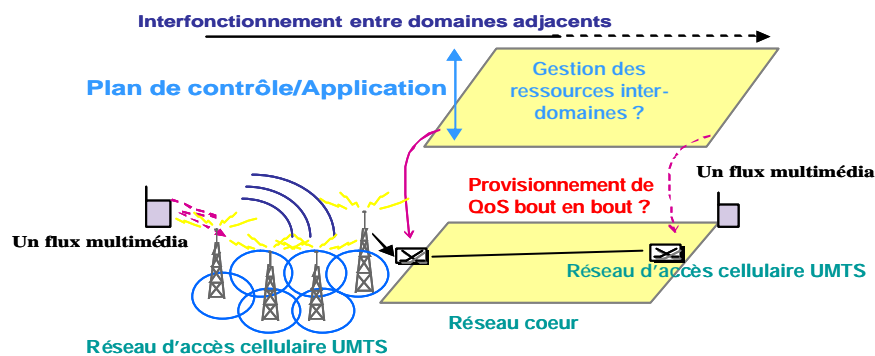


Figure 6.3 Plans de contrôle/application

6.5.4 Contrôle d'admission d'appel

Au-delà de l'allocation des ressources en local c'est à dire au niveau de réseau radio cellulaire, le mécanisme de contrôle d'admission d'appel entre domaines gère au niveau de chacun d'eux le partage de la bande passante pour garantir un des niveaux de qualité de service acceptable. Ce mécanisme se décompose en deux étapes :

- La première étape, au niveau du réseau radio, permet de fixer les paramètres statiques qui sont les seuils d'acceptation d'un appel [BEG03][CHE02]. Si l'allocation de ressources est acceptée au niveau du réseau radio de l'appelant, alors le contrôle d'admission sera lancé entre domaines afin de déterminer si la qualité de service requise peut être assurée de bout en bout.
- La seconde étape, entre les réseaux mobiles, les réseaux cœurs des opérateurs et les réseaux des fournisseurs de services, permet d'établir un consensus entre les différents processus de décision des domaines participants à l'application (radio,

cœur, fournisseurs de services). Ce consensus concerne la décision de réservation des ressources selon un des niveaux SLA possibles.

Pour gérer le problème d'allocation des ressources de bout en bout durant une requête, le réseau est divisé en zones appelées «zones assurées». Dans chaque zone, il existe un agent courtier des ressources RM [ELL02][BAK02] ayant les fonctions suivantes (voir figure 6.4) :

- Implantation des conditions d'accès aux ressources réseau (données seuils récupérées des différentes zones et enregistrées dans les SLA des routeurs de bord).
- Information des routeurs de périphérie de la présence de flux de trafic et de leurs caractéristiques.
- Interaction avec les agents courtiers des ressources d'autres domaines.

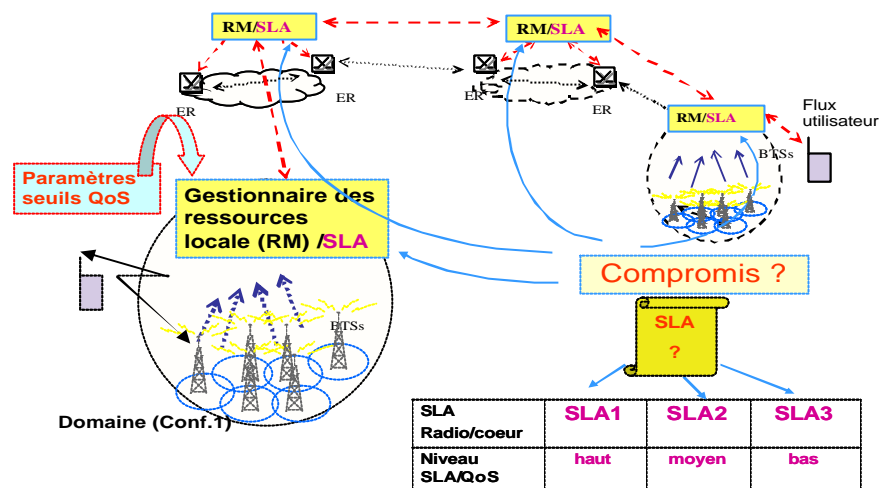


Figure 6.4 Réseaux radio mobile/DiffServ

6.6 Conclusion

Au cours de ce chapitre, nous avons défini le dimensionnement des ressources du système du point de vue capacité réseau, mobilité des utilisateurs, contrôle d'admission, etc. Ensuite, nous avons présenté les différentes méthodes d'allocation des ressources.

Les allocations que nous considérons dans nos travaux (voir chapitres 8,9 et 10) sont basées sur :

- La gestion de la mobilité des utilisateurs ;
- Les paramètres de la qualité de service ;
- Le contrat de SLA ;
- Le contrôle d'admission d'appel alimenté par un protocole de consensus ;
- La négociation des ressources ;
- Les préférences de l'utilisateur.

Les propriétés de l'allocation des ressources, sont présentées au niveau du paragraphe 6.4. Ces propriétés sont :

- La demande ;
- Contrôle et prise de décision ;
- Mise en application.

La dernière partie de ce chapitre a été consacrée à la présentation de l'état de l'art des travaux de la gestion de la qualité de service d'allocation des ressources. A partir de cet état de l'art, nous avons positionné nos travaux et notre contribution à travers cette thèse.

Références Bibliographiques

- [AHM03]Ahmed, T and al., "*QoS dynamic management of video flow in IP DiffServ architector,*" Proc. GRES'03, 2003, pp. 53-65.
- [BHA01]Bhatnagar, S. and al., "*Providing quality of service guarantees using only edge routers,*" Global Telecommunications Conference, 2001. Globecom'01. IEEE, Vol.4, 2001, pp. 2542-2548.
- [BAT98]Bathelt, A., "*Grafical models for analysis of design mobile wireless networks,*" PhD thesis ENST, 14 April 1998.
- [BEG03]Ben Rejeb, S., Choukair, Z. and Tabbane, S., "*Resource allocation and QoS management for cellular multimedia network,*" Proc. GRES'03, Brazil, February 2003, pp. 3-13.
- [BAK02]Bakiras, S. and Victor O. K. Li., "*Efficient resource management for end to end QoS guarantees in DiffServ network,*" Communications, 2002. ICC 2002 IEEE International Conference on, Vol.2, 2002, pp. 1220-1224.
- [CHAD98]Chahdra, P. and al., "*Network support for application oriented QoS,*" Proc. of Sixth International Workshop on Quality of Service (IWQoS'98), pp. 187-195.
- [CHA00]Chang, F. and al., "*Automatic configuration and run-time adaptation of distributed application,*" Ninth IEEE Intl. Symposium on Hight Performance Distributed Computing (HPDC), August 2000, pp. 11-20.
- [CHE02]Cheng, Y. and Zhuang, W., "*DiffServ ressources allocation for fast handoff in wireless mobile Internet,*" IEEE Communications Magazine May 2002, pp. 130-136.
- [CHE02]Cheng, Y. and Zhuang, W., "*DiffServ resources allocation for fast handoff in wireless mobile Internet,*" IEEE Communications Magazine May 2002, pp. 130-136.
- [DAH01]Dahmouni, H., "*Conception and planification of mobile networks: video-telephony application on UMTS/IP,*" work elaborated for Philips Consumer communications society - October 2001.
- [ELL02]Ellen W. Zegura and al., "*QoS routing for anycast communications: motivation and an architecture for DiffServ networks,*" IEEE Communication Magazine, June 2002.
- [FOS93]Foschini, G.J. and al., "*Channel cost of mobility,*" IEEE Transactions on vehicular technology, Vol. 42, No.4, November 1993.
- [LAG98]Lagrange, X., "*Performance analysis of hierarchical cellular network,*" PhD thesis, ENST 98 E 009.
- [MANV00]S.S.Manvi, "*QoS management by mobile agents in multimedia communication,*" IEEE 2000.
- [MON92]Montenegro, G. and al., "*Time dependent analysis of mobile communication trafic in a ring shaped service area with non uniform vehicule distribution,*" IEEE Transaction on vehicular technology, Vol.41, No.3, August 1992.

[NAN93]Nanda, S., "*Teletraffic models for urbain and suurbain microcells : cell sizes ans handoff rates,*" IEEE Transactions on vehicular technology, Vol.42, No.4, Nov 93.

[P906]<http://www.euroscm.de/public/projects/P900-series/p1008/default.asp>

[P1008]<http://www.euroscm.de/public/projects/P1000-series/p1008/default.asp>

[QUAS]QUASIMODO; <http://www.eurescom.de/public/projects/P900-series/906/P906.htm>

[STE95]Steel, R. and al., "*Teletraffic performance of GSM 900/DCS 1800 in street microcells,*" IEEE Communications magazine, March 1995.

[SES92]Sescar, I. and al., "*Rate location area updates in cellular environement,*" 42 th IEEE Vehicular technology conference 1992.

[TJEL01]Terje Tjelta, "*Future broadband radio acsess systems for integrated services with flexibles resources management,*" IEEE Communications Magazine - August 2001.

[TAB97]Tabbane, S., "*Location management methods for third-generation mobile systems,*" IEEE Communication , Magazine, August 1997.

[THO98]Thomas, R. and al., "*Influence of the moving of the mobile stations on the performance of a radio mobile cellular network,*" Proceeding of the Nordic seminar on digital land radio mobile communications, 1998.

[KAT94]Katzela, I. and Naghsineh, M., "*Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey,*" Proceeding of the IEEE, 1994, pp. 1430-1430.

[KHA02]Khalil, I. and Braun, T., "*Edge provisioning and fairness in VPN-Diffserv networks,*" the 9th International Conference on computer communication and network (ICCCN 2000), 2002, Vol.2, pp. 1220-1224.

[KUNG00] Shaw-Kung Jong, "*QoS considerations on the third generation (3G) wireless systems,*" Research challenges, 2000. Proceedings.Academia/Industry working conference on, 2000.

Partie 3 : Allocation des ressources multimédia

Chapitre 7

**Modèle de mobilité dans un réseau radio
mobile multimédia****7.1 Introduction**

Pour un service radio et plus particulièrement pour les services multimédia, la disponibilité de ressources fluctue souvent, en fonction de la charge, de la mobilité des usagers et de la variation de la qualité de canaux. Cette fluctuation est un problème important pour l'opérateur de réseau radio auquel nous nous intéressons au sein de nos travaux pour adapter l'utilisation des ressources en considérant essentiellement la mobilité des usagers. L'allocation de la bande passante pour les différents types de trafic se fait uniquement lors de l'occurrence dans la cellule d'un appel nouveau, un appel à effectuer ou un appel en handover.

La mobilité des utilisateurs a une influence importante sur la gestion de la qualité de service de l'allocation des ressources dans un réseau radio mobile multimédia. A chaque changement de cellule d'un utilisateur en communication, le système effectue un handover, afin de transférer la communication de la cellule courante de service vers la cellule destination. Un tel handover intercellulaire modifie la répartition du trafic dans le système et génère une charge de signalisation supplémentaire. On doit par conséquent considérer la mobilité des utilisateurs dans le dimensionnement du réseau et dans toute analyse de performance par la suite.

Pour bien dimensionner le système et analyser ses performances, la mobilité des utilisateurs doit être considérée dans les modèles de trafic choisis (cf. chapitre 3). Notons que ces modèles ne distinguent pas les différents mobiles ni leurs mouvements,

mais modélisent la mobilité macroscopique dans le système. Les modèles de trafic retenus (cf. chapitre 3) dans le cadre de nos travaux se basent sur les taux de transitions d'utilisateurs (les appels entrants/sortants) entre cellules comme paramètres de mobilité.

Dans le cadre de ce travail, nous proposons un modèle de mobilité pour déterminer les valeurs des taux de transition à partir des données réelles récoltées par des mesures effectuées sur le système à partir des compteurs OMC et des informations récupérées sur le réseau à partir des cartes géographiques telles que les longueurs des rues, les positions des BTS, les distances entre les BTS, les rayons des cellules, etc. Les compteurs OMC s'incrémentent à chaque fois qu'une action se déclenche dans le réseau. Dans notre travail nous nous intéressons aux compteurs liés au nombre de handovers échoués, de nombre de coupures de communication sur handover, de nombre de tentatives de handovers inter et intra-cellulaires, etc.

Ce modèle de mobilité nous permet par la suite de déterminer au niveau d'une cellule et également au niveau du système entier les paramètres seuils, au delà desquels un appel peut être refusé. Il permet ainsi de connaître les limites de réservation de ressources : le taux de handovers, la probabilité de rejet d'appels, le trafic offert, le grade de service, etc.

Dans ce chapitre, nous nous intéressons à développer un modèle de mobilité dans quatre modèles de trafic retenus tels que voix, vidéo-téléphonie, ftp et www (cf. chapitre 3). Ce modèle de mobilité utilise le graphe de rues pour déterminer la capacité du système et évaluer les paramètres de performance réseau. Ensuite nous présentons l'environnement considéré pour calibrer et appliquer ce modèle. Nous terminons ce chapitre par présenter l'organigramme de simulation ainsi que les paramètres de performances réseau conclus.

7.2 Développement de modèle de mobilité

Plusieurs modèles de mobilité ont été présenté dans la littérature (cf. chapitre 6). Nous proposons dans la suite une version simplifiée du modèle de mobilité dans un environnement multimédia, se basant sur les flux d'utilisateurs dans le réseau routier d'une configuration réelle donnée. Ce modèle se limite à étudier de manière macroscopique les dynamiques globales dans le système [BEM03].

Nous considérons un système en équilibre et une densité constante d'utilisateurs. Les mouvements des mobiles sont indépendants.

7.2.1 Les modèles graphiques

La théorie de graphes a trouvé des applications dans les domaines des réseaux radio mobiles. Trois aspects de la planification radio mobile sont traités à l'aide de modèles graphiques à savoir :

- Le graphe biparti : modélise les relations de couverture et d'interférences dans le système ;
- Le graphe d'interdiction : modélise les contraintes de réutilisation de fréquence pour une planification de fréquence ;

- Le graphe de rues : permet de modéliser la mobilité des utilisateurs dans les systèmes radio mobiles multimédia ainsi que le réseau de transport dans un réseau cœur DiffServ de 3^{ème} génération.

On étudie dans le cadre de cette thèse une version modifiée du graphe de rues. Cette version utilisée pour développer dans un premier temps un modèle de mobilité qui permet de prendre en compte des trafics multimédia (voir paragraphe 7.2.2) et dans un deuxième temps pour router un flux multimédia selon les niveaux SLA à travers plusieurs domaines adjacents dans un réseau cœur DiffServ (voir chapitre 10).

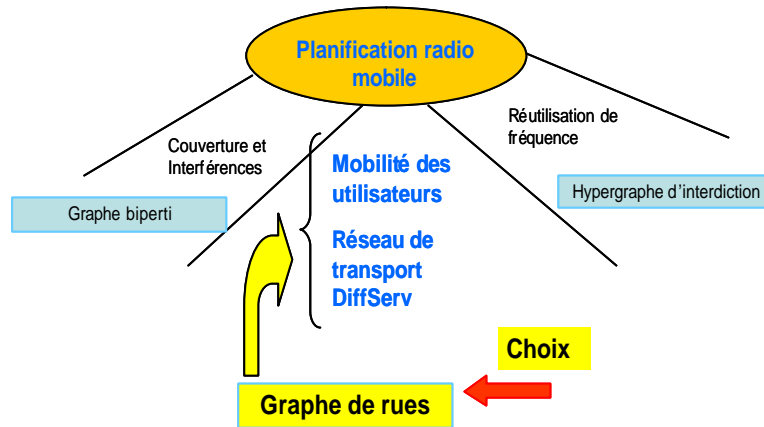


Figure 7.1 Les modèles graphiques pour la planification radio mobile

7.2.2 La définition du graphe de rues

Nous partons du principe que les utilisateurs dans une zone urbaine se déplacent surtout le long des rues [BAT98]. Ainsi, le graphe de rues $RG(X', R)$ permet la modélisation de la mobilité des utilisateurs à partir des flux d'utilisateurs dans celles-ci. Il représente ainsi le réseau routier dans la zone de service du système.

Dans ce graphe, une rue est représentée par une arête $r_{ik} = (x_i, x_k)$ qui lie les deux extrémités de la rue. Une rue à double sens est représentée par deux arêtes $r_{ik} = (x_i, x_k)$ et $r_{ki} = (x_k, x_i)$. On attribue à chaque arête r_{ik} un poids f_{ik} qui représente le flux d'utilisateurs dans la rue (voir figure 7.2). Le graphe des rues se compose donc d'un ensemble de nœuds $X' = \{x_i\}$ aux intersections des rues qui sont reliées par des arêtes directives de l'ensemble $R = \{r_{ik} = (x_i, x_k)\}$.

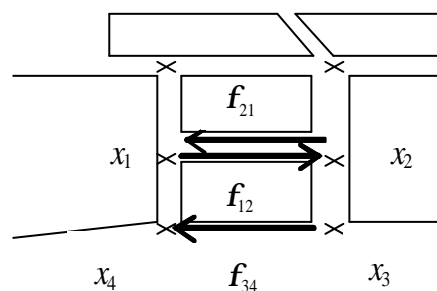


Figure 7.2 Graphe de rue dans une zone urbaine

7.2.3 Répartition des flux d'utilisateurs aux cellules

Les flux d'utilisateurs dans les rues permettent de déterminer les flux d'utilisateurs entre les cellules du réseau. Si une rue est couverte par plusieurs cellules, on répartit le flux d'utilisateurs dans la rue aux cellules couvrantes.

Les rues sont affectées aux différentes zones de couverture de BTSs. Une rue qui se trouve dans une zone de recouvrement est affectée à chacune des cellules couvrantes. On suppose qu'un utilisateur dans une telle rue est connecté avec une certaine probabilité à une des cellules couvrantes.

Le flux total \mathbf{f}_{tk} dans une rue $r_k=(x_i, x_k)$ (où x_i et x_k sont couverts par un nombre M_{tk} de cellules) est divisé dans des sous-flux $\mathbf{f}_{tk}^{(j)}$, où $\mathbf{f}_{tk}^{(j)}$ désigne les sous-flux attribués à la cellule j . Cette répartition peut s'exprimer par :

$$\mathbf{f}_{tk}^{(j)} = \mathbf{f}_{tk} \cdot \mathbf{p}_{tk}^{(j)} \text{ avec } \sum_j \mathbf{p}_{tk}^{(j)} = 1 \quad (7.1)$$

La répartition des flux est illustrée dans la figure 7.3 pour le cas d'une rue qui est couverte par deux cellules. Le flux \mathbf{f}_{12} attribué à (x_1, x_2) est réparti par exemple aux cellules x_n et x_p couvrantes : $\mathbf{f}_{12} = \mathbf{f}_{12}^{(n)} + \mathbf{f}_{12}^{(p)}$.

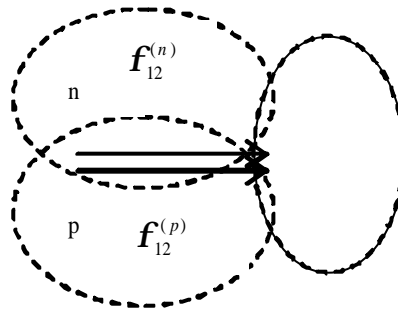


Figure 7.3 Répartition des flux d'utilisateurs sur les différentes cellules couvrantes

Dans la suite, on considère le cas : $\mathbf{p}_{tk}^{(j)} = \frac{1}{M_{tk}}$ et la répartition de flux d'utilisateurs est comme suit :

$$\mathbf{f}_{tk}^{(j)} = \frac{\mathbf{f}_{tk}}{M_{tk}} \quad (7.2)$$

\mathbf{f}_{tk} : la somme de flux des usagers dans les rues r_{tk} , est répartie entre les différentes cellules qui recouvrent ces rues sous forme des sous-flux $\mathbf{f}_{tk}^{(j)}$, $\mathbf{f}_{tk}^{(n)}$, etc.

$\mathbf{f}_{tk}^{(j)}$: les sous-flux des usagers dans les rues r_{tk} qui sont couvertes par la cellule j . Cette répartition peut s'exprimer par :

$$\mathbf{f}_{tk}^{(j)} = \frac{\mathbf{f}_{tk}}{M_{tk}} \quad (7.3)$$

Et le flux moyen des utilisateurs dans les rues r_{tk} est donné par :

$$\mathbf{f}_{tk} = \frac{n_{tk} v_{tk}}{l_{tk}} \quad (7.4)$$

où :

M_{tk} : le nombre de cellules qui couvrent les rues r_{tk} .

n_{tk} : le nombre d'utilisateurs dans les rues r_{tk} .

l_{tk} : les longueurs des rues r_{tk} .

v_{tk} : les vitesses des utilisateurs dans les rues r_{tk} .

Exemple 1 : soit une rue r_{12} couverte par deux cellules m et p (voir figure 7.4) :

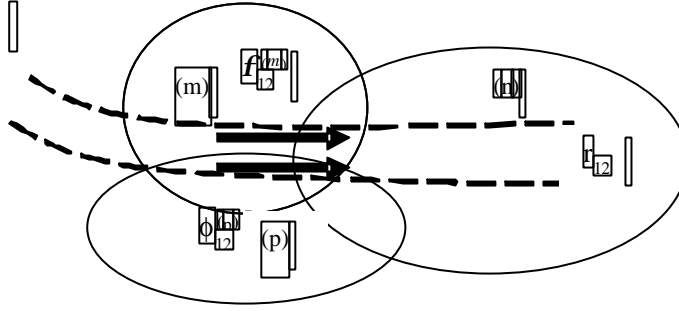


Figure 7.4 Répartition des flux d'utilisateurs entre cellules dans la rue r_{12}

Le flux total dans cette rue r_{12} est donné par :

$$f_{12} = f_{12}^{(m)} + f_{12}^{(p)}$$

avec $n=1, n'=1$

Si on généralise l'équation précédente pour les rues r_{tk} (cf. figure 7.3) on obtient l'équation suivante :

$$f_{tk} = f_{tk}^{(m)} + f_{tk}^{(n)} + f_{tk}^{(p)} + \dots$$

7.2.4 Flux de transition entre cellules

Dans la suite, on détermine les flux d'utilisateurs entre les cellules du réseau.

Les flux d'utilisateurs f_{tk} dans les rues du réseau sont répartis sur les différentes cellules du réseau. Les valeurs $f_{tk}^{(j)}$ introduites auparavant tiennent compte de cette répartition.

Afin de déterminer les flux entre cellules, on doit d'abord définir les emplacements où les utilisateurs qui se dirigent d'une cellule x_j vers une cellule x_n effectuent leur handover. On suppose dans la suite que le mobile effectue ce handover le plus tard possible, avant de quitter définitivement la zone de couverture de la cellule x_j . Par conséquent, le flux d'utilisateurs Θ_{jn} de la cellule x_j vers la cellule x_n est donné par la somme des flux $f_{tk}^{(j)}$ déterminés selon l'équation (7.2) dans les rues qui traversent la limite de la zone de couverture de x_j vers x_n :

$$\Theta_{jn} = \sum_{i \in j, k \in n, k \neq j} f_{ik}^{(j)} \quad (7.5)$$

Cette somme contient l'ensemble des rues qui lient un nœud x_i couvert par la cellule x_j à un nœud x_k qui est couvert exclusivement par la cellule x_n mais n'est couvert par la cellule x_j .

Exemple 2 : les rues r_{tk} sont couvertes par la cellule m . Le flux de transition de la cellule m vers la cellule n est donné par la figure 7.5 suivante :

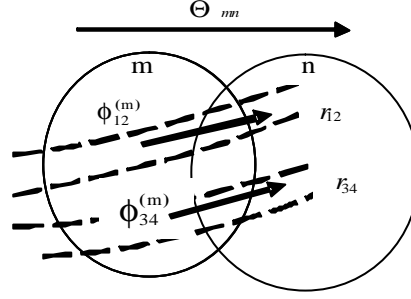


Figure 7.5 Le flux de transition de la cellule m vers la cellule n

Le flux Θ_{mn} est représenté par :

$$\Theta_{mn} = \sum_{tk} f_{tk}^m = f_{12}^{(m)} + f_{34}^{(m)} + \dots$$

Par le biais des valeurs Θ_{mn} , on connaît le nombre d'utilisateurs qui changent de cellule par unité de temps. On peut alors déterminer les taux de transitions d'utilisateurs entre cellules α_{mn} .

7.2.5 Taux de transition entre cellules

Avec les flux d'utilisateurs Θ_{jn} entre cellules, les valeurs des taux de transition d'utilisateurs entre cellules α_{jn} peuvent être déterminés pour des systèmes micro-cellulaires en tenant compte des différents axes de mobilité des utilisateurs.

On déduit le taux de transitions d'utilisateurs d'une cellule x_j vers une cellule x_n du flux d'utilisateurs Θ_{jn} et du nombre N_j d'utilisateurs dans la cellule x_j selon :

$$\alpha_{jn} = \frac{\Theta_{jn}}{N_j} \quad (7.6)$$

Pour une densité $s_{ik}^{(j)}$ constante dans une rue $r_{ik} = (x_i, x_k)$ de longueur l_{ik} , le nombre d'utilisateurs $N_{ik}^{(j)}$ en liaison avec la cellule x_j est donné par $N_{ik}^{(j)} = s_{ik}^{(j)} l_{ik}$. La densité d'utilisateurs $s_{ik}^{(j)}$ est donnée par $s_{ik}^{(j)} = \frac{f_{ik}^{(j)}}{v_{ik}}$, où v_{ik} est la vitesse des utilisateurs dans la rue.

Le nombre d'utilisateurs N_j dans la cellule j est donné par la somme des utilisateurs qui se trouvent dans des rues couvertes par j et attachées à la cellule j :

$$N_{tk} = \sum_{r \in j} s_{rs}^{(j)} I_{rs} \quad (7.7)$$

Le taux de transitions des utilisateurs de x_j vers x_n , peut alors être calculé par :

$$a_{jn} = \frac{\Theta_{jn}}{N_{tk}} = \frac{\sum_{t \in j, k \in n, k \neq j} f_{tk}^{(j)}}{N_{tk}} \quad (7.8)$$

7.2.6 Taux de transition global de la cellule j vers les cellules voisines

Les paramètres a_j sont liés à a_{jn} par la relation :

$$a_j = \sum_n a_{jn} = \sum_n \frac{\sum_{tk} f_{tk}^{(j)}}{N_{tk}^{(j)}} \quad (7.9)$$

7.2.7 Taux de handovers intercellulaires

En tenant compte des taux de transition dans le système, il est possible de déterminer les taux de handover inter-cellulaires. Nous nous basons pour cela sur les lois de services qui correspondent à chaque type de trafic voix, vidéo-téléphonie, www, ftp, notées m_{trafic} , m_{trafic} tels que m_{voice} , $m_{vidéo-téléphonie}$, m_{www} , m_{ftp} .

En tenant compte de la mobilité des usagers, les taux de services sont donnés par les équations suivantes :

$$m_j = m_{trafic} + a_{jn} \quad (7.10)$$

7.2.8 Charge de trafic

Le trafic total offert r_j dans la cellule j est donné par l'équation suivante :

$$r_j = \frac{I_j + I_j^H}{m_{trafic} + a_j} \quad (7.11)$$

où $I_j^H = \sum_f I_{jf}^H$: taux de handovers arrivant dans la cellule j . Ils sont calculés depuis

l'ensemble des cellules adjacentes. Les taux d'arrivées de handovers dans une cellule j sont donnés par les équations d'équilibre de flux suivantes :

$$I_j^H = \sum_k a_{kj} [(1 - P_{bk}) \frac{I_k}{m_{trafic} + a_k} + (1 - P_{dk}) \frac{I_k^H}{m_{trafic} + a_k}] \quad (7.12)$$

où :

I_k : représente les taux d'appels nouveaux pour chaque type de trafic.

P_{bk} : représente la probabilité de blocage des appels nouveaux.

P_{dk} : représente la probabilité de handovers avec $P_{dk} \ll 1$.

Pour le système entier, on présente les valeurs a_{jn} dans la matrice des taux de transitions : $T = (a_{jn})$. Ces matrices sont les données d'entrée du modèle de trafic multimédia.

7.3 Modèle d'environnement considéré

Dans notre étude, une zone urbaine est définie (voir figures 7.6 et 7.7), considérée comme étant la zone de service et comportant des sites potentiels auxquels des poids sont affectés. Cette zone (cf. figure 2.11) est décrite par des paramètres avec lesquels le modèle de mobilité est défini. Ce modèle de mobilité se base principalement sur l'étude des flux d'utilisateurs dans les axes routiers qui nous permettent de déterminer en premier lieu le taux de transition/utilisateur entre les différentes cellules du système, ce qui donne par la suite la probabilité qu'un utilisateur dans une cellule transite vers une autre cellule, et en deuxième lieu les taux de handovers inter-cellulaires.

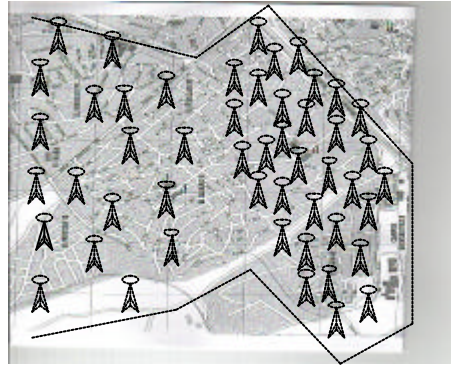


Figure 7.6 Configuration 1 (inspirée de la zone Hached Tunis, échelle : 1/12000, 50 cellules, 1BTS/cellule)

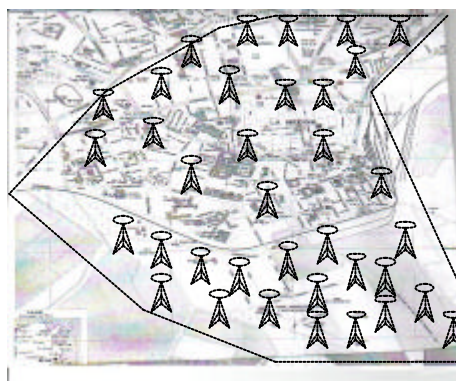


Figure 7.7 Configuration 2 (inspirée de la zone Sfax centre, échelle : 1/10000, 36 cellules, 1BTS/cellule)

Dans le cas du GSM/UMTS, l'opérateur dispose des mesures suivantes :

- Pour chaque cellule le nombre de handover à succès qui arrivent ou qui partent de chaque cellule adjacente ;

- Le nombre moyen des canaux dans chaque cellule ;
- La probabilité de blocage dans une cellule ;
- Le nombre de demandes d'appel multimédia dans le système ;
- La durée moyenne d'appel multimédia ;
- Le nombre d'appels multimédia permettant de déterminer le trafic offert et le trafic écoulé ;
- Le nombre d'utilisateurs dans le système qui peut être déduit du nombre d'appels, etc.

7.4 Calibration de modèle de mobilité

En nous basant sur les configurations 1 et 2 (cf. figures 7.6 et 7.7), sur le modèle de mobilité présenté ci-dessus et en nous appuyant sur une approche calibrée au sein de [T.TE02][BAT98][ETSI98][DAW99][BEG03][BEM03], le modèle de simulation à concevoir (cf. figure 7.10) est donc formé par un ensemble des paramètres réels résumés dans les tableaux 7.1, 7.2 et 7.3 :

Paramètres	Conf. 1	Conf. 2
Rayon de configuration	1.7 km	2 km
Nombre de cellules	50	36
Nombre de stations de base	50	36
Longueurs des rues	20 m à 1,7 km	20 m à 2.6 km

Tableau 7.1 Domaine de configuration (Cartes géographiques)

Paramètres	Conf. 1	Conf. 2
Période de mesure	12h à 13h	
Nombre des utilisateurs dans les rues	20 à 1500 utilisateurs	20 à 1700 utilisateurs
Charge de trafic voix	7.787 Erl	6.009 Erl
Capacité de cellule	28 canaux/cellule	45 canaux/cellule

Tableau 7.2 Mesures considérées (OMC)

Paramètres	Conf. 1	Conf. 2
Vitesse des mobiles dans les rues	3 à 60 km /h	
Probabilité de blocage d'appel	1%	

Tableau 7.3 Paramètres estimés (Bibliographie)

Dans le cadre de ce travail, les indicateurs clés à ajuster et à optimiser sont extraits à partir des deux configurations (cf. figures 7.6 et 7.7) placées l'une à Hached-Tunis, avec un rayon global 1.7 km et couverte approximativement par 50 cellules et l'autre au centre ville de Sfax (Sfax Port - Sfax Jedida-Medina) avec un rayon de 2 km, couverte

approximativement par 36 cellules. Dans chacune de ces deux configurations, on considère une BTS/ cellule, des grands axes routiers, des avenues, des petites rues, etc. Ainsi le trafic total considéré est énorme surtout dans les heures de pointe.

Hypothèses :

- La répartition régulière des cellules est choisie afin de dissocier les aspects de mobilité des aspects de répartition de cellules [BAT98] ;
- Le rayon d'une cellule : $R_{cell} = \frac{d_{BTS2-BTS1}}{2}$;
- Les utilisateurs mobiles sont servis par des stations de base (BTSs) ;
- Un site peut regrouper plusieurs BTSs ;
- Une rue peut être présentée par une arête r_{tk} ;
- La vitesse des utilisateurs mobiles dans les rues r_{tk} est v_{tk} ;
- La longueur des rues est présentée par l_{tk} ;
- Le nombre des utilisateurs dans les rues est n_{tk} ;
- La représentation de BTS sur la carte se fait en fonction de déploiement de données ;
- Une cellule contient c canaux voix, c' canaux vidéo, c'' canaux pour le web.

Les stations de base sont placées près des rues et couvrent leurs cellules associées. La mobilité des utilisateurs le long des rues est représentée par des flux.

Ainsi, le modèle de l'environnement est représenté comme suit (voir figure 7.8) :

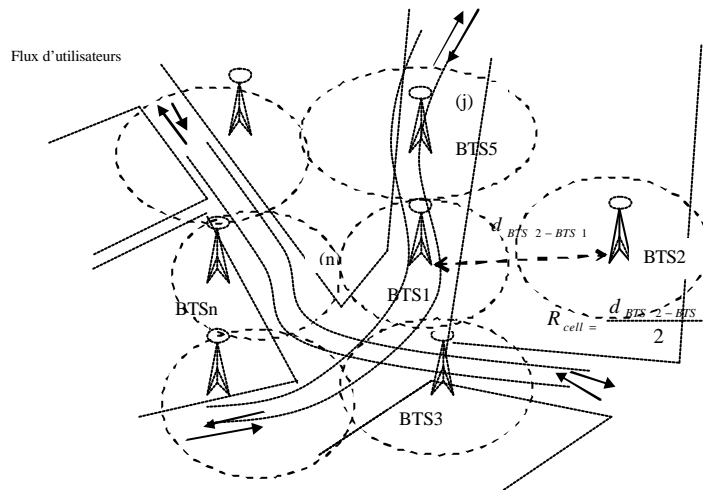


Figure 7.8 Modèle d'environnement considéré

Le modèle de mobilité considéré se base principalement sur l'étude des flux d'utilisateurs dans les axes routiers de la figure 7.9. Il est calibré par l'ensemble des paramètres réels et estimés résumés dans les tableaux 7.1, 7.2 et 7.3.

L'étude est effectuée pour l'heure chargée 12h et 13h.

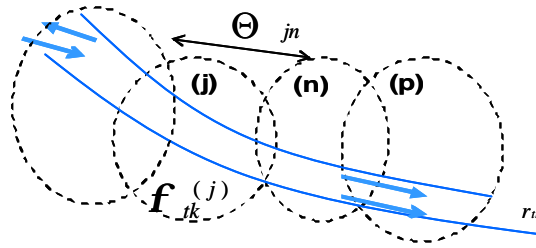


Figure 7.9 Modèle de mobilité adopté pour les axes routiers de l'environnement considéré

7.5 Expérimentations et interprétations des résultats

Dans les simulations réalisées (voir figure 7.10), les utilisateurs passent des appels voix, donnée, multimédia en fonction du modèles de trafic définis (cf. chapitre 3). Les utilisateurs se déplacent ainsi selon un modèle de mobilité calibré. Ils sont créés dans des sites composés de 3 à 6 stations de base. Les débits possibles varient de 8 Kbit/s à 2 Mbit/s.

L'objectif de la simulation est d'améliorer globalement la qualité du réseau et de s'assurer de l'efficacité de l'utilisation des ressources. La croissance du trafic dans le réseau nécessite en effet une continuité entre la phase de planification et le réseau opérationnel.

La première phase de la simulation pour le dimensionnement du réseau consiste à définir les indicateurs clés de performance. Ces derniers proviennent soit des mesures effectuées par le système, soit des mesures sur le terrain (cf. annexe E).

Il sera utile donc d'utiliser, pour les réseaux de troisième génération, les sites existants, installés par exemple pour le GSM, afin de partager les coûts des sites et des liens de transmission.

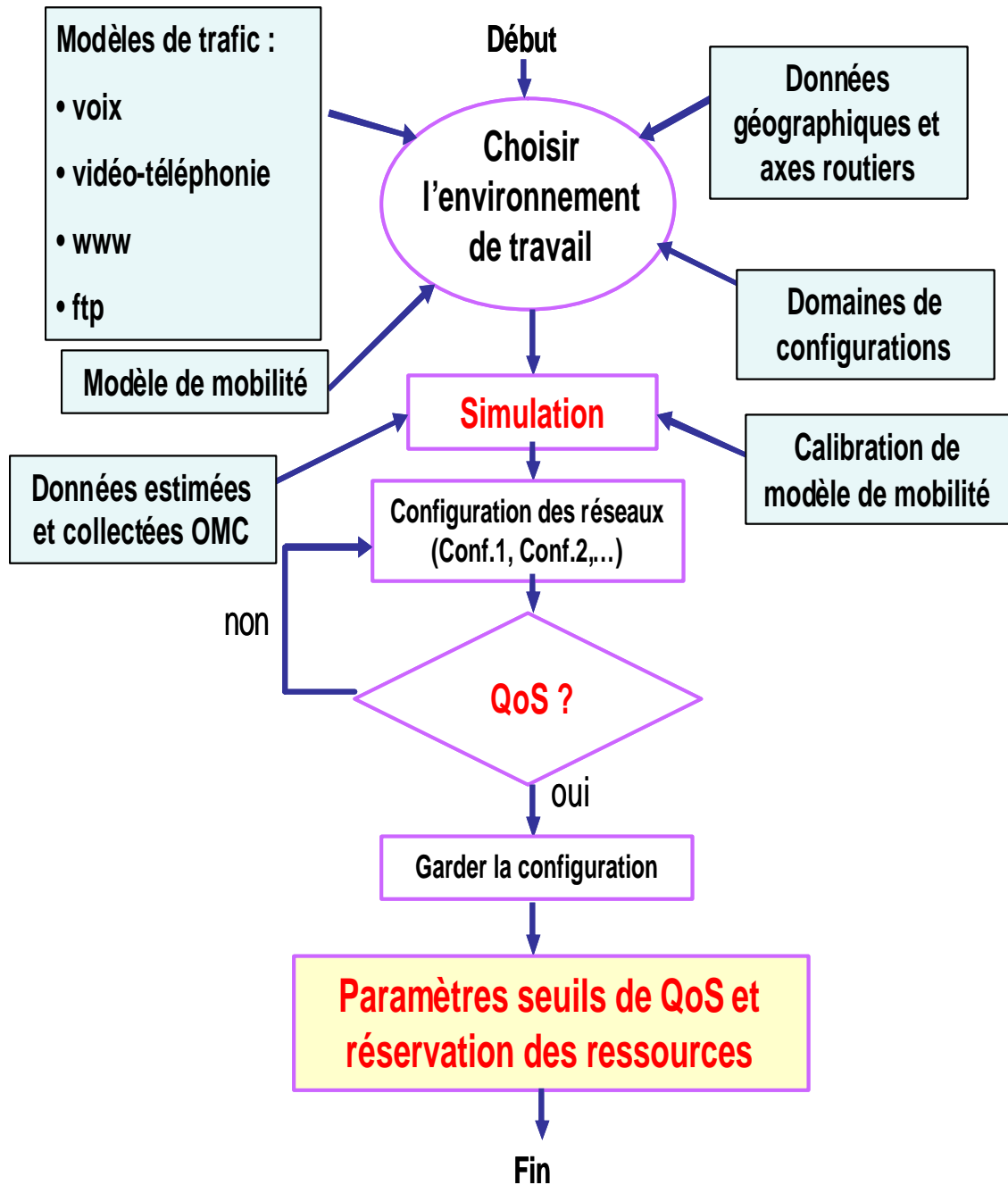


Figure 7.10 Organigramme de la procédure de planification et de l'évaluation des performances réseau lors de la phase de simulation

Selon les modèles de mobilité et de trafic, les flux d'utilisateurs dans les rues permettent la détermination des flux d'utilisateurs (appels entrants/sortants) entre les cellules du réseau (cf. équation (7.6)). Ces valeurs nous permettent de déduire les taux de handovers ainsi que la charge de trafic dans une cellule afin d'évaluer la charge globale dans le système.

La simulation est réalisée en utilisant le langage C. La figure 7.11 représente le résultat du calcul (cf. équation (7.4)) de variation des flux d'utilisateurs dans les deux configurations considérées (ex. Conf.1 et Conf. 2) en fonction des longueurs des rues. Les résultats présentés montrent que les flux d'utilisateurs évoluent avec le même profil

pour les deux configurations. Les valeurs augmentent avec les longueurs des rues jusqu'à une certaine limite et diminuent ensuite rapidement. Pour les deux configurations, les flux d'utilisateurs sont importants dans les rues qui sont localisés dans les zones denses (par exemple pour les rues qui ont des longueurs entre 0.08 et 1.3 km) là où les flux atteignent 1000 utilisateurs par seconde dans chaque cellule. Nous notons qu'il existe des rues où les flux d'utilisateurs sont maximaux et ces rues correspondent à un axe dynamique de fort trafic. Les valeurs des flux d'utilisateurs atteignent une valeur maximale de 1.3 utilisateur/cellule/seconde pour une longueur des rues égale à 0.52 km dans la configuration 1 (1.1 utilisateur/cellule/seconde pour des rues de longueur égale à 1.33 km dans la configuration 2). Au delà de ces valeurs maximales (longueurs des rues >1.44 km) les flux d'utilisateurs diminuent et tendent vers des valeurs minimales (respectivement 1 utilisateur/cellule/seconde dans la configuration 1 et 0.7 utilisateur/cellule/seconde dans la configuration 2). Nous pouvons conclure que les flux d'utilisateurs diminuent pour des rues plus longues que pour certaines autres qui pourraient être hors de la zone dynamique des deux configurations : rues loin des zones denses dans chacune des deux configurations.

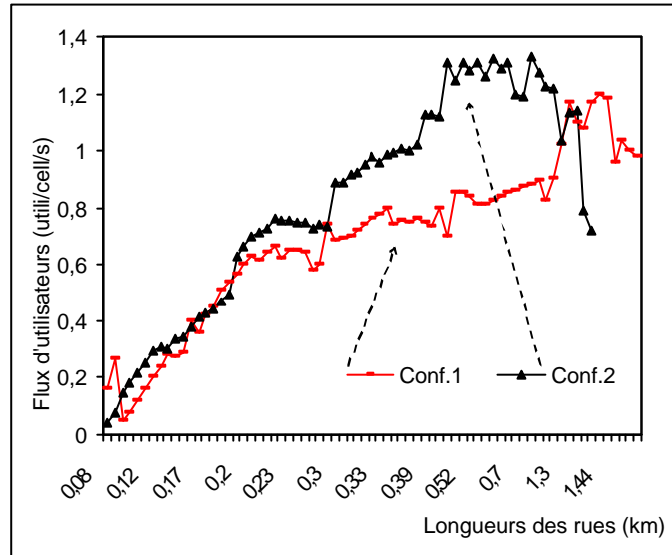


Figure 7.11 Flux d'utilisateurs $f_{tk}^{(j)}$ en fonction des longueurs des rues l_{tk}

En fonction des résultats obtenus dans la figure 7.11, les flux d'utilisateurs peuvent être modélisés par les équations suivantes conformément de ce qu'on trouve dans la littérature [LESIE] :

$$f_{tk}^{(j)} = (f_{(0.08)} + l_{tk} e^{\frac{l_{tk}}{l_a}}) \text{ util/cell/s} \quad (7.13)$$

où : $f_{(0.08)}$: variable

Le calcul des flux d'utilisateurs basé sur la figure 7.11, nous mène à l'évaluation de la densité d'utilisateurs par unité de surface (km^2). Les résultats sont présentés dans la figure 7.12 et montrent que la densité est maximale pour des vitesses d'utilisateurs moyennes (12 et 27 km/h). Par exemple pour la configuration 1, la valeur de la densité est égale à 0.07 utilisateur/ km^2 pour une vitesse qui varie entre 12 et 27 km/h. Cette

valeur devient beaucoup plus importante ($0.13 \text{ utilisateur/km}^2$) dans la deuxième configuration. Au-delà de ces valeurs ($>27 \text{ km/h}$), les valeurs de la densité diminuent et tendent vers une valeur minimale égale à $0.02 \text{ utilisateur/km}^2$). Nous concluons que la densité augmente pour des utilisateurs se déplaçant avec des vitesses moyennes (près des zones de services denses où les rues ne sont pas trop longues (voir figure 7.10) et diminue si les utilisateurs se déplacent avec des vitesses plus importantes (puisqu'ils sont lointains des zones de services denses). En effet, les résultats de la figure 7.11 sont logiquement complémentaires à ceux présentés sur la figure 7.12.

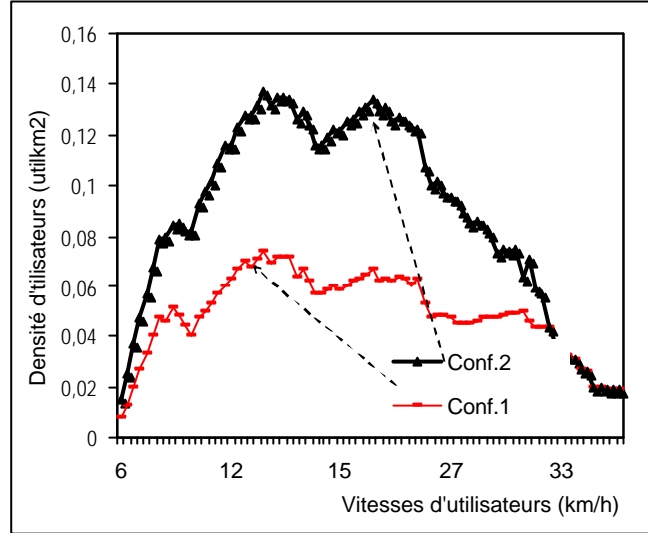


Figure 7.12 Densités d'utilisateurs $s_{ik}^{(j)}$ en fonction des vitesses v_{ik}

En fonction des résultats obtenus dans la figure 7.12, les densités d'utilisateurs peuvent être modélisées par les équations suivantes conformément à ce qu'on trouve dans la littérature [LESIE] :

$$s_{ik}^{(j)} = (s_{(6)} + v_{ik} e^{\left(\frac{v_{ik}}{v_a}\right)}) \text{ util/km}^2 \quad (7.14)$$

où : $s_{(6)} = 0.01$

Les transitions entre les cellules formalisées par l'équation (7.5) sont présentées dans la figure 7.13. Par exemple, pour la configuration 2 et avec une vitesse variant entre 6 et 33 km/h, les flux de transitions augmentent avec l'augmentation de la vitesse. Ils atteignent une valeur maximale constante égale à $0.04 \text{ utilisateur/cellule/seconde}$ dans la configuration 2 pour une vitesse égale à 27 km/h . Au-delà de ces valeurs, les flux de transitions entre cellules sont ceux des utilisateurs véhicules rapides qui se déplacent de la cellule courante à la cellule destination d'une façon régulière. Ils circulent dans les grandes rues, et avenues, placées généralement loin des zones de services denses (cf. figure 7.12). Les taux de transitions peuvent être bas selon la régularité de trafic. Pour les flux de trafic de faible vitesse, les véhicules se déplacent généralement avec moins d'ordre et moins de régularité entre cellules. Ces véhicules se déplacent dans un secteur dense où la densité des utilisateurs est haute (cf. figure 7.11) par exemple dans les petites rues, zones représentatives, zones administratives, etc. Cependant, le trafic le moins régulier, est caractérisé par des vitesses basses qui peuvent engendrer des flux de transitions assez bas.

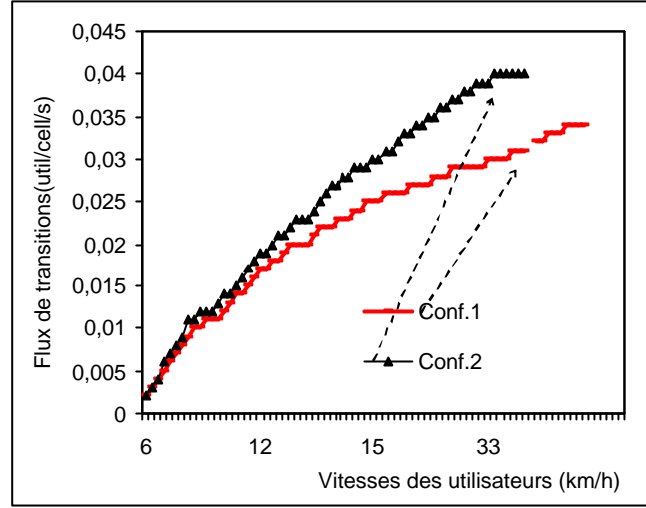


Figure 7.13 Flux de transitions $\Theta_{tk}^{(j)}$ entre cellules en fonction des vitesses v_{tk}

En fonction des vitesses d'utilisateurs (cf. figure 7.13), les flux de transitions peuvent être modélisés par les équations suivantes conformément à ce qu'on trouve dans la littérature [LESIE] :

$$\Theta_{tk}^{(j)} = (\Theta_{(6)} + d(v_{tk})^2) \text{ util/cell/s} \quad (7.15)$$

où : $d_{cte} = 10^{-4}$

$\Theta_{(6)} = 0.018$

7.6 Dimensionnement réseau et modèles de trafic

Nous nous concentrons sur les aspects de dimensionnement (voir figure 7.16) qui sont utilisés pour calibrer la capacité du système et optimiser la quantité de ressources allouées. Le dimensionnement du réseau (ex. Conf.1, cf. figure 7.5) utilise ainsi le calcul des taux de handovers en fonction de la charge de trafic selon les paramètres de mobilité \mathbf{a}_j (cf. équation (7.9)) calculés pour chaque type de service demandé.

Quatre scénarios (cf. chapitre 3, voir tableau 7.4) sont considérés dans la simulation :

- Scénario 1 : modèle voix ;
- Scénario 2 : modèle vidéo-téléphonie ;
- Scénario 3 : modèle www ;
- Scénario 4 : modèle ftp.

Pour chacun des ces quatre scénarios, nous calculons les valeurs des taux de handovers et de charge de trafic [BEG03] dans chaque cellule et dans le système entier afin d'avoir une couverture globale de tout le réseau. De plus, nous supposons que dans chaque cellule la probabilité de blocage de nouveaux appels est $P_b = 1\%$ et que la vitesse d'utilisateurs varie entre 6 et 33 km/h.

	Modèle de communication téléphonique	Modèle de flux multimédia : vidéo - téléphonie	Modèle de flux www	Modèle de flux ftp
Caractéristiques et statistiques	<ul style="list-style-type: none"> • Taux moyen d'appel : $I_{appel} = 5.510^5 \text{ appels}$ • Durée d'un appel : $\frac{1}{m_{appel}} = 150 \text{ s}$ • Durée de silence : $\frac{1}{m_{silence}} = 1.35 \text{ s}$ • Durée d'activité : $\frac{1}{m_{activité}} = 1.00 \text{ s}$ • Débit : 4 à 13 kb/s 	<ul style="list-style-type: none"> • Débit : 32 à 384 kb/s • Taux de séquence vidéo/jour : 4 • Durée moyenne d'une séquence vidéo : 120 s 	<ul style="list-style-type: none"> • Débit : 8 à 2048 kbit/s • Nombre d'appel : 5 appels/session • Temps de lecture : 4 à 12 s • Nombre de paquets : 10 p/appel • Durée inter-arrivée : 1.95 à 500 ms 	<ul style="list-style-type: none"> • Ex. Taille de paquet (dist. Cauchy) : 170 à 2140 octets • Durée : 0 à 20 min

Tableau 7.4 Modèles de trafic

Comme mentionné, nous optimisons les paramètres de handover de façon à maintenir et assurer une qualité de service en fonction des caractéristiques des sites GSM existants (ex. Conf.1, cf. figure 7.6), des trafics multimédia écoulés dans ces sites et des critères de déclenchement de handover.

Les compteurs OMC s'incrémentent à chaque fois qu'une action se déclenche dans le réseau. Dans notre travail nous nous intéressons aux compteurs liés au handover et spécifiquement ceux qui nous signalent l'existence d'un problème, soient les compteurs de nombre de handovers échoués, de nombre de coupures de communication sur handover, de nombre de tentatives de handovers inter et intra-cellulaires, etc. (voir annexe E).

Les figures 7.14 et 7.15 montrent que les taux de handover voix sont plus importants que ceux calculés dans les trois autres scénarios (vidéo-téléphonie, www et ftp). Ces valeurs atteignent une valeur maximale égale à $1e^{-3}s^{-1}$ pour les valeurs des taux de transitions $0.04s^{-1}$. Ces valeurs de taux de handovers qui sont calculées avec une vitesse $>27 \text{ km/h}$, correspondent aux zones périphériques [T.TE02][BEG01]. Au-delà de cette vitesse, les taux de handover pour les différents trafics restent presque constants. On peut expliquer cela en supposant alors que les utilisateurs sont des véhicules rapides qui se déplacent d'une façon régulière.

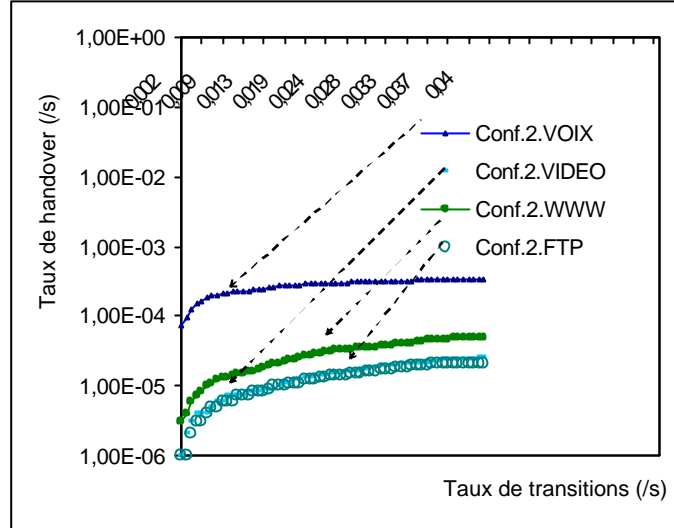


Figure 7.14 Taux de handovers I_H^j en fonction des taux de transition a_j
 $P_b = 1\%$ - vitesse : 6-33 km/h ; Conf. 1

Les taux de handovers en fonction des paramètres de mobilité (cf. figure 7.14) sont modélisés par les équations paraboliques suivantes conformément à ce qu'on trouve dans la littérature [LESIE] :

$$I_j^H = (I_{(0,002)}^H + d \cdot (a_j^2)) s^{-1} \quad (7.16)$$

où : $d = cte$, $I_{(0,002)}^H = \text{variable}$

La figure 7.15 montre que le trafic voix est considéré comme étant plus prioritaire que les autres trafics : vidéo-téléphonie, www, ftp. Il nécessite donc de plus de canaux au niveau de la station de base pour supporter une charge de trafic qui varie entre 50 et 100 mErl (voir figure 7.15).

Nous remarquons aussi que les deux scénarios, www et ftp, ont la charge de trafic la plus élevée selon la figure 7.15 (153 et 270 mErl). Nous déduisons que pour le trafic www ou ftp, le nombre des appels nouveaux augmente avec l'augmentation du nombre de handovers, cela peut mener à un risque de dégradation des services les plus prioritaires tels que la voix et la vidéo-téléphonie, et donc à une diminution de la qualité de service dans le système.

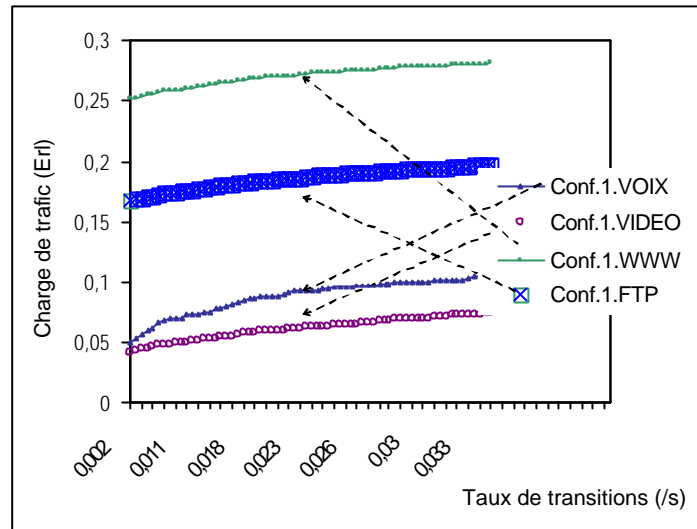


Figure 7.15 Charge de trafic r_j en fonction des taux de transitions a_j -
 $P_b = 1\%$ - vitesse : 6-33 km/h ; Conf.1

Les charges de trafic en fonction des paramètres de mobilité (cf. figure 7.15) peuvent être modélisées par les équations suivantes conformément à ce qu'on trouve dans la littérature [LESIE] :

$$r_j = d^*(a_j + r_{(0.002)}) \text{ Erl} \quad (7.17)$$

où : $d^* = cte$; $r_{(0.002)}$ = variable

Les paramètres du réseau radio de la configuration considérée [T.TE02][BEM03] (ex. Conf.1) sont résumés dans le tableau 7.4. Nous remarquons qu'avec une probabilité de blocage des appels nouveaux évaluée à 1% dans chaque cellule et des mesures réelles venant du terrain et du système, nous pouvons valider nos résultats de simulation en calculant la charge globale de trafic voix (le seul trafic pour lequel nous avons des mesures réelles). Cette valeur trouvée par la simulation est très proche de la valeur mesurée entre 12h à 13h pour la dite configuration.

M= 50 cellules – Capacité =1400 canaux – ex. Conf.1				
Flux de transitions globaux Θ (utilisateurs/cellule/s)	51.883 par simulation			
Taux de transitions globaux a (s^{-1})	1.650 par simulation			
	voix	Vidéo- téléphonie	www	ftp
Taux de handovers globaux I^H (s^{-1})	2.7310^{-3} par simulation	6.6510^{-4} par simulation	4.1210^{-3} par simulation	6.610^{-4} par simulation
Charge de trafic totale r (Erl)	6.60 par simulation	7.787 par mesure réelle	4.611 par simulation	20.154 par simulation

Tableau 7.5 Résultats globaux de simulation ; Conf.1

Les modèles de mobilité et de trafic sont appliqués aux deux configurations considérées. Ils nous ont permis d'évaluer les différents paramètres de performance : la charge de trafic, les taux de handovers, le niveau de service (GoS), etc. (voir figure 7.16).

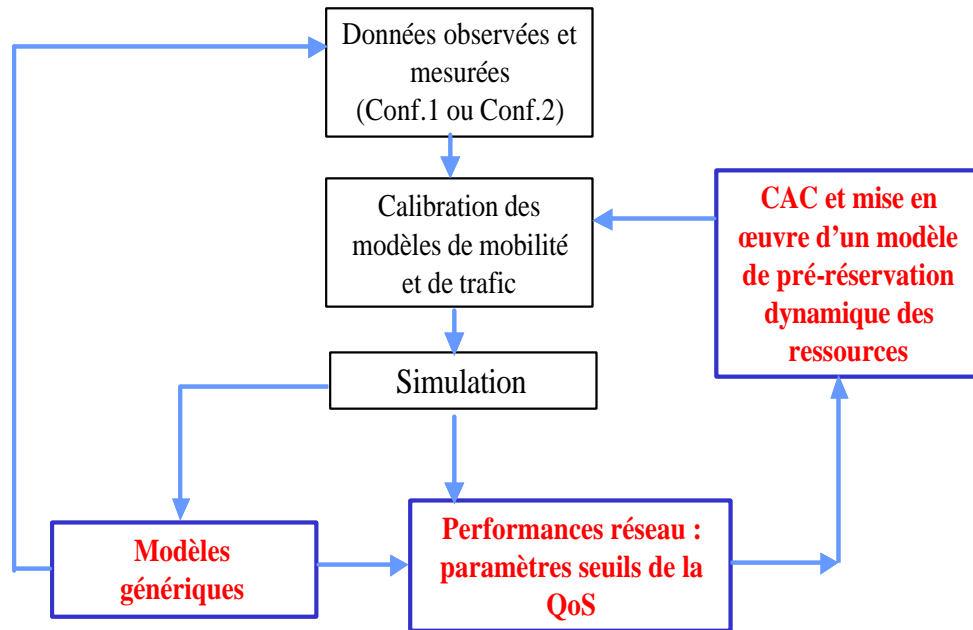


Figure 7.16 Résultats de simulation

Nous remarquons ainsi que les courbes obtenues pour ces différents paramètres dans les deux configurations opérationnelles Conf.1 et Conf.2 ont des profils similaires et les mêmes caractéristiques. En effet, nous pouvons conclure que ces modèles peuvent être appliqués à d'autres zones de services semblables. Les formules génériques conclues pour chaque figure restent valables et elles peuvent être calibrées en fonction des paramètres suivants : période de mesure, rayon de zone de services, distance entre stations de base, etc.

7.7 Conclusion

Dans ce chapitre et dans le contexte général de planification des réseaux radio mobiles cellulaire multimédia, nous avons développé un modèle de mobilité dans quatre modèles de trafic tels que voix, vidéo-téléphonie, ftp et www. Ce modèle est basé sur la théorie de graphe. Il est calibré par des mesures réelles récoltées du système et du terrain de deux zones de services opérationnelles (ex. Conf. 1 et Conf. 2)

Ce modèle de mobilité nous a amené à déterminer les différents paramètres de performance réseau tels que le taux de handovers, la probabilité de rejet d'appel, la charge de trafic, les niveaux de service, etc. Ces paramètres seront utilisés par la suite comme critères de la qualité de service pour l'acceptation des appels multimédia nouveaux (cf. chapitre 8). L'idée consiste à développer un mécanisme d'allocation de ressources, dont l'objectif est de maintenir, autant que possible les valeurs de paramètres de la qualité de service continues, conformes aux seuils fixés pour le réseau GSM existant.

Au cours d'une simulation, les résultats graphiques obtenus au sein de ce chapitre nous ont amené à former des formules génériques. Ces formules peuvent être utilisées par l'opérateur réseau radio. Ils peuvent les exploiter pour modéliser les paramètres de performance réseau pour d'autres zones de services semblables. Il suffit de les reparamétriser en fonction des paramètres suivants : période de mesure, rayon de zone de services, distance entre stations de base, etc.

Dans le chapitre suivant, nous allons exploiter les paramètres de performance réseau conclus. Ces paramètres seront considérés comme des valeurs seuils d'acceptation d'un appel. Si l'allocation de ressources est acceptée au niveau du réseau radio de l'appelant, alors le contrôle d'admission sera lancé entre domaines afin de déterminer si la qualité de service requise peut être assurée de bout en bout.

Références Bibliographiques

- [BAT98]Bathelt, A., “*Grafical models for analysis of design mobile wireless networks*,” PhD thesis ENST, 14 April 1998.
- [BEG03]Ben Rejeb, S., Choukair, Z. and Tabbane, S., “*Resource allocation and QoS management for cellular multimedia network*,” Proc. GRES’03, Brazil, February 2003, pp. 3-13.
- [BEM03]Ben Rejeb, S., Tabbane, S. and Choukair, Z., “*Mobility model for multimedia mobile radio network*,” Proc.MMNS’03, LNCS 2839 September 2003, pp. 183-195.
- [DAW99]Dawood, A.M. and Ghambari, M., “*Universal mobile telecommunications system (UMTS)*,” Selection procedures for the choice of radio transmission technologies of the UMTS; IEEE transactions on multimédia, Vol. 1, No.1, March 1999.
- [ETSI98]ETSI, “*Universal mobile telecommunications system (UMTS)*,” Selection procedures for the choice of radio transmission technologies of the UMTS; UMTS 30.03, version 3.2.0, April 1998.
- [T.TE02]Tunisia Telecom, “*Network planning mobile radio system*,” Tunis-Hached and Sfax cities; Technical Report, May 2002.
- [TAB97]Tabbane, S., “*Location management methods for third-generation mobile systems*,” IEEE Communication , Magazine, August 1997.
- [LESIE]Lesieur, L., Lefebvre, J. “*Analyse Mathématique*,” Armand Colin, Collection U- Paris. 1967.

Chapitre 8

**Gestion de QoS dans un réseau radio
mobile cellulaire****8.1 Introduction**

La gestion de l'allocation des ressources consiste à décider de l'admission d'un appel nouveau en fonction de la disponibilité des ressources, de qualité de service et également en fonction des possibilités de dégradation de services les plus demandeurs en ressources tel que par exemple la vidéo-téléphonie.

Dans ce chapitre, un mécanisme de contrôle d'admission d'appel est proposé lors de l'allocation des ressources afin d'adapter et d'assurer un service multimédia mobile tout en optimisant l'usage des ressources de l'opérateur. Ce mécanisme est implanté au niveau du gestionnaire des ressources (RM) à l'entrée du domaine radio. Il est basé essentiellement, sur l'état du réseau où il est implanté et plus particulièrement sur les paramètres seuils statiques conclus du modèle de mobilité (cf. chapitre 7). Ces paramètres sont considérés comme des informations contractuelles au sein du *Service Level Agreement* (SLA). Le SLA est adopté par la suite pour gérer les niveaux de qualité de service possibles (haut, moyen et bas).

L'objectif principal de ce mécanisme est de maximiser le taux d'acceptation des appels et l'utilisation des ressources tout en contrôlant la charge du réseau de telle sorte que les différents niveaux de garantie de services demandés soient satisfaits.

Dans ce chapitre, nous allons présenter tout d'abord les paramètres seuils de qualité de service obtenus à partir des modèles de trafic et de mobilité développés dans le

chapitre précédent. Puis nous exposerons les fonctions liées au contrôle d'admission d'appel local. Nous terminerons ce chapitre par l'application de ce contrôle d'admission à notre approche en exposant la technique de l'allocation des ressources locale pour des services multimédia tel que par exemple la vidéo-téléphonie.

8.2 Performance réseau - paramètres de QoS

Les performances du système en terme d'acceptabilité des appels dans le respect des contraintes de qualité de service sont décrites par le degré de service GoS qui traduit la probabilité qu'un appel soit bloqué ou retardé pendant une durée supérieure à un seuil (i.e., la probabilité de rejet d'appel handover, la probabilité de blocage d'un appel nouveau) [KIM98][LUO99][BEG03][BEM03][MAR]. Ces résultats constituent le prolongement des résultats déjà obtenus au sein des modèles de trafic et de mobilité développés dans le chapitre 7 (voir figure 8.1).

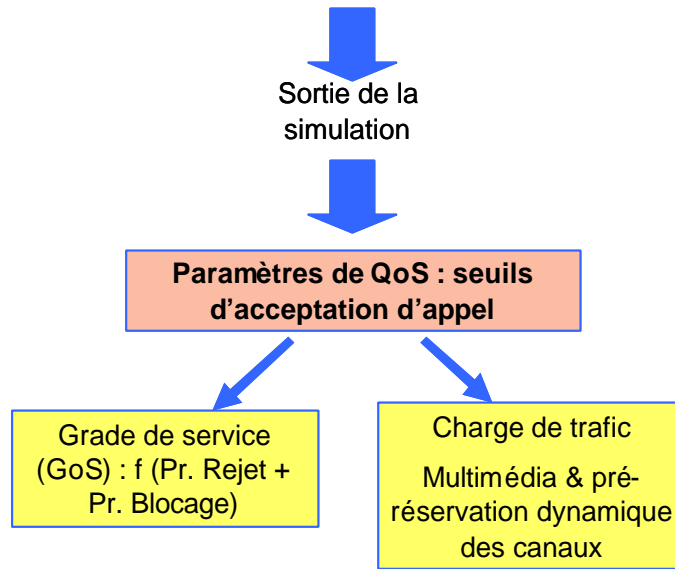


Figure 8.1 Sortie de simulation - paramètres seuils d'acceptation d'appel

8.2.1 La probabilité de rejet d'appel handover

Un utilisateur mobile peut traverser plusieurs cellules durant sa communication. A chaque changement de cellule, le système risque de perdre la communication en cours suite à un échec de handover. Afin de connaître la probabilité avec laquelle une communication arrive à terme sans interruption prématurée, nous calculons la probabilité de rejet d'appel en handover. Cette probabilité pour un type de trafic demandé après un $k^{\text{ème}}$ handover est donnée par l'expression suivante [BEG03][MAR] :

$$P_{P_j^H}(k) = [P_j^H (1 - P_{bj})]^{k-1} P_j^H P_{bj} \quad (8.1)$$

Avec : $k=1,2,3,\dots$

Et la probabilité de rejet sera donc exprimée par :

$$P_{dj} = \sum_{k=1}^{\infty} P_{h_i}(k) = \sum_{k=1}^{\infty} [P_j^H (1 - P_{bj})]^{k-1} P_j^H P_{bj} = \frac{P_j^H P_{bj}}{1 - [P_j^H (1 - P_{bj})]} \quad (8.2)$$

où $P_j^H = \frac{a_j}{a_j + m_{trafic}}$ (cf. chapitre 7) est la probabilité de handover interrompu avant de l'accomplissement d'appel.

L'équation (8.2) permet de tracer la figure 8.3 qui donne la probabilité de rejet d'appels en handover de quatre types de trafics en fonction de la charge r_j (Erl). Par exemple, pour le rejet des appels handovers voix, nous observons des faibles valeurs de probabilité dans la zone où la charge est faible. Ces valeurs demeurent quasi constantes à partir d'une certaine valeur de la charge ≥ 0.19 Erl. Cette probabilité est considérée comme étant une limite au delà de laquelle la plupart des appels nouveaux pourraient être rejetés.

Nous remarquons ainsi que la voix admet la probabilité de rejet la plus faible, égale à $7 \cdot 10^{-5}$, par rapport aux autres types de trafic et plus particulièrement par rapport au trafic vidéo qui admet la probabilité la plus grande qui est de l'ordre de 0.031. Cette dernière valeur peut engendrer une dégradation ou un blocage du service. Ainsi, la valeur de la probabilité de rejet pour les appels voix ($7 \cdot 10^{-5}$) engendre un meilleur degré de service (plus faible) égal à 0.027 alors qu'il est de 0.035 pour le trafic vidéo-téléphonie (voir figure 8.3).

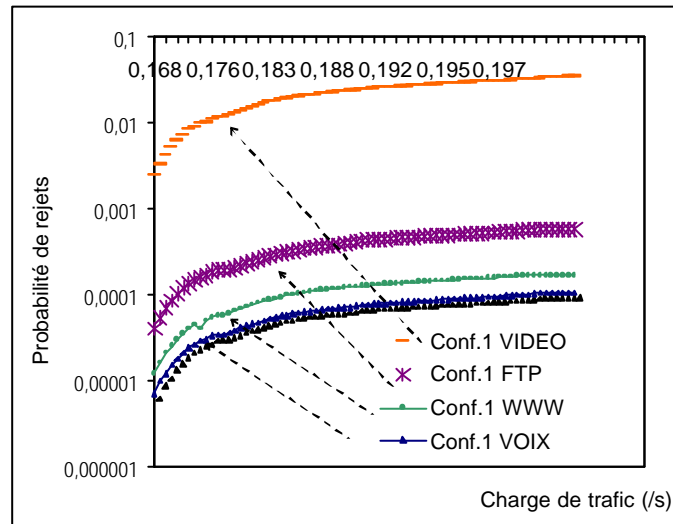


Figure 8.2 Probabilité de rejet d'appels handovers P_{dj} en fonction de la charge de trafic r_j
 $P_b = 1\%$ - vitesse : 6-33km/h ; Conf.1

8.2.2 Le degré de service GoS

Les critères de qualité de service les plus importants pour les utilisateurs de réseaux radio mobiles sont la disponibilité de ressources et leur fiabilité. La disponibilité de ressources est caractérisée par la probabilité de blocage d'appels nouveaux P_{bj} . Cette probabilité est souvent estimée au sein des travaux à 1% dans chaque cellule. La fiabilité de la liaison est décrite par la probabilité de rejet d'appel P_{dj} d'une communication en cours. Le degré de service est défini comme étant la somme pondérée de ces deux probabilités [KIM98][BEG03][BEM03].

Comme résultats conclus du processus de simulation effectué dans le chapitre 7, la figure 8.4 illustre la variation du degré de service en fonction des taux de mobilité a_j (taux de transitions entre cellules). Les courbes obtenues montrent que l'augmentation du taux de mobilité entraîne une augmentation du degré de service qui se traduit notamment par l'augmentation de la probabilité de rejet d'appel pour chaque classe de trafic et plus particulièrement pour le trafic vidéo-téléphonie [BEG03][BEM03] (3,5% dans le cas de la figure 8.3 : valeur maximale pour un taux de mobilité de $0.033s^{-1}$, calculé pour une vitesse supérieure à 33 km/h). Pour un taux de mobilité de $0.002s^{-1}$, calculé pour une faible vitesse de 6 km/h, le degré de service GoS est également faible et varie entre 0.2% et 0.5% pour les quatre types de trafic. Nous pouvons donc conclure que le meilleur GoS est donné pour une faible mobilité.

Nous observons d'après la courbe de la figure 8.4 que les valeurs de GoS sont plus élevées pour les trafics vidéo vu qu'ils admettent une probabilité de rejet d'appel importante (voir figure 8.2). De plus, de par ses besoins en ressources, le trafic vidéo-téléphonie peut provoquer une forte dégradation de la qualité des autres trafics et une gestion de ressources est alors nécessaire pour améliorer le degré de service et par conséquent la qualité de service.

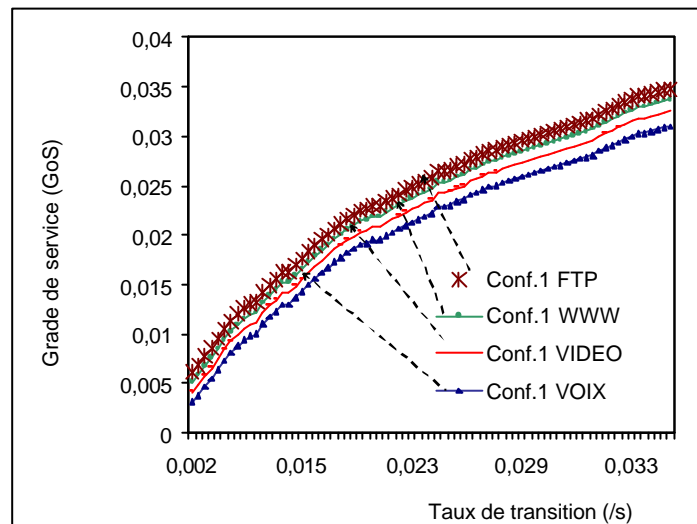
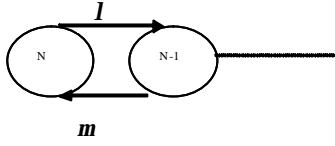


Figure 8.3 GoS_j en fonction des taux de transitions a_j
Vitesse : 6-33km/h ; Conf. 1

8.2.3 Pré-réservation dynamique des canaux (DPCr)

Afin de gérer de façon efficace la réservation de canaux multimédia, nous avons adopté un mécanisme de réservation dynamique des canaux DCPPr [LUO99]. Nous tenons compte des caractéristiques et des paramètres qui jouent sur la réservation des ressources, trouvés par simulation tels que la charge et les taux de handovers. Les valeurs considérées sont celles calculées pour la configuration 1 du réseau réel (cf. chapitre 7) [T.TE02][BEG03]. L'objectif majeur de ce mécanisme est d'offrir un service prioritaire pour les handovers afin de diminuer les probabilités des appels interrompus d'une communication en cours et de garantir un degré de service acceptable.

La réservation de canaux avec le mécanisme DCPr est une réservation de type *pool* [LUO99] qui se fait d'une façon aléatoire sans mémoire et qui dépend des taux de handovers entrants/sortants entre cellules. Le nombre de canaux N réservé suivant le GoS obtenu dans le réseau est donné par un calcul conforme au diagramme de transitions suivant, qui donne lieu aux équations globales suivantes (voir figure 8.4) :



$$\begin{aligned} \mathbf{m}p_0 &= \mathbf{l}p_1 \\ \mathbf{m} + \mathbf{l}p_i &= \mathbf{m}p_{i-1} + \mathbf{l}p_{i+1} ; i = 1, 2, \dots, N \\ \mathbf{l}p_N &= \mathbf{m}p_{N-1} \end{aligned}$$

Figure 8.4 Diagramme de transition

L'état régulier de probabilité est donné par l'équation suivante :

$$P[X = i] = \frac{(1 - \mathbf{r})\mathbf{r}_j^i}{1 - \mathbf{r}_j^{N+1}} \quad (8.4)$$

D'après [BEM03], la charge de trafic est donnée par l'équation suivante :

$$\mathbf{r}_j = \frac{\mathbf{l}_j + \mathbf{l}_j^H}{\mathbf{m}_{trafic} + \mathbf{a}_j} \quad (8.5)$$

Pour obtenir un degré de service GoS acceptable, chaque station de base détermine le nombre minimal N_j de canaux qui peut satisfaire les demandes des usagers :

$$P[X = 0] = \frac{(1 - \mathbf{r}_j)\mathbf{r}_j^i}{1 - \mathbf{r}_j^{N+1}} \leq P_{d_j} \quad (8.6)$$

$$N_j \geq \ln \frac{P_{d_j} + \mathbf{r}_j - 1}{P_{d_j}} / \ln \mathbf{r}_j \quad (8.7)$$

Un appel nouveau ou un handover d'une cellule à une autre cherche à se connecter à la BTS la plus proche. Chaque BTS d'un réseau cellulaire est souvent sollicitée par plusieurs demandes avec différentes caractéristiques de trafic et des usagers qui s'attendent à ce que leurs demandes soient satisfaites avec un degré de QoS acceptable.

Considérons un réseau mobile cellulaire (ex. Conf.1, cf. chapitre 7) qui supporte des caractéristiques de trafic différentes tels que la voix, la vidéo-téléphonie, www et le ftp. Pour chaque type de trafic, la réservation de canaux se fait selon les taux d'appels nouveaux entrants ainsi que les taux de handover entre cellules. Les résultats calculés sont trouvés par simulation et sont présentés dans le tableau 8.1 :

service	Nombre de canaux N réservés par BTS dans une cellule	Nombre de canaux N à réserver selon les spécifications de l'ETSI
voix	1.7	1
vidéo-téléphonie	5.79	5
www	3.89	4
ftp	4.46	4

Tableau 8.1 Pré-réserve dynamique de canaux DPCr dans la Conf.1

Les résultats obtenus sont présentés dans le tableau 8.1 et comparés avec les recommandations de l'ETSI [MAR]. L'ETSI propose pour chaque appel de transmission de données l'allocation de quatre canaux, pour chaque appel voix l'allocation d'un canal et pour chaque appel vidéo la réservation de cinq canaux (1 canal voix + 4 canaux de données).

Notons que le nombre de canaux disponibles dans chaque cellule est égal à 28 (ex. Conf.1, tableau 8.2). Nous supposons que l'établissement de connexions exige un nombre de canaux pour chaque type d'appel selon les résultats calculés au sein du tableau 8.1. Nous utiliserons ces valeurs par la suite comme premier critère de refus d'admission des appels nouveaux en cas d'indisponibilité de canaux.

8.3 Contrôle d'admission d'appel local

Pour traiter le problème de gestion des ressources, nous considérons un gestionnaire de ressources (*Resource Manager* ; RM) implanté dans un routeur de bordure, à l'entrée du réseau radio, au niveau du SLA (ex. Conf.1, voir figures 8.5 et 8.6). Ce gestionnaire de ressources communique avec les stations de base pour décider de l'acceptation des demandes d'appels caractérisées par les valeurs des paramètres de mobilité et de qualité de service. Lors d'un appel qui se traduit par une invocation de ressources, le gestionnaire de ressources doit décider si l'occupation du réseau permet d'accepter l'appel pour le niveau de qualité de service demandé : haut, moyen, bas. Il s'agit d'optimiser l'acceptation des appels pour maximiser l'utilisation de canaux en contrôlant la charge du réseau de telle sorte que les différents niveaux de garantie de services demandés soient satisfaits.

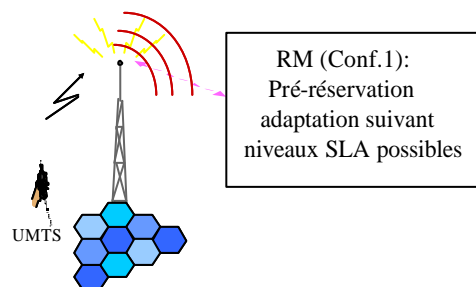


Figure 8.5 Réseau radio mobile (Conf.1) - Contrôle d'admission local

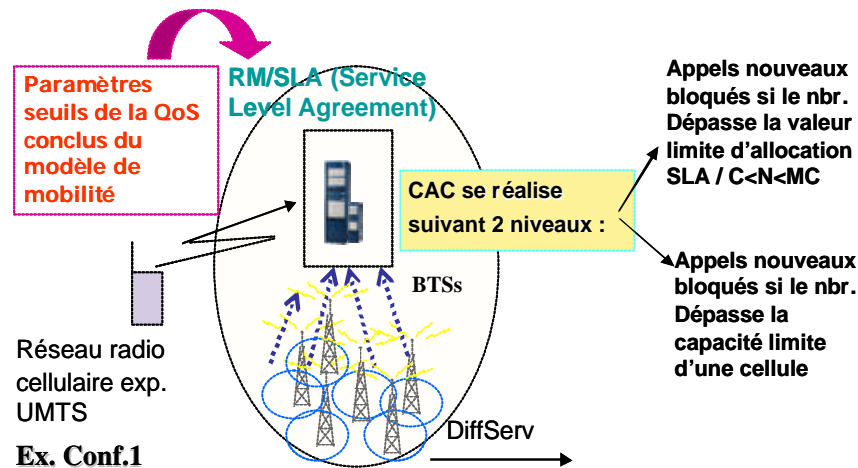


Figure 8.6 Données statiques récupérées du réseau cellulaire
Contrôle d'admission d'appel, BTSs/SLA1 ; Conf.1

Le contrôle d'admission d'appel (CAC) décidé par ce gestionnaire est donc basé sur l'état du réseau et plus particulièrement sur la probabilité de rejet d'appels handovers, critère pour une garantie de qualité de service. En effet, il est préférable pour la satisfaction des utilisateurs de bloquer les appels nouveaux plutôt que d'interrompre un appel en cours, suite à un échec de handover. Pour cela, les appels handovers doivent être servis en priorité par rapport aux appels nouveaux. Plusieurs travaux [CHE02][NAG96] ont étudié ce problème de priorité et ont proposé des solutions en réservant deux canaux de garde dans chaque station de base pour les appels handovers afin de diminuer la probabilité de rejet d'appels en cours et d'évaluer par la suite le degré de service GoS, considéré comme étant l'un des indicateurs clés de la disponibilité de ressources et de la fiabilité de connexion.

Le contrôle d'admission dans le domaine est donc réalisé suivant deux niveaux. Les appels nouveaux sont bloqués par le contrôle d'admission si le nombre des appels dépasse la valeur limite d'allocation de ressources dans le SLA ($C < N_j < MC$) ou encore le nombre d'appels dépasse la capacité limite d'une cellule. Dans ce cas, les appels handovers sont rejetés ainsi que les appels nouveaux [CHE02][NAG96].

Afin d'offrir une qualité de service garantie par le SLA, certains travaux proposent, par exemple, pour le cas d'un trafic multimédia, de coder la vidéo selon trois niveaux [CHE01][CHEN02] : L_1, L_2, L_3 . Le SLA utilise donc trois configurations contractuelles (1,2 et 3) pour offrir trois niveaux de QoS possibles (haut, moyen et bas) (voir tableau 8.4). Chaque niveau correspond à une probabilité de perte (t_1, t_2 et t_3) engendrée par le tampon du routeur de bordure ainsi qu'une bande passante effective optimale développée dans [CHE01][CHEN02] : b_1, b_2 et b_3 ($b_1 > b_2 > b_3$).

8.3.1 Application à notre approche

Le contrôle d'admission d'appel (CAC) pour l'allocation des ressources est effectué par un agent RM, basé essentiellement sur notre configuration d'étude et dont nous spécifions le rôle et l'activité conformément à notre approche (ex. Conf.1, cf. figure

8.6). Il doit connaître le nombre de stations de base, les ressources allouées pour chaque service et dans chaque station de base, et le niveau de service pré-alloués par le SLA. Ainsi il est informé de l'état du réseau (cf. chapitre 7) : modèle de mobilité d'usagers, statistiques des modèles de trafic, paramètres de qualité de service, etc.

Le contrôle d'admission d'appel se fait donc au niveau du RM du domaine de la Conf.1 (cf. figure 8.6) pour chaque type de trafic, suivant des seuils pré-définis (cf. figures 8.2 et 8.3 tableaux 8.2 et 8.4). Notre domaine d'étude contient $M = 50$ cellules. Chaque cellule peut supporter $C = 28$ appels. Les appels nouveaux sont rejetés par le domaine du contrôle d'admission si le nombre de ces appels est supérieur à $n = 1120$ (80% du capacité totale du domaine [CHE01]) d'appels du service courant dans le domaine (voir tableau 8.2).

Afin de valider notre travail en utilisant les paramètres trouvés (voir figures 8.2 et 8.3, tableaux 8.2 et 8.4) comme seuils statiques pour le contrôle d'admission d'appels, nous comparons nos résultats de simulation à une technique d'allocation des ressources, réalisée dans le même cadre et avec le même type de trafic et avec une réservation de canaux de garde pour les appels handovers [CHEN02][NAG96][OLI97] (voir figure 8.7). Ensuite, nous utilisons une technique de dégradation de la bande passante selon les niveaux spécifiés au sein du SLA pour offrir une QoS dégradée (voir figure 8.7, tableau 8.3) suivant trois niveaux de service : haut, moyen et bas.

/ domaine	Conf.1 (domaine 1)			
Nombre de cellules (dans le domaine)	50			
Capacité d'une cellule	28			
Nombre des appels nouveaux (80% de la capacité d'un domaine)	1120			
Nombre des appels handovers (10% de la capacité d'un domaine)	140			
La probabilité de blocage des appels nouveaux	1%			
/domaine	voix	vidéo-téléphonie	www	ftp
Les taux de handovers (s^{-1})	2.7310^{-3}	6.6510^{-4}	4.1210^{-3}	6.610^{-4}
La probabilité de rejet d'appel	0.005	1.608	0.008	0.027
Le grade de service GoS	0.078	1.521	0.081	0.098

Tableau 8.2 Données statiques et paramètres de QoS récupérés du domaine (Conf.1)-
Résultats de simulation

Ce travail est effectué en utilisant par exemple le modèle de trafic vidéo-téléphonie, qui est le plus difficile à servir par rapport au trafic voix du point de vue débit et durée de connexion [CHEN02][BEM02] (voir tableau 8.3).

/ domaine	bibliographie
Nombre des cellules (dans le domaine)	20
Capacité d'une cellule	20
Nombre des appels nouveaux (80% de la capacité d'un domaine)	320
Nombre des appels handovers (10% de la capacité d'un domaine)	40
	vidéo-téléphonie
Le taux de service moyen d'un appel (unité du temps)	0.5
Le taux de service dans la station de base (unité du temps)	0.1
La probabilité de rejet d'appel	0.09
La charge de trafic dans le domaine (Erl)	20

Tableau 8.3 Paramètres de QoS - bibliographie

Le processus de simulation dans le chapitre 7, nous permet par la suite de déduire la figure 8.7. Elle présente la courbe de la probabilité de rejet d'appels handovers vidéo-téléphonie en fonction de la charge r_j (Erl). Nous constatons que, dans le cas du rejet des appels handovers avec réservation de canaux de garde, les valeurs de la probabilité sont faibles dans la zone où la charge est faible. Nous déduisons qu'il y a une forte possibilité que le nombre d'appels nouveaux enregistrés dans le domaine de contrôle d'admission (domaine 1) soit inférieur à 1120 (voir figure 8.7, tableau 8.3). Donc le domaine peut continuer à accepter un nombre important d'appels [CHEN02] jusqu'à un certain seuil au-delà duquel la probabilité de rejet dépassera un certain seuil pour lequel il rejettera les appels nouveaux. Ainsi, nous remarquons que la probabilité de rejet d'appel augmente si la charge de trafic augmente. Par exemple pour une valeur de charge maximale, la plupart de requêtes nouvelles vont être rejetées par le domaine de contrôle d'admission qui est limité par une charge de trafic qui ne dépasse pas approximativement n/M [CHEN02], donc une limite pour la probabilité de rejet d'appel égale à 0.031 (0.000015 : voix, 0.0001 : www, 0.00015 : ftp (cf. figure 8.2)) [BEG03] et qui correspond à un GoS égale à 0.035 (voir figure 8.3). Cette valeur sera alors considérée comme étant une limite au-delà de laquelle la plupart de requêtes nouvelles pourraient être rejetées par le domaine de contrôle d'admission. Cette valeur de probabilité est calculée avec une probabilité de blocage des appels nouveaux de l'ordre de 1%. Elle sera considérée par la suite, ainsi que d'autres paramètres seuils, comme un paramètre seuil de négociation des ressources offertes par le gestionnaire des ressources (RM), le SLA de la configuration 1, et du réseau cœur DiffServ de l'opérateur.

La figure 8.7 représente les courbes de probabilités de rejet d'appel en fonction de la charge de trafic entre les simulations effectuées sur notre modèle de trafic et de mobilité et les données issues de notre étude bibliographique.

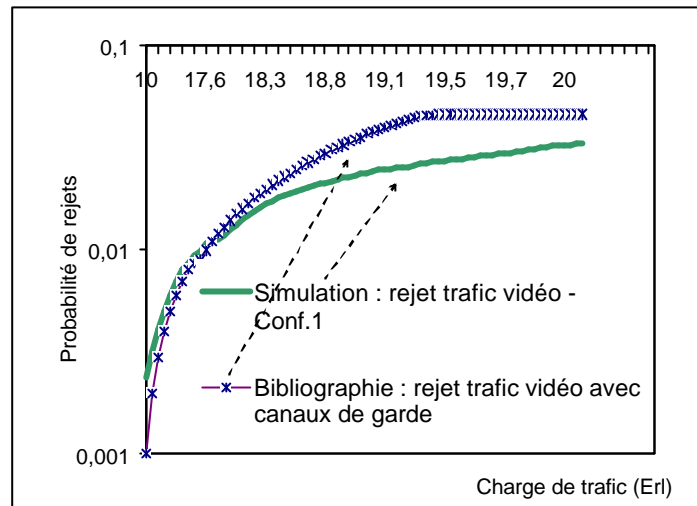


Figure 8.7 Contrôle d'admission d'appel pour le trafic vidéo-téléphonie - Comparaison : Bibliographie - simulation

8.3.2 Gestion d'allocation des ressources locale

La gestion de l'allocation des ressources consiste à décider de l'admission d'un appel nouveau en fonction de la disponibilité des ressources, de qualité de service et également en fonction des possibilités de dégradation de services les plus demandeurs en ressources. La disponibilité des ressources est prise en considération au niveau du sous-domaine de visibilité (SD : *Scope Domain*) du l'utilisateur mobile. Ce sous-domaine de visibilité est constitué par les stations qui détectent la demande de l'utilisateur. Le domaine radio admet un contrôle d'admission local (LARC) [CHEN02]. Le LARC gère l'allocation des canaux dans chaque cellule. Par exemple, si la charge de trafic vidéo-téléphonie/cellule est faible (0.1 et 0.176 Erl, voir figure 8.7) [OLI97] et que les ressources sont disponibles, alors un appel nouveau ou handover est admis avec une bande passante b_1 , c'est-à-dire avec un niveau de qualité de service élevé. Si la charge de trafic augmente ($> 0.176 < 0.2$ Erl) (voir figure 8.7) [OLI97], un appel nouveau ou handover ne pourra donc pas être admis avec une bande passante b_1 , le LARC fait diminuer alors l'allocation des ressources pour passer de b_1 à b_2 . Les appels sont alors acceptés par la suite avec un niveau moyen de qualité de service (bande passante égale à b_2).

Ce mécanisme de dégradation se résume donc comme suit (voir tableau 8.4) : si la charge de trafic augmente au delà d'un certain degré, le LARC diminue la valeur de la bande passante allouée pour le trafic existant (qui passera par exemple de b_1 à b_3) pour accepter globalement le maximum d'appels. Par la suite, si la charge diminue, les niveaux de services dégradés pourront recouvrir leurs niveaux initiaux par augmentation de la bande passante allouée qui pourra alors passer de b_3 vers b_1 [CHEN02].

Les paramètres indiqués au sein du tableau 8.4 sont par la suite utilisés pour paramétrer le domaine d'enregistrement du contrôle d'admission d'appel de la configuration 1. Ils constitueront les paramètres critiques d'acceptation des appels nouveaux. Le domaine utilise ces paramètres pour déduire la valeur de la bande

passante à allouer afin de diminuer les probabilités de blocage et de rejet d'appel. La valeur allouée pour chaque appel vidéo-téléphonie doit être ajustée entre b_1 , b_2 et b_3 . La capacité à allouer pour le trafic vidéo dans chaque cellule est égale à $C b_1$ paquets par seconde ou sinon $C (b_1 / b_2)$ paquets par seconde. Ainsi, le domaine de contrôle peut accepter jusqu'à $0.8M (C b_1)$ paquets par seconde avec un niveau de QoS haut, $0.8M (C b_1 / b_2)$ avec un niveau de QoS moyen et $0.8M (C b_1 / b_3)$ avec un niveau de QoS bas [CHEN02] (tableau 8.4).

Charge de trafic (Erl)	$0.1 < r < 0.176$	$0.176 < r < 0.2$	$r > 0.2$
Probabilité de rejet d'appel	0.01	0.022	0.031
GoS	0.005	0.02	0.035
Couches	L_1	L_2	L_3
Niveau QoS - SLA	Haut	moyen	bas
Probabilité de perte	$t_1 = 10^{-10}$	$t_2 = 10^{-4}$	$t_3 = 10^{-1}$
Bande passante effective	$b_1 = 142.45$ packet/s = 60.39 kbps	$b_2 = 119.82$ packet/s = 50.80 kbps	$b_3 = 112.65$ packet/s = 47.76 kbps
Contrôle d'admission d'appel (CAC)	$0.8 M(C b_1)$ = 159544 paquet/s = 0.6 gbps	$0.8 M(C b_1 / b_2)$ = 1331.53 paquet/s = 0.5 mbps	$0.8 M(C b_1 / b_3) =$ 1416.28 paquet/s = 0.6 mbps

Tableau 8.4 Paramètres seuils récupérés du réseau radio mobile du Conf.1 - Résultats simulation

8.4 Conclusion

Ce chapitre explicite notre approche de gestion de la qualité de service de l'allocation des ressources locale (i.e. niveau radio cellulaire). Cette approche est basée essentiellement sur l'état du réseau et plus particulièrement sur les paramètres seuils conclus du modèle de mobilité (cf. chapitre 7). Au-delà de l'allocation des ressources en local, le mécanisme de contrôle d'admission d'appel doit gérer l'allocation des ressources de bout en bout c-à-d entre domaines afin de garantir un des niveaux de qualité de service acceptable.

Pour gérer le problème d'allocation des ressources de bout en bout durant une requête, le réseau est constitué par un ensemble de domaines (i.e. radio, cœur, fournisseurs de services) qui participent à l'acheminement de cette requête. Dans chaque domaine, il existe un gestionnaire de ressources ayant les fonctions suivantes :

- Implantation des conditions d'accès aux ressources réseau (données seuils récupérées des différentes zones et enregistrées dans les SLA des routeurs de bord).
- Information des routeurs de périphérie de la présence de flux de trafic et de leurs caractéristiques.
- Interaction avec les agents courtiers des ressources d'autres domaines.

Les réseaux radio, les réseaux cœurs des opérateurs et les réseaux des fournisseurs de services, permettront d'établir un consensus entre les différents processus de décision des domaines participants à l'application. Ce consensus concerne la décision de réservation des ressources selon un des niveaux SLA possibles.

Références Bibliographiques

- [AJM01]Aimin, S. and al., “*Weighted fairness guarantee for scalable DiffServ assured forwarding*,” Communications, (ICC 2001), IEEE International Conference on, Vol.8, 2001, pp. 2365-2369.
- [BAK02]Bakiras, S. and Victor O.K.Li, “*Efficient resource management for end to end QoS guarantees in DiffServ network*,” Communications, (ICC 2002), IEEE International Conference on, Vol.2, 2002, pp. 1220-1224.
- [BEG03]Ben Rejeb, S., Choukair, Z. and Tabbane, S., “*Resource allocation and QoS management for cellular multimedia network*,” Proc. GRES’03, Brazil, February 2003, pp. 3-13.
- [BEA03]Ben Rejeb, S., Choukair, Z. and Tabbane, S., “*Modeling end-to-end QoS management and resource reservation for multimedia mobile radio network*,” Proc. AICCSA’03, July 2003.
- [BEM03]Ben Rejeb, S., Tabbane, S. and Choukair, Z., “*Mobility model for multimedia mobile radio network*,” Proc.MMNS’03, LNCS 2839 September 2003, pp. 183-195.
- [CHEN02]Cheng, Y. and Weihua, Z., “*DiffServ resources allocation for fast handoff in wireless mobile Internet*,” IEEE Communications Magazine May 2002, pp.130-6.
- [CHE01]Cheng, Y. and Weihua, Z., “*Optimal buffer partitioning for multiclass Markovian traffic source*,” Proc. IEEE Globecom’01, Vol. 3, 2001, pp.1852-56.
- [DAH01]Dahmouni, H., “*Conception and planification of mobile networks: video-telephony application on UMTS/IP*,” work elaborated for Philips Philips Consumer communications society, October 2001.
- [ELL02]Ellen, W.Z.. and al., “*QoS routing for anycast communications : motivation and an architecture for DiffServ networks*,” IEEE Communication Magazine, June 2002.
- [FRA01]Frank, Y. and Norvad, S., “*A priority-oriented call admission control paradigm with QoS re-negotiation for multimedia services in UMTS*,” Vehicular Technology Conference, VTC 2001 Spring. IEEE VTS 53rd, Vol.3, pp. 2021-2025.
- [IETF]http://www.ietf.org.
- [IETF1]http://www.ietf.org/html.charters/intserv-charter.html.
- [KIM98]Kim, J.S. and al., “*Determining velocity threshold for handoff control in hierarchically structured networks*,” Personal, Indoor and Mobile Radio Communications, the ninth IEEE International Symposium on, Vol.1, 1998, pp. 350- 354.
- [LUO99]Luo, X. and al., “*A dynamic channel pre-reservation scheme for handoffs with GoS guarantee in mobile networks*,” Computers and Communications, Proceeding. IEEE International Symposium on, 1999, pp. 404- 408.
- [MAR]Marsan, M.A. and al., “*Performance analysis of cellular mobile communication networks supporting multimedia services*,” This article was supported by the Italian National Research Council (CNR) and the Italian Ministry for University and Scientific Research.
- [NAG96]Naghshineh, M. and Anthony S.A., “*Design and control of micro-cellular networks with QoS provisioning for data traffic*,” Communications, ICC’96, Conference Record,

Converging Technologies for Tomorrow's Applications , IEEE International Conference on, Vol.1, 1996, pp. 464-468.

[NGU02]Nguyen, T.M.T. and al., "*COPS- SLS: a service level negotiation protocol for the Internet*," IEEE Communications Magazine, May 2002.

[OLI97]Olivier, T. W. Y and al., "*Adaptative resource allocation for prioritized call admission over an ATM- based wireless PCN*," IEEE Journal on selected areas in communications, Vol. 15, N° 7, September 1997.

[T.TEL02]Tunisia Telecom, "*Network planning mobile radio system*," Tunis Hached and Sfax cities; Technical Report, May 2002.

Chapitre 9

Algorithmes de consensus et approche consensuelle**9.1 Introduction**

L'approche de la gestion de la qualité de service de l'allocation des ressources de bout en bout doit respecter les contraintes liées à la diversité des besoins des applications tels que délai, gigue, pertes, bande passante, etc., et à l'intégration des différents média. Elle est mise en œuvre au niveau de serveurs de règles par des courtiers de ressources RM implantés au niveau des modules de prise de décision. Les prises de décisions concernent, pour chaque domaine, l'acceptation des demandes d'appels caractérisées par des paramètres de mobilité et de trafic et le niveau de SLA à offrir. Il s'agit pour l'opérateur d'accepter suffisamment d'appels pour maximiser l'utilisation des infrastructures tout en contrôlant la charge du réseau de telle sorte que les différents niveaux de garantie de service demandés soient satisfaits. A chaque session multimédia, il faudra décider si l'occupation du réseau permet d'accepter le niveau de qualité de service demandée par la session : bande passante, taux de perte, etc.

La qualité de service est définie par la façon de voir et juger le service de la part de l'utilisateur qui diffère de celle de l'opérateur. La gestion de la qualité de service demeure une tâche difficile en raison des divers types des mobilités des utilisateurs et des services. Pour décrire la vitesse et la fiabilité dans le cas de la transmission de données, la qualité de service se traduit par quelques paramètres tels que : le débit de transmission, le délai de transit, etc. Ces paramètres sont peut être suffisants pour définir la qualité de service quand le réseau de communication est utilisé principalement

pour la transmission de données sans contrainte temporelles. Mais lorsque les contraintes temporelles deviennent de plus en plus importantes pour ce type d'applications, il faut que tous les éléments du système (terminaux, BTS, routeurs, commutateurs, serveurs, etc.) participent pour assurer une performance garantie.

En s'appuyant sur les études et les approches adoptées au sein de [HUR99][CHAND96][RYA97][LAR00], le système de gestion de la qualité de service de l'allocation des ressources de bout en bout à définir est donc formé par un ensemble de nœuds connectés à travers un réseau de communication (i.e. les nœuds se synchronisent et collaborent pour échanger des messages à l'intérieur du réseau de communication). L'idée est de faire collaborer ces nœuds ensemble d'une façon consensuelle. Le consensus est implémenté dans le gestionnaire de ressources de chaque nœud pour décrire la négociation et la renégociation de la qualité de service. La renégociation est basée sur l'hypothèse que si les utilisateurs ne puissent pas acquérir les ressources nécessaires pour obtenir leur niveau de qualité de service le plus haut, ils peuvent spécifier qu'ils acceptent une admission à un niveau de service inférieur, plutôt que le blocage.

Dans ce chapitre nous présentons les caractéristiques des modèles de consensus proposés dans la littérature tels que le *DWD* (Consensus centralisé asynchrone avec échéance de temps ; *Decision with Deadline*), le *Losange S* (consensus distribué asynchrone avec échéance de temps en présence d'un détecteur de défaillance) et le *Frame work* (consensus asynchrone avec échéance de temps en présence d'un détecteur de défaillance). Puis, nous présentons les phases d'exécution de service pour le transférer au travers un réseau. Au cours de cet exécution un consensus devra être nécessaire pour l'accord sur la disponibilité de ressources entre domaines qui participent à l'acheminement de ce service. Pour ce faire, nous exposons par la suite notre approche consensuelle. Elle se base à la mise en place d'un consensus capable de contrôler l'allocation des ressources de bout en bout. Nous nous sommes limité au choix de protocole *Losange S* vu qu'il est compatible aux systèmes télécom distribués asynchrones temps réel. Nous terminons ce chapitre par l'application de notre approche consensuelle.

9.2 Les modèles de consensus

Les modèles de consensus proposés dans la littérature sont définis comme suit :

9.2.1 Consensus centralisé asynchrone avec échéance de temps (ex. DWD)

Le système est composé de n nœuds. Chaque nœud héberge un objet et est supposé avoir [RAY97] :

- Une mémoire locale ;
- Une horloge hardware ;
- Un local d'enregistrement stable.

Pour évaluer les performances du système, deux valeurs d'échéances sont fixées :

- L'instant d'échéance noté d ;

- La durée d'ordonnancement noté S .

A l'instant t , un objet veut diffuser une procédure aux autres objets du système dans le but d'avoir une décision (cf. figure 9.1). Pour cela, il définit ses propres échéances avant de commencer, puis il envoie un vote de type a ou de type b :

- Si l'objet reçoit x votes de type a , tout en considérant ses échéances et les échéances propres des autres processus, alors il décidera A ;
- Si l'objet reçoit y votes de type b , tout en considérant ses échéances et les échéances propres des autres processus, alors il décidera B ;
- Sinon, il va donner une valeur estimée moyenne E .

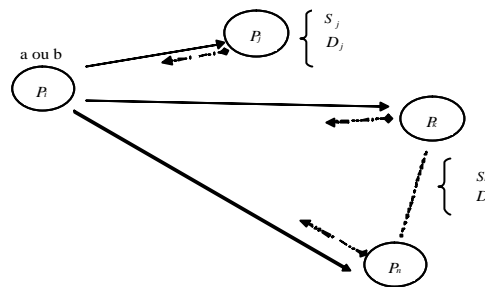


Figure 9.1 Consensus DWD

Les paramètres considérés lors de l'exécution de ce protocole sont présentés dans le tableau suivant :

L'identité du processus	p_i	p_j	p_k	...	p_n
L'instant de l'échéance	d_i	d_j	d_k	...	d_n
La durée d'ordonnancement	s_i	s_j	s_k	...	s_n

Tableau 9.1 Les paramètres considérés lors de l'exécution de protocole DWD

Le problème du consensus est caractérisé par les cinq propriétés suivantes :

- *Terminaison* : un objet non défaillant qui veut voter peut décider tout en considérant ses propres échéances ;
- *Validité* : la valeur décidée doit être A , B ou E ;
- *Coordination* (accord) : A et B sont deux valeurs de décisions exclusives, c'est à dire, si un objet décide A , un autre objet ne peut pas décider B ;
- *Justification* : si un objet décide A après avoir reçu x votes de type a , cela signifie que la diffusion et la décision se sont déroulées de façon correcte (c'est à dire que le message est arrivé à temps et avant même d'atteindre les échéances considérées pour chaque objet) ;
- *Obligation* : soit S , le dernier instant pour diffuser une procédure (pour une décision) et soit D , la première échéance à considérer, alors :
 - si x objets votent a , soit X vrai,
 - si y objets votent b , soit Y vrai,

- si $D-S \geq d$, soit X vrai, les objets sont corrects et la communication arrive à temps pour tous les objets : la décision finale doit être A ,
- si $D-S \geq d$, soit Y vrai, les objets sont corrects et la communication arrive dans le temps pour tous les objets : la décision finale doit être B .

L'algorithme se déroule en trois étapes :

- *Vote type a ou type b*: p_i diffuse une procédure $DWD (vote, S_i, D_i)$ avec D_i, S_i : les échéances considérées pour lui même et pour tous les processus.
- *Wait* : p_i reste en attente de recevoir :
 - réponse : x votes type a
 - ou réponse : y votes type b
 - ou réponse : un processus défaillant
- *Décide* : p_i décide sa propre valeur v_i (en tenant compte de l'état global) :
 - v_i : réponse : A (tout en considérant ses échéances)
 - v_i : réponse : B (tout en considérant ses échéances)
 - v_i : réponse : E indique si un ou plusieurs processus sont en panne

On suppose que :

- Il existe un seul processus /site (processeur).
- Chaque processus a une mémoire locale, une horloge physique, etc.
- Les messages sont diffusés à l'instant t . Si cet instant est supérieur à une échéance d alors le processus est considéré défaillant.
- Chaque processus peut voter en envoyant oui ou non aux autres processus.

L'exactitude des propriétés de ce protocole est définie par :

- *Terminaison* : un objet participe à une décision avec une procédure d'échéances, décide en tenant compte de ses propres échéances.
- *Validité* : la valeur décidée par un objet soit A , B ou E .
- *Coordination (accord)* : Il est impossible qu'un objet reçoive x propositions a et qu'un autre objet reçoive y propositions b .
- *Justification* : si un objet décide A , c'est que le dernier x objet est voté a .
- *Obligation* : cette propriété élimine le troisième cas qui considère la décision E . Il est possible que la décision soit A ou B .

La bonne configuration de ce protocole est caractérisée comme suit :

Soit X l'état de x objets qui votent a , alors :

- la durée entre la plus haute valeur de temps S pour commencer et la plus basse échéance D est $\geq d$,

- les messages échangés entre les objets doivent être dans \mathbf{b} temps c'est à dire que le temps de transfert doit être $<$ à une valeur bien définie \mathbf{d} ,
- les messages qui ont un temps de transfert $> \mathbf{d}$ sont considérés défaillants.

9.2.2 Consensus distribué asynchrone avec échéance de temps en présence d'un détecteur de défaillance (ex. Losange S)

Ce modèle a été discuté au sein de [HUR99][CHAN96][LAR00]. Soit le problème d'établissement d'un consensus avec détecteur de défaillance dans un système réparti asynchrone. Chaque processus peut, soit s'exécuter correctement, soit s'interrompre brutalement (et définitivement) suite à une panne. Il devient nécessaire d'apporter des solutions au problème d'accord entre processus. *Chandra* et *Toueg* ont augmenté le modèle asynchrone en introduisant la notion de détecteur de défaillance dont le but est de résoudre le problème de la distinction entre les processus lents, très lents et défaillants. Ils considèrent un détecteur associé à chaque processus qui est chargé de détecter les défaillances externes en définissant un délai appelé «délai de garde».

Chandra et *Toueg* ont défini huit classes de détecteurs de défaillance dont : S (le plus rapide), W , D , etc. Tous ces détecteurs sont caractérisés par les deux objectifs suivants :

- *Complétude* : détection des processus réellement arrêtés ;
- *Exactitude* : vise à limiter les suspicions erronées.

La classe S est la classe la plus faible qui peut résoudre le consensus en s'appuyant sur le paradigme de coordinateur tournant. Les objectifs visés par *Chandra* et *Toueg*, en construisant les algorithmes de consensus tout en se basant sur l'introduction des détecteurs de défaillances, sont les suivants :

- Faire en sorte que chaque processus vote en diffusant un message, soit pour favoriser une décision durant une tentative courante, soit pour abandonner cette tentative pour commencer une autre ;
- Maîtriser la terminaison des protocoles exécutés en gérant au mieux une durée maximale d'exécution ;
- Limiter dans le temps la durée des tentatives tout en synchronisant leurs déroulements ;
- Déterminer pour un temps maximal souhaité le meilleur couple (nombre de tentatives, durée de tentatives, etc.) ;
- Prendre en compte des paramètres caractérisant l'environnement d'exécution (fonction de répartition des délais de transmission de messages, probabilité de perte de messages, probabilité de pannes, etc.) pour réaliser ces calculs ;
- Assurer la vivacité du protocole (c'est à dire s'assurer que les défaillances soient bien détectées) ;
- Assurer la sûreté du protocole.

Le protocole s'exécute en effectuant des tours asynchrones (cf. figure 9.2). Chaque tour est coordonné par un processus qui joue le rôle d'un coordinateur ($p_i = p_c$). Chaque tour s'exécute en quatre phases :

- Phase (1): chaque processus envoie son estimation courante : $estimate_p(p, r_p, estimate_p, ts_p)$ au coordinateur c_p avec « r_p » le numéro de tour, « ts_p » le temps associé à la valeur estimée.
- Phase (2) : le coordinateur choisit une valeur estimée « $estimate_c$ » avec un temps associé le plus large. Cette valeur sera diffusée pour tous les processus.
- Phase (3) : chaque processus a deux possibilités pour répondre :
 - le processus reçoit la valeur $estimate_c$ et envoie un *ack* au c_p pour lui indiquer qu'il a adopté cette valeur pour sa propre estimation,
 - en consultant son propre détecteur de défaillance D_p , ce dernier lui indique que le (c_p) n'est peut pas voter, à ce moment là, le processus envoie un *nack* au c_p .
- Phase (4) : le coordinateur c_p attend les réponses *ack* ou *nack*. Si toutes les réponses sont *ack* alors c_p sait que tous les processus ont adopté leurs estimations à la valeur estimée $estimate_c$ et que cette dernière valeur sera considérée par la suite la valeur décidée.

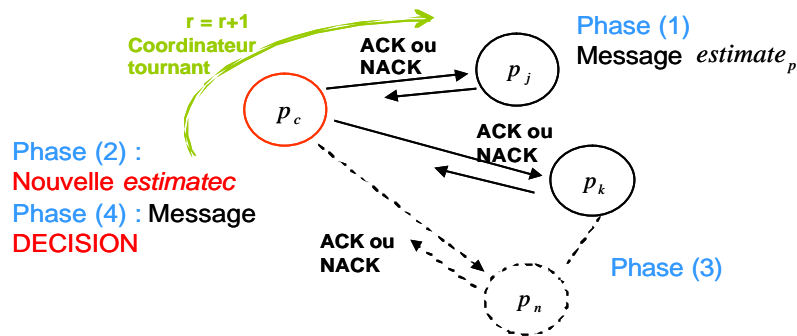


Figure 9.2 Consensus Losange S

Signalons ici que chaque processus est équipé par un détecteur de défaillance (D_p) qui permet d'indiquer les comportements de processus qui lui est associé. Les paramètres considérés lors de l'exécution de ce protocole sont mis dans le tableau suivant :

L'identité du processus	p_i	p_j	p_k	p_n
Le détecteur de défaillance	D_i	D_j	D_k	D_n
Le numéro du tour	r_i	r_j	r_k	r_n
La valeur estimée	$estimate_i$	$estimate_j$	$estimate_k$	$estimate_n$
L'instant de diffusion	ts_i	ts_j	ts_k	ts_n

Tableau 9.2 Les paramètres considérés lors de l'exécution de protocole Losange S

9.2.3 Consensus de type frame work

Pour résoudre les problèmes de coordination, [HUR99] propose un cadre général d'application (*frame work*) qui généralise les protocoles de coordination (d'accord). Il

est basé sur des fonctions principales qui sont reliées entre eux par des règles. Les fonctions sont définies par :

- *GET* : le processus commence par proposer une valeur par défaut appelée : \perp , puis le protocole demande à l'application de proposer une nouvelle valeur dont le but est d'avoir une valeur d'accord. Cette valeur d'accord peut être égale à \perp ou bien à une valeur proposée par n'importe quel processus. Au fur et à mesure que le protocole évolue, il devient difficile pour les processus de changer leurs propositions vu que le temps devient différent de zéro et le protocole arrête de faire appel à *GET*.
- $<$: lorsque la fonction est appelée deux fois ou plus par le protocole, la fonction $<$ équilibre un groupe des valeurs par rapport aux autres en se basant sur la signification du valeur.
- *F* : définit les valeurs à considérer dans le vote et les valeurs non considérées.
- *ACCEPTABLE* : c'est la fonction qui permet d'accepter des valeurs, de rejeter d'autres et de contrôler la signification des valeurs proposées pour acceptation ou refus.
- *EXCUSED* : cette fonction intervient lorsque le coordinateur ne peut pas attendre pour longtemps. Le manque de proposition provient d'un processus quelconque.

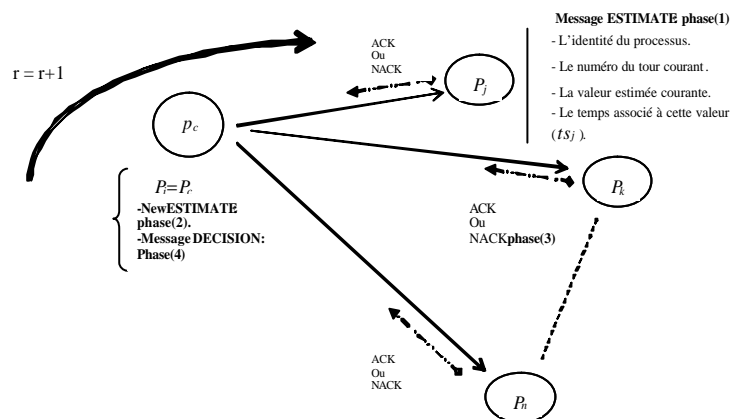
Le protocole est basé sur le principe de coordination tournant et a pour objectif de converger vers une décision finale. Les tours effectués par le protocole sont successifs asynchrones. Chaque nouveau tour est coordonné par un processus qui essaye d'imposer une valeur de décision. Chacun de ces processus p_i a sa propre variable locale est_i qui représente la valeur estimée courante. Un temps ts_i est associé à cette valeur.

Le protocole commence tout d'abord avec une valeur par défaut ($est_i = \perp$) qui correspond à $ts_i = 0$. Cette valeur sera changée au fur et à mesure de l'évolution du protocole et convergera vers la valeur finale. Durant le tour r , la coopération entre les processus est basée sur un schéma de communication centralisé : chaque message peut être envoyé ou reçu du coordinateur courant, le message qui sera envoyé pour décision durant le tour r ne sera pris en compte que par les processus qui s'exécutent durant le même tour. Chaque tour est divisé en quatre phases (voir figure 9.4) :

- Phase (1) : chaque processus envoie un message au coordinateur courant indiquant son estimation (noté p_c qui peut être un processus parmi les processus du système) dans le but d'avoir la valeur finale. Ce message envoyé est appelé «*messagesESTIM*» et contient quatre valeurs :
 - l'identité du processus qui a envoyé ce message par exemple p_i ,
 - le numéro du tour courant r ,
 - la valeur estimée courante est_i ,
 - le temps associé à cette valeur ts_i ,

Cette phase s'exécute au début du tour et à chaque fois qu'il y aura une nouvelle valeur proposée par un processus p_i .

- Phase (2) : tous les «*messagesESTIM*» sont envoyés vers le coordinateur courant. A chaque fois que ce coordinateur reçoit «*messagesESTIM*», il exécute les trois actions suivantes :
 - Il met les identités de processus qui ont envoyé les messages dans un état «*received_from_i*». Cela lui permet de connaître facilement les processus qui n'ont pas envoyé leurs messages.
 - Il adapte la valeur variable à une nouvelle valeur estimée «*new_est_i*», qui sera proposée par la suite pour tous les processus ;
 - Le message «*NEW_ESTIM* » qui sera diffusé par le coordinateur contient trois points : l'identité de l'envoyeur ($p_c = p_i$), le numéro de tour courant (r_i) et la nouvelle valeur estimée (*new_est_i*).
- Phase (3) : chaque processus attend la nouvelle valeur d'estimation qui provient du coordinateur. Il recevra soit un message qui indique que le coordinateur est en panne, soit la nouvelle estimation. Si un processus refuse cette nouvelle valeur, alors il envoie un acquittement négatif «*nack*» et s'il l'accepte il en envoie un positif.
- Phase (4) : le coordinateur attend des messages d'acquiescement de tous les processus. Si la majorité des messages reçus sont positifs alors il diffuse le message *Decision* qui contient la valeur décidée puis il recommence pour réaliser le 2^{ème} tour.

Figure 9.3 Consensus *frame work*

Les paramètres considérés lors de l'exécution de ce protocole sont mis dans ce tableau :

L'identité du processus	p_i	p_j	p_k	...	p_n
Le numéro du tour	r_i	r_j	r_k	...	r_n
La valeur estimée	$ESTIM_i$	$ESTIM_j$	$ESTIM_k$...	$ESTIM_n$
L'instant de diffusion	ts_i	ts_j	ts_k	...	ts_n

Tableau 9.3 Les paramètres considérés lors de l'exécution de protocole *frame work*

Le protocole est caractérisé par les deux propriétés suivantes :

- *Terminaison* : les processus proposent plusieurs valeurs significatives pour avoir une valeur acceptable.
- *Sûreté* : à travers ce protocole il y aura une possibilité de copier les différentes décisions prises par les coordinateurs durant les tours effectués.

Exemple : soit deux décisions prises par les deux coordinateurs du tour n et n_2 avec $n < n_2$. Quatre cas sont à considérer :

- les deux décisions sont prises après la collection de la majorité des estimations (*M-decision*),
- les deux décisions sont les premières décisions à considérer (*E-decision*),
- la première décision est *E-decision* et la 2^{ème} est *M-decision*,
- la première décision est *M-decision* et la 2^{ème} est *E-decision*.

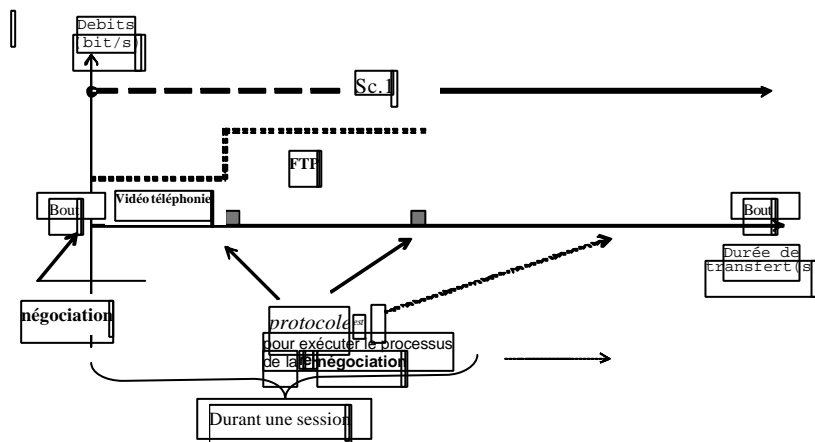
9.3 Phase d'exécution de services

9.3.1 Phase de négociation et de re-négociation d'un media

Les sessions multimédia manipulent différents types de média : voix, vidéo-téléphonie, etc. Chacun de ces types a des besoins spécifiques en terme de qualité de service (cf. chapitre 3). Pour transférer ces informations au travers d'un réseau, il est nécessaire que la session négocie avec le réseau hors ligne puis en ligne la qualité de service nécessaire pour satisfaire la demande désirée [BLA98][DUR00][CAR03][FLE02].

Lors d'une session d'appel, un utilisateur mobile peut établir et maintenir plusieurs connexions simultanément. Il est également possible de modifier ces caractéristiques durant une connexion grâce à une procédure de re-négociation.

Au niveau des schémas ci-après présentés, figurent des scénarios suivant le type, le débit et la durée du transfert d'une application demandée durant par exemple un appel :



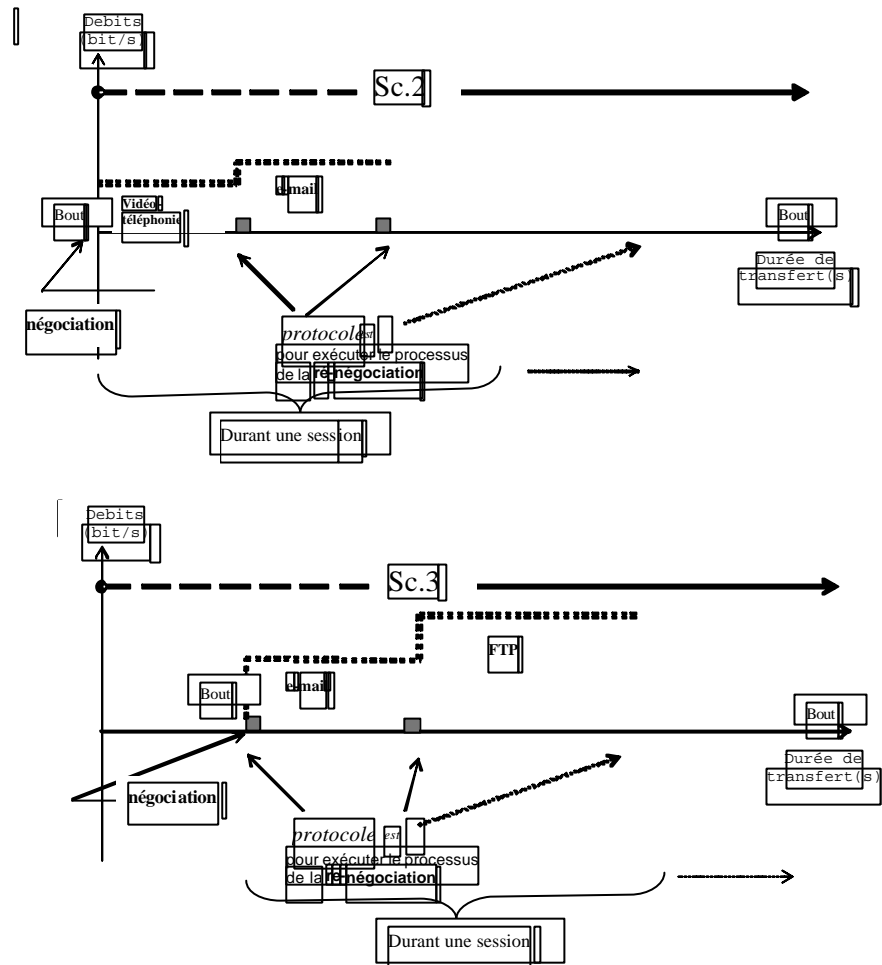


Figure 9.4 Phase d'exécution de services le long d'une session

Au cours de la phase de négociation ou re-négociation d'un média, un protocole d'allocation devra être nécessaire et qui pourra être utilisé pour déterminer :

- L'accord sur la disponibilité de ressources ;
- L'affectation équitable de ressources ;
- La négociation et la re-négociation ;
- Le contrôle de la qualité de bout en bout ;
- La détection des défaillances du point de vue du fonctionnement du protocole au niveau de chaque processus, pannes au niveau de l'exécution de logiciels tournants inter-blocage, faute de paramétrage, etc.

9.3.2 Les caractéristiques d'un flux multimédia

Un flux multimédia est une session d'appel à temps réel mis sous forme d'une succession des média. Ces média peuvent tolérer une certaine dégradation : ex. vidéo-téléphonie. Les média sont affectés par des niveaux SLA possibles dans chaque gestionnaire de ressources qui participe à la pré-réservation d'une route complète entre la source et la destination. Une plage commune décidée par tous les gestionnaires de ressources sera affectée à la session.

Les sessions arrivants au niveau de chaque gestionnaire de ressources sont considérées comme des média appartenant aux différentes classes de service (M1(S1), M1(S2), M1(S3), etc.) (voir figures 9.5 et 9.6). Ils se distinguent par exemple par :

- Des processus d'arrivée différents ;
- Des temps de services différents ;
- Un ordonnancement dans la file d'attente en fonction de leur classe; etc.

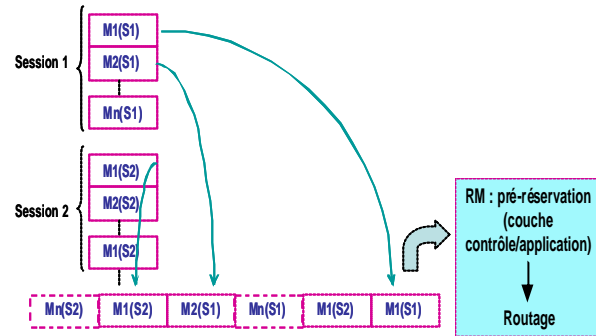


Figure 9.5 Ordonnement des média dans la file d'attente en fonction de leur classe

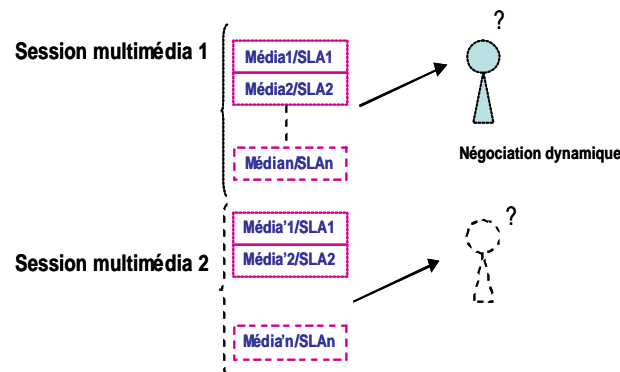


Figure 9.6 Sessions multimédia- négociations dynamiques temps réel

Dans la suite de notre travail et au niveau du chapitre suivant, les médias des classes 1 et 2 (*Conversational* et *Interactive*) sont considérés comme étant des services à faible bande (i.e. durée d'appel courte), alors que les médias des classes 3 et 4 (i.e. *background* et *Streaming*) sont considérés des services à large bande (cf. chapitre 3).

La qualité de service accordée aux médias des classes 1 et 2 doit être meilleure que celle accordée aux média des classes 3 et 4. Les premiers se verront accorder un service « *garantie* » tandis que les seconds auront un service « *au mieux* ».

9.4 Contrôle d'admission d'appel de bout en bout

L'allocation des ressources est basée sur un mécanisme de contrôle d'admission d'appels de bout en bout.

Le rôle majeur du mécanisme de contrôle d'admission d'appel est de contrôler les différents gestionnaires de ressources qui participent à l'acheminement d'une session demandée parce qu'il existe des contraintes à imposer dans le réseau, telles que :

- L'intégration des différents média ;
- La qualité limitée des ressources ;
- La diversification des besoins des applications (délai, gigue, pertes, bande passante, etc.).

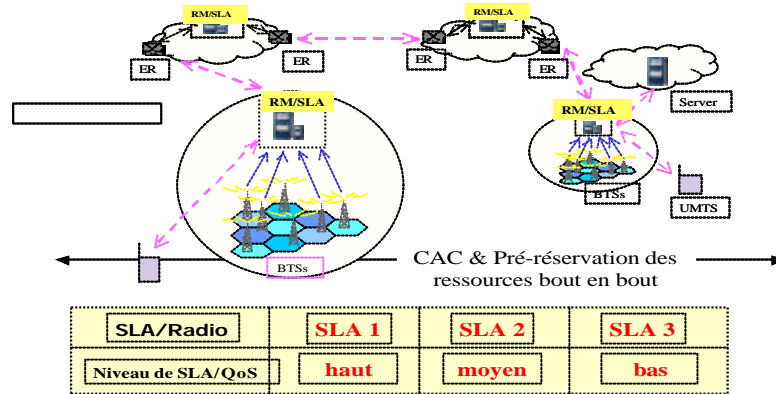


Figure 9.7 CAC/SLA - Pré-réserve des ressources intra et inter-domaines

Chaque gestionnaire de ressources participant à la décision construit sa propre politique qui lui donne une vue globale sur l'état du réseau (%SLA offert, pré-réserve des chemins, routages, le nombre des domaines qui participent à l'application, etc.) (voir figures 9.8 et 9.9, cf. Chapitre 10). Les routeurs de bordure (cf. chapitre 5) peuvent appliquer la politique correspondante pour admettre les flux multimédia pouvant accepter une qualité de service dégradée.

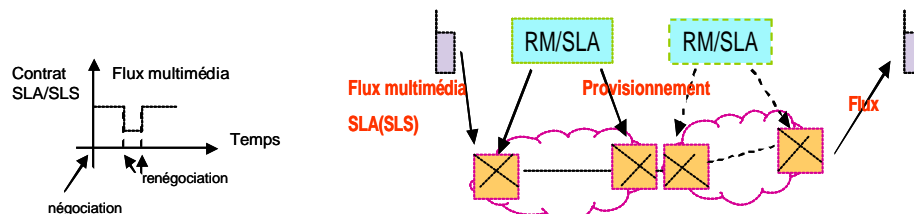


Figure 9.8 RM coordinateur - vue globale au sein de son domaine et les domaines voisins

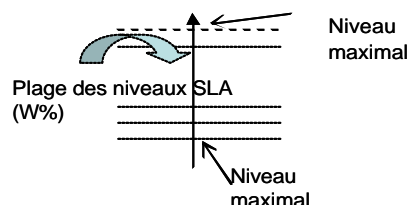


Figure 9.9 Plage commune décidée par tous les RMs

9.4.1 Table de routage

L'architecture DiffServ (cf. chapitre 5) définit quatre types d'éléments qui constituent le chemin à emprunter par la session multimédia lorsqu'elle passe par les routeurs [JIA01].

Ces éléments sont :

- Classificateur de trafic ;
- Élément d'actions ;
- Élément de mesures ;
- Gestionnaires de file d'attente.

Ces différents éléments peuvent être contrôlés par un système à base politique à temps réel.

Une table de routage (cf. tableau 9.4) est enregistrée au sein de chaque gestionnaire de ressources (RM). Cette table l'informe sur l'état et la capacité de ressources au sein de son domaine. Elle est représentée comme suit :

Route	Interface 1	Interface 2	Interface k
Route 1	C(1,1)résiduel	C(1,2)résiduel	C(1,n)résiduel
Route 2	C(2,1)résiduel	C(2,2)résiduel	C(2,1)résiduel
....
Route m	C(m,1)résiduel	C(m,2)résiduel	C(m,1)résiduel

Tableau 9.4 Capacité allouée par les interfaces routeurs

où $C(i,k)$ résiduel est la capacité allouée par les interfaces routeurs j .

9.4.2 Problématique

Si les ressources sont indisponibles ou bien non satisfaisantes auprès de quelques gestionnaires de ressources alors deux cas possibles se présentent :

- Soit le système change le chemin au niveau de ces domaines pour essayer d'en négocier d'autres domaines voisins, afin de satisfaire l'utilisateur ;
- Soit que le système accepte des niveaux de service inférieurs offerts par ces domaines en décidant un niveau de service moyen qui peut satisfaire l'utilisateur, en essayant de l'améliorer par la suite.

Dans la figure 9.10, le chemin réservé peut être alloué aux différents média de la même session. Chaque gestionnaire de ressources pourra allouer un niveau différent de SLA au même média. En effet, tout au long du chemin réservé, ces ressources peuvent être de niveaux différents du fait que les domaines sont hétérogènes et gérés par différents protocoles, ce qui engendre un problème de gestion des ressources allouées.

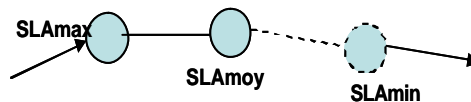


Figure 9.10 Niveaux différents de SLA au même média

Si l'un des gestionnaires ne dispose pas de ressources, le média peut être rejeté. Ce rejet peut engendrer par la suite le rejet de toute la session.

Plutôt que de rejeter toute la session, une solution pourrait consister à acheminer le média d'une même session par une autre route dans le même domaine associé au gestionnaire de ressources défaillant, soit par un domaine voisin en raison de la garantie insuffisante.

9.4.3 Approche consensuelle

Pour résoudre cette problématique, nous adoptons au sein de notre approche le mécanisme d'admission d'appel (CAC) dans les réseaux radio mobile afin d'améliorer la gestion des ressources en adaptant les ressources allouées aux applications multimédia (en terme de réquisition et d'acceptabilité des niveaux de service) et aux caractéristiques du réseau cellulaire.

Par conséquent, cette approche est basée sur un algorithme consensuel entre tous les processus de système. Cet algorithme peut être utilisé pour :

- Fixer les paramètres d'échéance de temps au cours de la diffusion d'un message entre les processus afin de distinguer les processus lents, corrects et défaillants ;
- Etudier les problèmes de consensus rencontrés lors d'une négociation, renégociation et répartition de ressources suivant les débits, durée et les scénarios d'usages considérés.
- Elaborer une adaptation des algorithmes distribués temps-réel aux spécificités des systèmes télécoms ;
- Sélectionner un algorithme dont la complexité reste cohérente par rapport aux contraintes spécifiées par les organismes de normalisation des télécoms ;
- Traduire les paramètres abstraits en des vecteurs de paramètres du réseau ;
- Mettre en place cet algorithme en se basant sur des techniques d'informatique de télécoms ;
- Concevoir un protocole qui gère le comportement des processus (défaillance, lenteur, etc.) lors d'une décision (consensus) pour envoyer ou recevoir un message avec un débit bien déterminé ;
- Trouver des solutions pour les pannes (les défaillances) qui peuvent être les causes d'un mauvais fonctionnement d'un protocole d'accord dans un processus par exemple, faire fixer un détecteur qui joue le rôle d'un *policing*.
- Etudier les problèmes de consensus rencontrés lors d'une négociation, renégociation et répartition de ressources suivant les débits, durée et les scénarios d'usages considérés.

A l'instant de la demande, le gestionnaire de ressources local définit le profil de l'utilisateur et la destination (il fixe l'adresse IP de l'utilisateur destination (i.e. le nombre de RM qui participent à l'application)) et fait un contrôle d'admission d'appel local au sein de son domaine. Si les ressources sont disponibles, le gestionnaire de ressources local établit un chemin au sein de son domaine puis envoie à l'aide de l'algorithme de consensus et le protocole de signalisation de bout en bout COPS (*Common Open Policy Services*) (cf. chapitre 5) un message RAV (vote pour allocation de ressources de bout en bout) auprès des autres domaines voisins :

- Si le gestionnaire de ressources local reçoit S votes type SLA1 (i.e. S nœuds qui ont votés SLA1 (i.e. SLAmax)), il vérifie les niveaux offerts puis il décide SLA1 ;

- Si le gestionnaire de ressource local reçoit S' votes type SLA2 (i.e. S' nœuds qui ont votés SLA2 (i.e. SLAmoy)), il vérifie les niveaux offerts puis il décide SLA2 ;
- Sinon, le gestionnaire de ressources local peut décider le niveau SLA3 (i.e. SLAmin).

L'algorithme de consensus s'exécute suivant trois étapes :

- Vote ;
- Attente ;
- Décision.

Il doit vérifier les cinq propriétés suivantes :

- Terminaison ;
- Validité ;
- Coordination ;
- Justification ;
- Obligation.

9.4.4 Discussion

Les protocoles de consensus cités ci-dessus seront utiles pour gérer l'allocation des ressources de bout en bout d'une session demandée (cf. chapitre 10), notamment nous sommes limité au choix de protocole *Losange S* vu que ce protocole est plus compatible avec les systèmes télécom distribués, asynchrones, temps réel par rapport aux deux autres protocoles de consensus (voir tableau 9.5) [BEA03].

Par exemple pour une demande d'un media d'une session demandée, on peut associer à chaque gestionnaire de ressource qui participe a ce consensus une base de donnée qui contient le profil de l'utilisateur, le niveau de SLA estimé, le temps de diffusion de vote, le temps de réception de vote, la valeur finale verrouillée et donc le chemin alloué au sein de son domaine. Chaque gestionnaire peut connaître la liste des coordinateurs tournants présents en cours de vote ainsi que le nombre des RM défaillants (i.e. le nombre des gestionnaires qui n'ont pas votés) (voir figure 9.11).

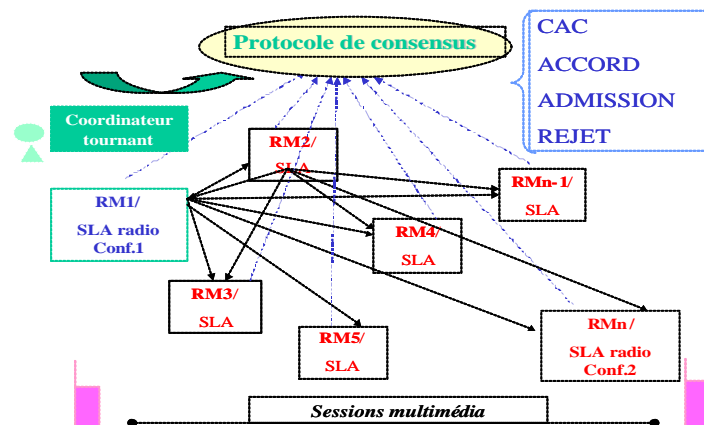


Figure 9.11 Mise en place de consensus *Losange S*

Ce protocole peut converger vers une décision finale, verrouillée dans un temps fini (ex. une plage % SLA minimale par exemple pour le premier média en essayant de voter pour un SLA plus haut plus tard). Il sera mis en oeuvre et comparé dans le chapitre 10 avec celui du *DWD* pour évaluer la rapidité de convergence vers un consensus portant sur la réservation d'un chemin dynamique avec un niveau de QoS acceptable.

	Avantages	Inconvénients
<i>Losange S</i>	<ul style="list-style-type: none"> – Exécution à temps réel. – Chaque RM peut voter : oui ou non. – Modèle asynchrone : chaque RM peut jouer le rôle de coordinateur de vote : modèle compatible avec les systèmes télécom (asynchrones, répartis,...). – La valeur décidée sera verrouillée après avoir lieu plusieurs tours asynchrones. 	<ul style="list-style-type: none"> – Même s'il y avait des RM défaillants, le système converge vers une valeur verrouillée.
<i>DWD</i>	<ul style="list-style-type: none"> – Exécution à temps réel. – Le système peut converger vers une décision finale : <ul style="list-style-type: none"> - si la majorité des réponses sont A, alors RM décide A. - si la majorité des réponses sont B, alors il décide B, - sinon, il décide E. – Chaque RM peut voter : oui ou non. 	<ul style="list-style-type: none"> – Modèle centralisé : un seul processus (le coordinateur) qui gère le vote auprès des autres processus. – Au début de l'exécution, le système fixe une échéance de temps globale. – Même s'il y avait des RM défaillants, le système converge vers une valeur verrouillée.
<i>Frame work</i>		<ul style="list-style-type: none"> – Modèle n'est pas compatible avec les systèmes distribué télécom. Il est considéré ni distribué ni centralisé. – Les processus proposent plusieurs valeurs significatives pour avoir une valeur acceptable. – Le système ne peut se converger vers une valeur finale verrouillée. – Chaque nouveau tour est coordonné par un processus qui essaye d'imposer une valeur de décision.

Tableau 9.5 Comparaison des modèles de consensus du point de vue compatibilité avec les systèmes télécom

9.4.5 Application de notre approche consensuelle

En se basant la littérature et les principes de base des modèles de consensus théoriques cités au niveau du paragraphe 9.2., nous pouvons appliquer notre approche de contrôle et de gestion de la qualité de service de l'allocation de ressources en utilisant par exemple le protocole *Losange S* qui sera comparé par la suite par rapport au protocole *DWD* du point de vue rapidité de convergence (voir chapitre 10).

9.4.5.1 *Losange S*

Le mécanisme *Losange S* se base sur le mécanisme d'un coordinateur tournant. Il se réalise en deux étapes. La deuxième étape s'exécute en 3 époques asynchrones (voir chapitre 10) :

Etape 1 : Contrôle d'admission

Conditions à respecter :

- Chaque RM peut estimer une plage des niveaux SLA ($w\%$) ;
- Chaque RM peut jouer le rôle d'un coordinateur. Donc il peut voter pour recevoir une décision sur un niveau SLA demandé ;
- Chaque coordinateur est distingué par un numéro ;
- Chaque rond est distingué par un numéro.

Etape 2 : Pré-réservation des ressources au sein de RM et établissement d'un chemin.

L'algorithme s'exécute suivant trois époques asynchrones. Chaque époque peut être réalisée par des ronds asynchrones :

- Première époque: chaque RM peut estimer une plage des niveaux SLA ($w\%$) ;
- Deuxième époque: une plage de SLA est décidée. Cette plage sera verrouillée par le coordinateur. Aucune autre plage de décision n'est possible ;
- Troisième époque: les RM décident la plage verrouillée.

Chaque rond s'exécute en quatre phases :

- Phase 1 : à l'instant de la demande, chaque RM estime une plage des niveaux SLA ($w\%$). Il envoie son estimation au coordinateur c ainsi que le numéro du rond dans lequel il l'a adopté ;
- Phase 2 : le coordinateur c_{RM_i} collecte les estimations. Il choisit une plage avec la plus large ts_{RM_i} . Il la renvoie par la suite à tous les RM en tant que leur nouvelle estimation $estimate_{RM_c}$;
- Phase 3 : chaque RM peut avoir deux réponses possibles :
 - Dès qu'il reçoit $estimate_{RM_c}$ envoyé par le coordinateur c_{RM_i} , il lui renvoie un «ack» pour lui indiquer qu'il a bien adopté $estimate_{RM_c}$ en tant que sa propre estimation ;
 - Si non il lui renvoie un «nack».

Phase 4 : le coordinateur c_{RM_i} attend les réponses «ack ou nack». Si toutes les réponses sont «ack», alors il est bien informé que la majorité des gestionnaires changent leurs propres estimations pour adopter la valeur $estimate_{RM_c}$. Cette valeur sera donc par la suite verrouillée.

9.4.5.2 DWD

L'algorithme DWD peut être réalisé en deux étapes (voir annexe D) :

Etape 1 : Contrôle d'admission

Conditions à respecter :

- La capacité et la plage des niveaux SLA ($w\%$) à allouer par chaque gestionnaire de ressources (ex. RM1) ;

- Avant de lancer le consensus, le système fixe une échéance de temps globale d ;
- RM1 diffuse un message de vote pour un niveau de qualité de service SLA (SLA1 ou SLA2 ou SLA 3) à fixer ;
- Après diffusion du message de vote, RM1 attendra Td pour recevoir l'ensemble des réponses ;
- RM1 recevra le message de vote à l'instant Tr . La durée de transfert $Tr-Td$ doit être $< d$;
- Si RM1 recevra les messages de vote avec une durée de transfert $Tr-Td > d$. Ce que ces messages sont considérés défaillants.

Etape 2 : Pré-réservation des ressources au sein de RM et établissement d'un chemin.

9.5 Conclusion

Dans ce chapitre nous avons défini les caractéristiques des trois modèles de consensus tels que le *Losange S*, le *DWD* et le *Frame work*. Nous nous sommes limité au choix de protocole *Losange S* vu qu'il est le plus compatible aux systèmes télécom distribués asynchrones temps réel. Ce consensus fixe des valeurs seuils critiques relatifs à la composante temporelle tels que les scénarios d'usage, la bande passante, la durée d'exécution, le délai de transfert, etc. et à la composante de qualité de service tel que la durée maximale d'établissement du consensus (voir chapitre 10). Ces valeurs seuils seront fixées au niveau du courtier des ressources RM afin de résoudre le problème d'allocation des ressources entre domaines : contrôle d'admission d'appel, pré-réservation des ressources, négociation, renégociation, etc.

Références Bibliographiques

- [BEJ96]Bejani, J.E., “*Supervision of QoS on real time distributed system*,” PhD thesis, ENST Paris, 24 October 1996.
- [BLA98]Blake, S. and al., “*An architecture for differentiated service*,” RFC 2475, December 1998.
- [BEA03]Ben Rejeb, S.,Choukair, Z. and Tabbane, S., “*Modeling end-to-end QoS management and resource reservation for multimedia mobile radio network*,” Proc. AICCSA’03, July 2003.
- [BEL00]Bellili, H., “*Radio resource allocation methodology for GSM-Multimedia networks*,” PhD thesis, U. Paris 6, 15 September 2000.
- [BHA01]Bhatnagar, S. and al., “*Providing quality of service guarantees using only edge routers*,” Global Telecommunications Conference, 2001. Globecom’01. IEEE, Vol.4, 2001, pp. 2542-2548
- [CAR03]Cardoso, K., “*COPS and QoS management for UMTS mobile network*,” Proc. GRES’03, pp. 227-239.
- [CHAN96] Chandra, T.D. and Toueg, S., “*Unreliable failure detectors for reliable distributed systems*,” Journal of ACM, Vol. 43, No. 2, March 1996, pp. 225-267.
- [CHAN99]Chandra, T.D., “*Unreliable failure detectors for reliable distributed systems*,” Journal of the ACM, Vol.43, N° 2, March 1996, pp.225-267.
- [DUR00]Durham, D. and al., “*The COPS (Common Open Policy Service) protocol*,” RFC 2748, January 2000.
- [HUR99]Hurfin, M. and Raynal, M., “*A simple and fast asynchronous consensus protocol based on a weak failure detector*,” Distributed Computing, 12(4), pp. 209-223, 1999.
- [LAR00]Larrea, M. and al., “*Optimal implementation of the weakest failure detector for solving consensus*,” Proceedings of the 19 th IEEE symposium on reliable distributed system (SRDS’02).
- [LEB98]Leboucher, R.L., “*Algorithm for real time distributed system*,” PhD thesis, ENST Paris, 28 September 1998.
- [MOS99]Mostefaoui, A. and Raynal M., “*Solving Consensus Using Chandra-Toueg's Unreliable Failure Detectors: a Generic Quorum-Based Approach*,” Proc. 13th Int. Symposium on Distributed Computing (DISC’99), (Formerly WDAG), Springer-Verlag LNCS (P. Jayanti Ed.), pp. 49-64, Bratislava (Slovakia), 1999.
- [RAY97]Raynal, M., “*Real time dependable decision in timed asynchronous distributed systems*,” Institut of Electrical and Electronics Engineer 1997.
- [FLE02]Flegkas, P. and al., “*A policy-based quality of service management system for IP DiffServ networks*,” IEEE Network, Special Issue on Policy-Based Networking, Mar./Ap. 2002, pp. 50-56.

Chapitre 10

**Gestion de QoS dans un réseau radio mobile
de bout en bout****10.1 Introduction**

Dans le chapitre précédent, et pour résoudre la problématique de l'allocation des ressources pour un flux multimédia donné, nous avons défini le mécanisme de contrôle d'admission d'appel de bout en bout. Ce mécanisme est basé sur un protocole de consensus tel que *Losange S*. Le rôle majeur de ce mécanisme est de contrôler les différents gestionnaires de ressources qui participent à l'acheminement d'une session demandée.

Nous nous intéressons dans ce chapitre à la mise en œuvre d'un algorithme de contrôle d'admission d'appel (CAC) entre les gestionnaires de ressources des différents domaines. Cet algorithme permet de faire converger ces gestionnaires vers un consensus portant sur la réservation d'un chemin dynamique avec un niveau de qualité de service acceptable. Le chemin réservé aux média est déterminé à la fois par la connaissance de ressources disponibles dans le réseau (les niveaux de service contractuels spécifiés au sein des SLA) et les demandes de la qualité de service de ces média.

10.2 CAC de bout en bout

Le gestionnaire d'allocation de ressources de l'appelant (ex. Conf.1 ; cf. chapitre 8) évalue la *Capacité d'Acceptation de Nouveaux Appels par Cellule* (CANAC) puis, dans le cas où cette capacité le permet, il étudie le respect des contraintes de qualité de

service (ex. r , P_d , GoS ; cf. chapitre 8, voir figure 10.1) une fois l'appel est accepté. Dans le cas où l'ensemble de ces conditions est satisfait, l'appel est considéré par le gestionnaire d'allocation de ressources de l'appelant comme étant acceptable (cf. chapitre 8) et il lance alors une demande d'acceptation d'allocation de ressources aux autres réseaux participant à la session (voir procédure CANAC) [BEA03].

Procédure CANAC(SLAI)

Si CANAC (SLA1) < (r , P_d , GoS)

Alors appel accepté en local avec SLA1 (niveau haut)

et lancement de la demande d'acceptation aux RM des autres domaines.

Elaboration d'un « consensus » pour SLA1

Sinon élaboration d'un « consensus » pour SLA2

Sinon élaboration d'un « consensus » pour SLA3

Sinon si CANAC (SLA2) < (r' , P_d' , GoS')

Alors appel accepté en local avec SLA2 (niveau moyen)

et lancement de la demande d'acceptation par les RM des autres domaines.

Elaboration d'un « consensus » pour SLA2

Sinon élaboration d'un « consensus » pour SLA3

Sinon si CANAC (SLA3) < (r'' , P_d'' , GoS'')

Alors appel accepté en local avec SLA3 (niveau bas)

et lancement de la demande d'acceptation par les RM des autres domaines.

Elaboration d'un « consensus » pour SLA3

Sinon appel rejeté.

Figure 10.1 Procédure CANAC(SLAI)

10.2.1 Mise en place de CAC de bout en bout et approche consensuelle

Une des étapes la plus importante du contrôle d'admission d'appel multimédia concerne le lancement d'un vote auprès des domaines concernés par l'application sur le niveau de service SLA (i.e. SLA1 (SLAmax), SLA 2 (SLAmoy) et SLA3 (SLAmin)) à réserver pour satisfaire l'appelant (voir figure 10.2). Cette décision doit être prise par les différents gestionnaires des ressources des différents réseaux qui participent à l'application. L'algorithme de consensus [RAY97][BEA03] fonctionne selon le principe suivant (cf. chapitre 9) :

- Le gestionnaire des ressources local de l'appelant RM (A) décide de son propre vote puis diffuse éventuellement un message de sollicitation de votes auprès des autres RMs ;
- Si RM (A) reçoit en retour x votes de type *acceptation* avant une certaine échéance qui tient compte de ses échéances et des échéances propres des autres RMs (voir chapitres 8 et 9), alors il décide de l'acceptation globale de la réservation de ressources ;
- Si RM (A) reçoit y votes de type *refus* avant une certaine échéance alors il conclut à un refus au niveau global ;
- Sinon, il donne une valeur estimée moyenne.

L'algorithme du consensus réparti sur les différents gestionnaires impliqués par une opération de réservation de ressources fixe des valeurs seuils critiques relatifs à la composante temporelle et à la composante de qualité de service au niveau du

gestionnaire des ressources RM (cf. chapitre 8) afin de résoudre le problème d'allocation des ressources entre domaines : réservation des ressources, négociation, renégociation, etc.

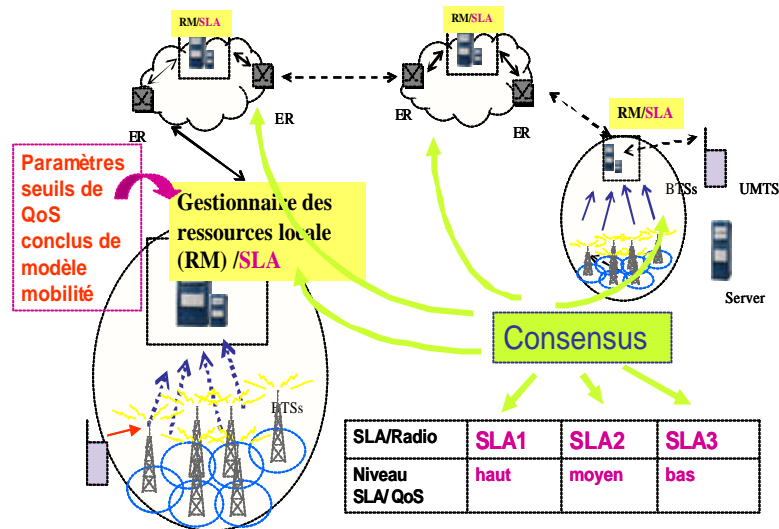


Figure 10.2 Mise en place de CAC de bout en bout

Nous proposons par la suite le schéma synoptique de contrôle d'admission d'appel de bout en bout suivant :

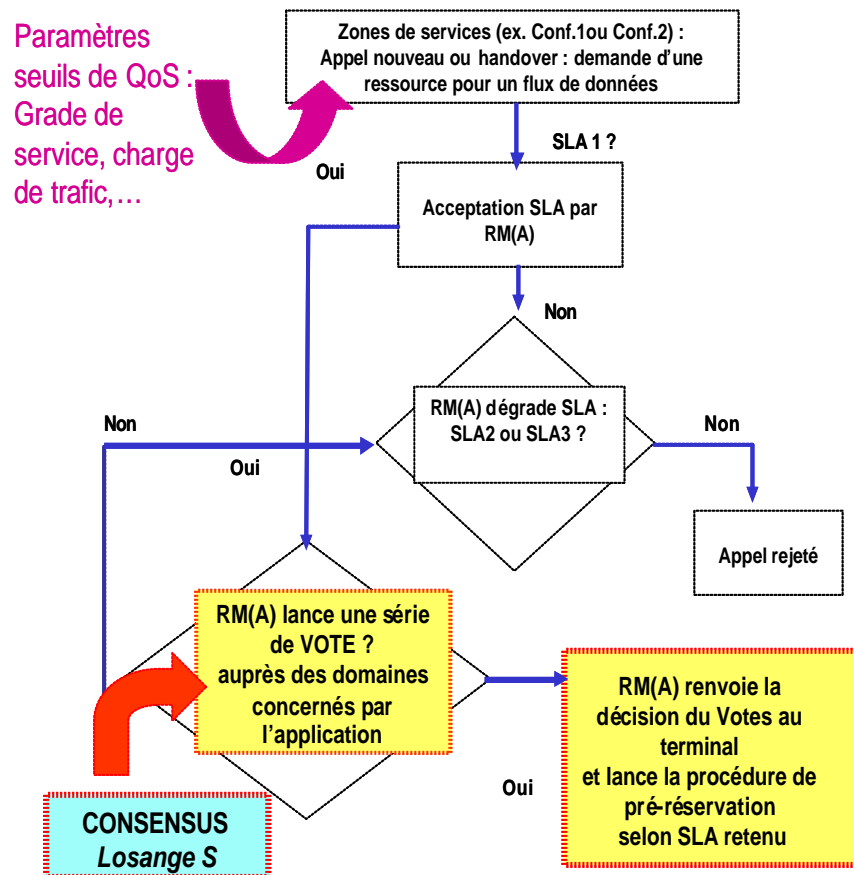


Figure 10.3 Organigramme du contrôle d'admission d'appel de bout en bout

10.2.2 Exemple de CAC de bout en bout

Considérons par exemple le cas d'une demande de session multimédia pour un service vidéo-téléphonie et étudions l'allocation de ressources. La figure 10.4 schématise le contrôle d'admission entre les serveurs SLAs pour la réservation de ressources.

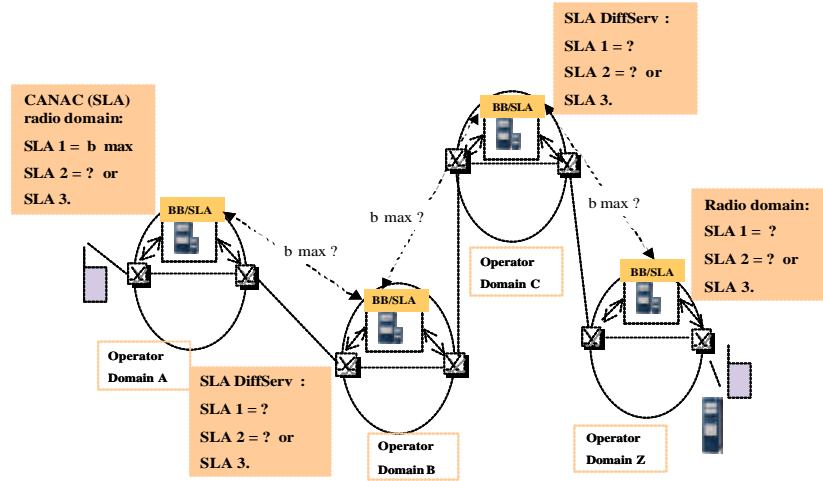


Figure 10.4 Contrôle d'admission d'appel intra et inter-domaines

Le débit b_{max} demandé par le domaine de l'appelant par exemple correspond à SLA1 et à la quantité maximum de trafic à négocier entre les gestionnaires des ressources impliqués dans l'application.

Le déroulement d'une requête suit les étapes suivantes :

- Un terminal mobile A informe son gestionnaire local RM (A) du besoin de communiquer avec B avec un service vidéo par exemple de débit : b_{max} ;
- RM (A) envoie un message COPS (cf. chapitre 5) au domaine d'enregistrement SLA de son domaine radio mobile d'authentification, de vérification de la disponibilité de ressources et de demande d'information sur les seuils d'acceptation (cf. chapitre 8) ;
- Si l'appel est accepté, le serveur RM (A) diffuse une requête COPS aux autres RM des domaines concernés par cette application dans le but de décider collégialement de l'admission de l'appel et de l'allocation de ressources. Ce message inclut le SLA de la demande avec les niveaux de SLA acceptables (SLA1 ou SLA2 ou SLA3) (cf. chapitres 5 et 8) ;
- Les différents RMs contactés vérifient la disponibilité des ressources d'après le contrat SLA demandé. Selon la disponibilité de ressources, les RMs répondent par un message indiquant le niveau de SLA qu'ils peuvent allouer à cette demande en fonction des niveaux de SLA possibles au sein du contrat qu'ils ont reçus (SLA1 ou SLA2 ou SLA3) ;
- L'algorithme d'élaboration d'un consensus permet alors de trouver un accord concernant le niveau de SLA pour lequel les différents intervenants pourront fournir des ressources.

10.3 Mise en œuvre du consensus

10.3.1 Théorie de graphe

Un réseau de transport est un graphe $G = (V, E)$ connexe, comportant un point d'entrée source et un point de sortie destination. V représente l'ensemble des domaines adjacents. E représente l'ensemble des connexions/chemins radio mobiles. L'arc (i, j) (bande passante) représente le chemin le plus court entre un domaine en i et le domaine final en j (voir figure 10.5) [BERR05][BEAI04].

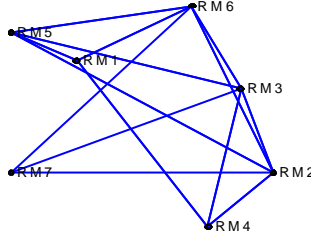


Figure 10.5 Exemple d'un modèle de réseau

Nous modélisons notre réseau selon les notations suivantes :

- Nous disposons de RM domaines adjacents répartis sur la zone de couverture et offrant des niveaux SLA tel que $SLA = \{x1, x2, \dots, xp\}$;
- Les RM étant munies de disponibilités qui représentent les débits maximaux offerts par chacune des sources (débit maximum offert par un domaine) ;
- Chacun des utilisateurs demande un service particulier depuis les P services disponibles et chacun nécessite un maximum de débit pour être satisfait ;
- Bande passante de chaque liaison entre les différents nœuds de notre réseau est assimilée à la capacité des arcs c_{ij} ;
- S_{ij} : la série des votes en fonction du nombre des domaines votants et des paramètres de qualité de service ;
- K : le nombre de domaines votants ;
- s_k : le domaine source ;
- d_k : le domaine destination ;
- $j(i, j)$: le flux de donnée de l'arc (i, j) .

L'algorithme de consensus se réalise en se basant sur une approche mathématique. Cette approche se base sur le modèle du réseau adopté plus haut. La série des votes réalisée par le coordinateur de l'algorithme est exprimée par les expressions suivantes :

$$S_{ij} = \text{Maximise} \sum_{(i,j) \in E} \sum_{k=1}^k x_{ij}^k \quad (10.1)$$

$$\text{Avec} \quad \sum_{j: (s_k, j) \in E} x_{s_k, j} = 1 \quad ; \quad k=1, \dots, K$$

$$\sum_{i: (i, d_k) \in E} x_{i, d_k} = 1 \quad ; \quad k=1, \dots, K$$

La série S_{ij} converge si est seulement si pour tout $c_{ij} > 0$; il existe un entier k tel que pour $k < K$ on ait :

$$S_{ij} = \text{Maximise} \sum_{(i,j) \in E} \sum_{k=1}^k x_{ij}^k \leq c_{ij} \quad (10.2)$$

10.3.2 Exécution de l'algorithme de consensus – négociation dynamique des média

L'algorithme de consensus (cf. chapitre 9) se place au niveau de la couche contrôle/application. Il s'exécute entre les différents gestionnaires de ressources qui participent à l'acheminement des média de bout en bout.

Procedure Consensus *Losange* $S (RM_i, r_{RM_i}, S_{ij}, ts_{RM_i})$

Process RM_i runs the following :

$estimate_{RM_i} \leftarrow S_{ij}$ { denotes RM 's estimate of the decision value }

$state_{RM_i} \leftarrow undecided$

$r_{RM_i} \leftarrow 0$ { r_{RM_i} denotes the current round number }

$ts_{RM_i} \leftarrow 0$ { the round in which $estimate_{RM_i}$ was last updated, initially 0 }

cobegin

|| **Task 1 : { Rotate through coordinator until decision is reached }**

while $state_{RM_i} = undecided$

$r_{RM_i} \leftarrow r_{RM_i} + 1$

$c_{RM_i} \leftarrow (r_{RM_i} \bmod K) + 1$ { K : is the RM 's number (from source until destination) } { c_{RM_i} is the current coordinator }

Phase 1 : { All resources managers K send $estimate_{RM_i}$ to the current coordinator }

RM_i sends ($RM_i, r_{RM_i}, estimate_{RM_i}, ts_{RM_i}$) to c_{RM_i}

Phase 2 : { The current coordinator gather $K - n$ estimates and proposes a new estimate }

if $RM_i = c_{RM_i}$ **then**

wait until [for $K - n$ resource manager RM_j : received ($RM_j, r_{RM_i}, estimate_{RM_j}, ts_{RM_j}$) **from** RM_j]

$msgs_{RM_i}(r_{RM_i}) \leftarrow \{$

$RM_j, r_{RM_i}, estimate_{RM_j}, ts_{RM_j} \}$ | RM_i received ($RM_j, r_{RM_i}, estimate_{RM_j}, ts_{RM_j}$) **from** RM_j }

$t \leftarrow$ largest ts_{RM_j} such that ($RM_j, r_{RM_i}, estimate_{RM_j}, ts_{RM_j}$) $\in msgs_{RM_i}(r_{RM_i})$

$estimate_{RM_i} \leftarrow$ select one $estimate_{RM_j}$ such that ($RM_j, r_{RM_i}, estimate_{RM_j}, t$) $\in msgs_{RM_i}(r_{RM_i})$

RM_i sends ($RM_i, r_{RM_i}, estimate_{RM_i}$) to all resources managers

Phase 3 : { All resources managers wait for the new estimate proposed by the current coordinator }

wait until [received ($c_{RM_i}, r_{RM_i}, estimate_{RM_c}$) **from** c_{RM_i}]

if [received ($c_{RM_i}, r_{RM_i}, estimate_{RM_c}$) **from** c_{RM_i}] **then**

$estimate_{RM_i} \leftarrow estimate_{RM_c}$

$ts_{RM_i} \leftarrow r_{RM_i}$

RM_i sends (RM_i, r_{RM_i}, ack) to c_{RM_i} { RM_i suspects that c_{RM_i} can vote }

else RM_i sends ($RM_i, r_{RM_i}, nack$) to c_{RM_i} { RM_i suspects that c_{RM_i} cannot vote }

Phase 4 : { The current coordinator wait for $K-n$ replies. If these replies indicate that $k-n$ resources managers adopted its estimate, the coordinator sends a request to decide }

if $RM_i = c_{RM_i}$ then

wait until [for $K-n$ resources managers RM_j : received (RM_j, r_{RM_i}, ack) or ($RM_j, r_{RM_i}, nack$)]

if [for $K-n$ resources managers RM_j : received (RM_j, r_{RM_i}, ack) then RM_i recasts (($RM_i, r_{RM_i}, estimate_{RM_i}, decide$))]

|| **Task 2 :** { When RM_i receives a decide message, it decides }

when RM_i redelivers ($RM_j, r_{RM_j}, estimate_{RM_j}, decide$) for some RM_j

decide on $estimate_{RM_j}$

$state_{RM_i} \leftarrow decided$

end.

Figure 10.6 Procédure consensus *Losange S*

10.3.2.1 RM correct - pré-réservation des ressources immédiates

A cours de l'exécution de consensus, le coordinateur tournant décide des niveaux de service SLA ($w\%$) pour les médias de chaque session (i.e. ressources disponibles). Des chemins dynamiques et temporaires sont donc alloués pour chacune (voir figures 10.7 et 10.8).

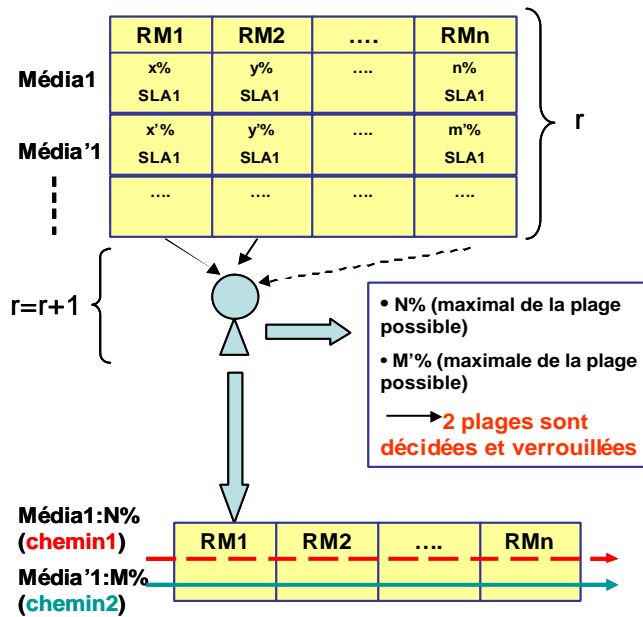


Figure 10.7 Exécution du protocole de consensus : ressources disponibles dans le réseau : réponse immédiate : ex. table de routage 1

10.3.2.2 RM défaillant – négociation des ressources intra et inter-domaines

En plus, au cours de l'exécution du consensus, le coordinateur tournant fixe le nombre et la localisation des gestionnaires des ressources défaillants (ex. RM2, RM3 et RM4) (voir figure 10.8).

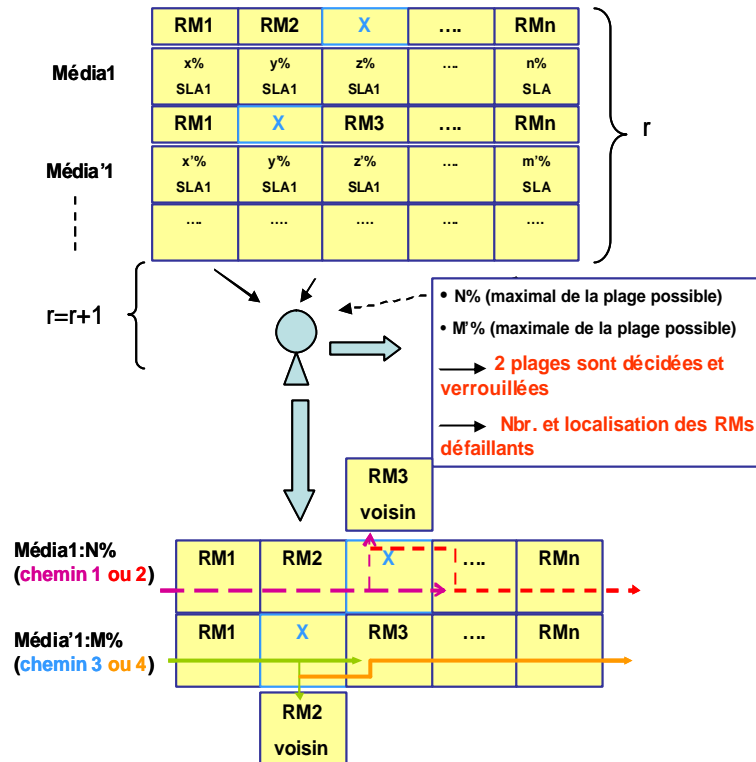


Figure 10.8 Exécution du protocole de consensus : ressources non disponibles auprès de quelques RM : ex. RM2, RM3 et RM4 ; table de routage 2

Les chemins choisis par les coordinateurs sont :

- Les chemins les plus courts ;
- Les chemins les moins chargés (probabilité de blocage des sessions est faible).

10.3.3 Expérimentations et interprétation des résultats

10.3.3.1 Votes successifs – chemins alternatifs – convergence de votes

Application : session vidéo-téléphonie.

K : 10 ; nombre des domaines concernés par l'application (i.e. nombre des votants)

S_{ij} : série de votes successifs (phase 1, phase 2, phase 3 et phase 4)

d : < 150 ms ; délai d'exécution de bout en bout

Au cours d'un contrôle d'admission d'appel de bout en bout et au cours d'une session demandée (voir figure 10.9), l'algorithme de consensus est mis en œuvre à travers des contrats SLA (SLAmin, SLAmoy ou SLAmax) (cf. chapitre 8). Durant une succession des votes réalisés par les votants qui participent à cette application, l'algorithme converge vers une décision pour chaque contrat SLA conclu. La convergence permet une pré-réservation pour ceux qui ont voté l'allocation des ressources et par conséquent d'envisager des domaines alternatifs dans le cas des domaines indisponibles.

Nous remarquons que pour les contrats SLA, le nombre des votants qui ont votés pour indiquer leur disponibilité des ressources pour SLAmin par exemple est

pratiquement égale à 6 domaines sur l'ensemble des domaines qui participent à l'application (10 domaines). Nous pouvons conclure dans ce cas, qu'on peut former alors un chemin avec 6 domaines qui peuvent router par la suite cette application.

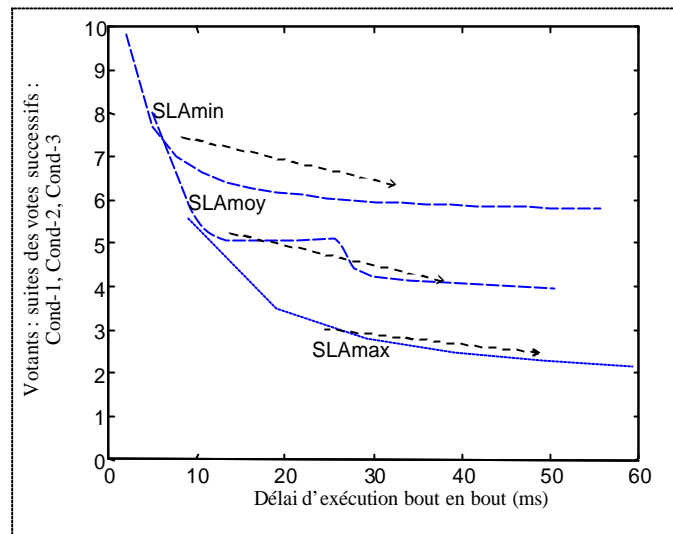
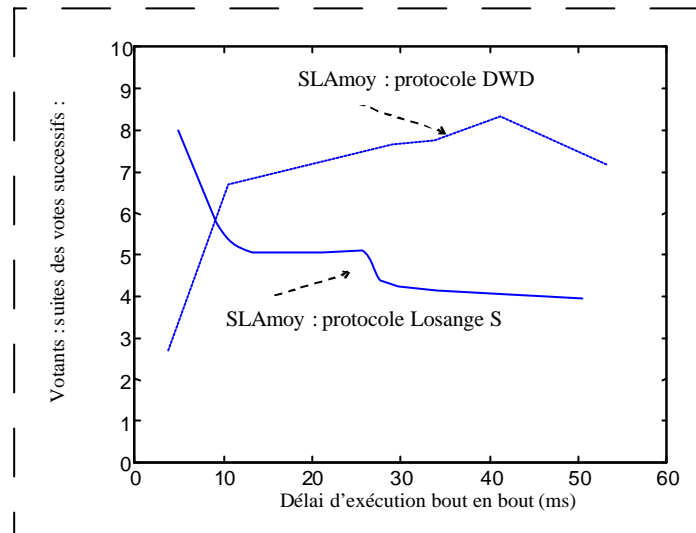


Figure 10.9 Convergence de consensus vers K domaines concernés par l'application - Routage pour SLAmin, SLAmoy, SLAmax

La figure 10.10 compare le consensus *Losange S* à *DWD* (cf. chapitre 9, annexe D) du point de vue convergence de vote. Par exemple pour une décision de vote demandée par l'application, nous remarquons que le *Losange S* converge plus rapidement que le deuxième consensus après une somme des votes successifs et durant un délai d'exécution de bout en bout qui ne dépasse pas le 150 ms. Cette convergence de vote rapide est due à ce que le *Losange S* dans un premier temps tient compte des domaines défaillants (i.e. les domaines qui n'ont pas voté ou bien qui sont en pannes) puisqu'il est alimenté par un détecteur de défaillance. Et dans un deuxième temps il est considéré pour gérer des systèmes télécom asynchrones distribués.

Afin de mieux gérer les systèmes télécom distribués on peut utiliser le consensus *Losange S*. D'après les figures 10.9 et 10.10 le consensus *Losange S* est apparaît plus flexible et plus compatible que le consensus *DWD*.

A cours de l'exécution de consensus *Losange S*, le coordinateur décide des niveaux de service SLA ($w\%$) pour les média de chaque session (ressources disponibles). Des chemins dynamiques et temporaires sont donc alloués pour chacune. En plus, au cours de l'exécution du consensus, le coordinateur fixe le nombre et la localisation des RM défaillants.

Figure 10.10 Comparaison entre le consensus *Losange S* et le consensus *DWD*

L'algorithme de consensus est donc nécessaire pour résoudre le problème de l'indisponibilité des ressources au sein de quelques domaines concernés. Il utilise une suite de votes successifs sur des ensembles disjoints de votants. Il permet de faire des votes sur plusieurs étapes jusqu'à convergence avec réservation pour ceux qui ont voté l'allocation de ressource et par conséquent d'envisager des domaines alternatifs dans le cas des domaines indisponibles.

10.3.3.2 Combinaison de consensus et de la chaîne de Markov : contrôle d'admission et modélisation de système

Le coordinateur du consensus doit alors avoir une vue globale sur l'état de ressource au sein des RM corrects et défaillants (cf. chapitre 8, voir figure 10.11).

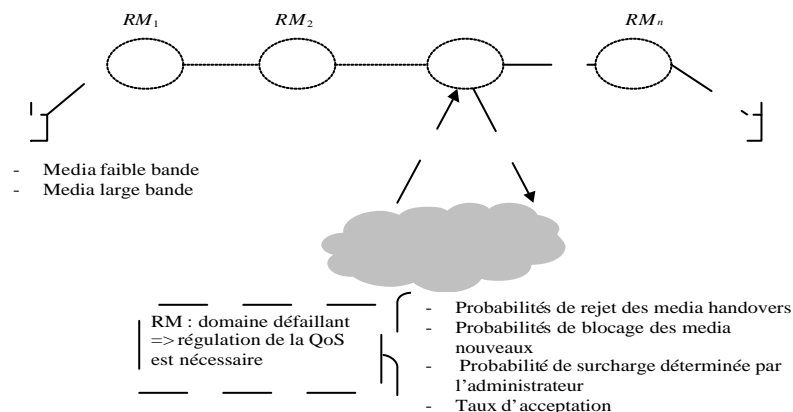


Figure 10.11 Vue globale sur l'état de ressources au sein des RM corrects et défaillant

Donc le choix des chemins alternatifs doit être contrôlé par le module du contrôle d'admission (voir figure 10.12) au sein de chaque coordinateur. Ce dernier, pour qu'il accepte une session nouvelle, il faut qu'il y ait des ressources suffisantes au sein de son domaine et au niveau des domaines voisins.

Pour optimiser les performance de RM sur un niveau SLA alloué ou bien un chemin établi, le contrôle d'admission se base sur des paramètres de contrôle de la qualité de service pour des services hétérogènes [MISIC97][MISIC98][BEAI04]. Le réglage de ces paramètres est nécessaire. La régulation est choisie pour deux types de services : à faible bande et à large bande.

La régulation des paramètres de la qualité de service se fait au niveau de chaque RM. Elle se base sur le calcul de la probabilité de surcharge P_{ov} , là où la somme des canaux occupés par les sessions actives et les canaux réservés pour celles de handovers dépassent la capacité limite du domaine.

Quand un événement de surcharge se produit, la session nouvelle commence à emprunter la bande passante réservée pour les futurs handovers. Ainsi la probabilité de surcharge peut être considérée comme la probabilité où les canaux réservés pour les handovers sont empruntés par les sessions nouvelles.

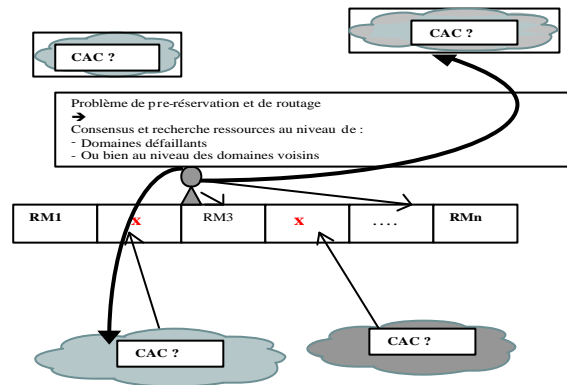


Figure 10.12 Consensus - ressources à récupérer – recherche des chemins alternatifs

Le système de la figure 10.13 peut être modélisé par une chaîne de Markov à deux dimensions. Cette dernière modélise les deux types de service considérés. Là où le service à faible bande est caractérisé par les paramètres I_1, h_1, m_1 , le service à large bande est caractérisé par I_2, h_2, m_2, B_2 . On suppose que la durée de la session m_j ainsi que *dwell time* h_j pour chaque type de service j suit une distribution exponentielle et le taux d'arrivée des sessions I_j suit un processus de Poisson.

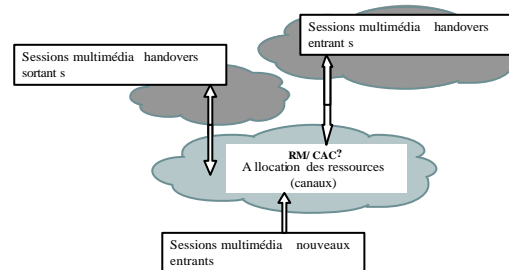


Figure 10.13 CAC - régulation des paramètres de la QoS

Une chaîne de Markov qui représente notre système par exemple pour un seul type de service est illustrée dans la figure 10.14 [MISIC97][MISIC98][BEAI04]. Les états

dans la chaîne représentent le nombre des sessions continues dans le domaine ainsi que le nombre de sessions arrivants des domaines voisins.

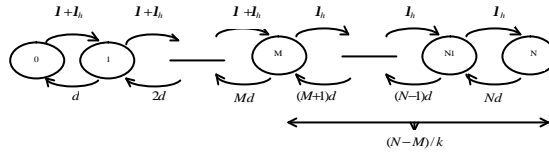


Figure 10.14 Chaîne de Markov modélise un seul type de service (avec $d=h+m$)

La valeur $M = r_{\max} N$ représente le nombre maximum des canaux permis pour tout taux d'arrivée I sous une certaine politique de contrôle d'admission pour un seuil P_{ov} donné. Notons que M est également l'état où le système fonctionne sous un seuil P_{ov} déterminé par l'administrateur de réseau. Au delà de l'état M , seulement les sessions de transfert (handovers) sont acceptées.

Les probabilités de blocage et de rejet pour ce système sont donc données par [MISIC97][MISIC98][BEAI04] :

$$P_B(I, m, h, M, N) = \frac{g^M \sum_{i=M}^N \frac{1}{i!} b^{i-M}}{\sum_{i=0}^M \frac{g_i}{i!} + g^M \sum_{i=M+1}^N \frac{1}{i!} b^{i-M}} \quad (10.3)$$

$$P_{hd}(I, m, h, M, N) = \frac{\frac{1}{N!} g^M B^{N-M}}{\sum_{i=0}^M \frac{g_i}{i!} + g^M \sum_{i=M+1}^N \frac{1}{i!} b^{i-M}} \quad (10.4)$$

avec : $g = \frac{I + I_h}{m + h}$, $b = \frac{I_h}{m + h}$, N est le nombre des canaux dans le domaine, et I_h est le taux d'arrivée des sessions handovers avant que l'algorithme d'admission commence à cesser d'admettre les sessions nouvelles.

i. Performance système basé sur la probabilité de surcharge P_{ov} - deux types de services analysés (faible bande et large bande)

Pour calculer rapidement les limites approximatives de la qualité de service et la valeur relative de P_{ov} , nous employons la chaîne de Markov bidimensionnelle basé sur l'analyse approximative (voir figures 10.15 et 10.16).

Les taux d'admission globaux (i.e. les services admis) et l'utilisation correspondante de domaine par les services à faible bande et large bande sont donnés par l'ensemble des équations suivantes :

Le taux d'arrivée global Λ admis dans le domaine ainsi que la capacité d'utilisation du domaine en terme de canaux sont exprimés par les équations suivantes [MISIC97][MISIC98] [BEAI04] :

$$\Lambda_1 = I_1 (1 - P_{B1}) + \Lambda \frac{h_1}{h_1 + m_1} (1 - P_{hd1}) \quad (10.5)$$

$$\Lambda_2 = I_2(1 - P_{B2}) + \Lambda \frac{h_2}{h_2 + m_2}(1 - P_{hd2}) \quad (10.6)$$

$$r_1 = \frac{\Lambda_1}{N(h_1 + m_1)} \quad \text{et} \quad r_2 = \frac{\Lambda_2}{N(h_2 + m_2)}$$

Donc, les taux de réservation prédits pour les handovers sont donnés par :

$$n_1 = I_1(1 - P_{B1}) \frac{B_1}{m_1 + h_1 P_{hd1}} \quad (10.7)$$

$$n_2 = I_2(1 - P_{B2}) \frac{B_2}{m_2 + h_2 P_{hd2}} \quad (10.8)$$

Et le nombre des canaux utilisés :

$$r_{tot} N = (r_1 + r_2) N = \frac{n_1}{B_1} + \frac{k n_2}{B_2} \quad (10.9)$$

En considérant le nombre de canaux réservés (équations (10.7) et (10.8)), la probabilité de surcharge peut être exprimée par l'équation suivante :

$$P_{ov} = e^{-(n_1 + n_2)} \sum_{q=[N(1-r_{tot})]_{q_2=0}}^{\infty} \sum_{k=0}^{[q_2]} \frac{n_2^{q_2}}{q_2!} \frac{n_1^{(q-k)q_2}}{(q-kq_2)!} \quad (10.10)$$

Le but de l'analyse de performance précédente est de décrire les propriétés de la régulation des paramètres de la probabilité de surcharge P_{ov} . Si P_{ov} est égale à 0, une bande passante est réservée pour satisfaire les faibles probabilités de rejet de handovers. Si P_{ov} varie entre $\{0,1\}$ la bande passante réservée diminue afin de satisfaire la probabilité de rejet et conserver un niveau maximum d'utilisation dans le domaine.

ii. Probabilité de blocage, probabilité de rejet et utilisation

On définit $k = \frac{I_1}{I_2}$ le rapport de taux d'arrivée de deux types de service. On suppose que dans un système, la valeur seuil choisie P_{ov} , limite le nombre maximal des canaux utilisés dans le domaine à $M = r_{\max} N$, pour n'importe quel taux d'arrivée I et par exemple pour un seul type de service. La valeur M peut être trouvée pour une valeur particulière de P_{ov} et par le calcul de r_{\max} en fonction de n_{\max} . De plus, on suppose que dans un système avec deux types de service et avec k constant, le nombre maximum des canaux utilisés sont respectivement $M_1 = r_{1\max} N$ et $M_2 = r_{2\max} N$. On suppose ainsi que $m_1 = r_1 N$ et $m_2 = r_2 N$ sont les nombres moyens des canaux occupés par les deux types de service, à faible bande et large bande. Ils sont déterminés respectivement pour des valeurs particulières des taux d'arrivée, I_1 et I_2 .

L'analyse complète de performance du système peut être réalisée pour ces deux types de service en utilisant la chaîne de Markov à deux dimensions (domaine avec à peu près 25 canaux [MISIC97][BEM03] et $k = 5$).

Pour plus de deux types de service, l'analyse correspondante de Markov deviendra très pénible et longue. Par conséquent, nous employons une analyse approximative qui découple le système en différents sous-systèmes où les différents types de service sont analysés séparément.

En particulier, les deux types de service à faible et à large bande peuvent être analysés séparément en utilisant les chaînes de Markov représentées dans les figures 10.15 et 10.16. Les états dans les chaînes représentent le nombre des sessions dans les deux sous-systèmes. Ces derniers modélisent les deux types de services considérés [MISIC97][MISIC98] [BEAI04] :

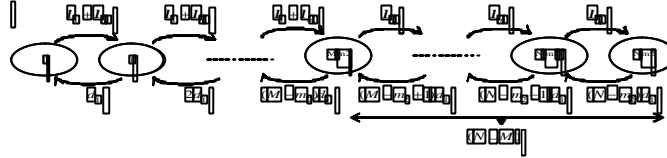


Figure 10.15 Chaîne de Markov modélise le service type 1
(ex. faible bande : *conversational* et *interactive*)

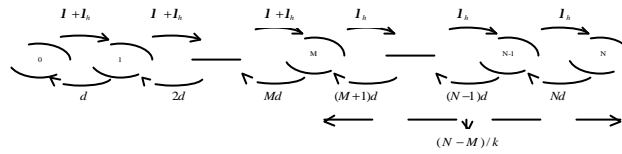


Figure 10.16 Chaîne de Markov modélise le service type 2
(ex. large bande : *background* et *streaming*)

Notons que les deux états dans les deux chaînes dans lesquelles la régulation de P_{ov} commencera, sont la probabilité de blocage et l'utilisation, notés respectivement, $M - m_2$ et $\frac{M - m_1}{k}$.

Les probabilités de blocage pour les deux types de service (ou bien les deux chaînes) sont calculées en se basant sur les simplifications et les hypothèses considérées dans les sections précédentes. Elles sont données par les équations suivantes [MISIC97][MISIC98][BEAI04] :

$$P_{B1} = 1 - (1 - P_{B1,h}) \prod_{i=1}^S (1 - P_{B1,i} P_{B2,i})^{6i} \quad (10.11)$$

$$P_{B2} = 1 - (1 - P_{B2,h}) \prod_{i=1}^S (1 - P_{B2,i} P_{B1,i})^{6i} \quad (10.12)$$

où $P_{B1,h} = E(\frac{I_1}{m_1}, M - m_2)$ et $P_{B2,h} = E(\frac{I_1}{k(m_2 + h_2 P_{hd2})}, \frac{M - m_1}{k})$.

Les probabilités où une session nouvelle (à faible bande ou large bande) peut être bloquée dans le domaine. Pareillement $P_{B1,i} = E(\frac{m_1}{k}, M - m_2)$ et $P_{B2,i} = E(\frac{m_2}{k}, \frac{M - m_1}{k})$ les probabilités où les sessions nouvelles seront bloquées à cause de la non disponibilité des domaines voisins pour satisfaire les deux types de service.

Puisque le système est prévu pour satisfaire les handovers, la valeur de P_{hd1} sera très petite et peut être négligée pour le calcul de m_1 et m_2 . On suppose que les intensités de trafic sont égales dans tous les domaines. Les relations entre le nombre moyen des

canaux utilisés pour les deux types de services, m_1 et m_2 sont données par [MISIC97][MISIC98][BEAI04] :

$$m_1 = \frac{I_1}{m_1} (1 - P_{B1,h}) \prod_{i=1}^S (1 - P_{B1,i} P_{B2,i})^{6i} \quad (10.13)$$

$$m_2 = \frac{I_1 k}{k(m_2 + h_2 P_{hd2})} (1 - P_{B2,h}) \prod_{i=1}^S (1 - P_{B2,i} P_{B1,i})^{6i} \quad (10.14)$$

La simulation est réalisée en utilisant un algorithme pour développer un consensus concernant le niveau de service en conjonction avec le modèle Markovien avec deux dimensions (deux types de services, faible bande et large bande). Les deux figures qui suivent (voir figures 10.17 et 10.18) représentent le blocage des appels et le rejet de handovers. Les résultats montrent une tendance générale observée pendant les expériences. Nous pouvons conclure que notre approche réduit les probabilités de blocage et de rejet. Quand un consensus est établi, les niveaux de service peuvent décroître du SLA1 au SLA2 ou SLA3 pour limiter ces rejets.

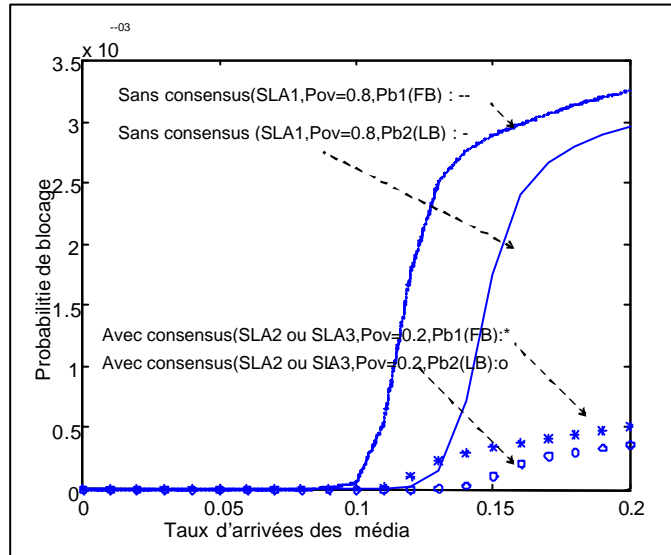


Figure 10.17 Probabilité de blocage en fonction de taux d'arrivée d'appels

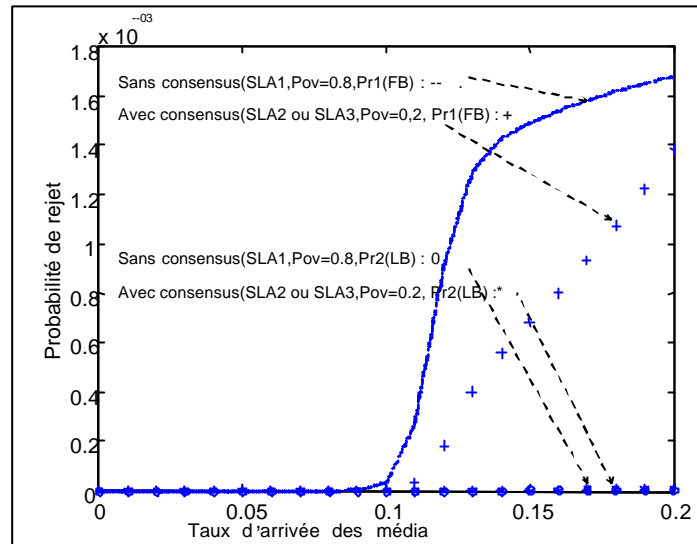


Figure 10.18 Probabilité de rejet en fonction de taux d'arrivée d'appels

En comparaison (voir figure 10.19), le taux d'utilisation reste approximativement le même. Pour les sessions multimédia, il est nécessaire d'éviter le rejet d'un nouveau média dans la même session.

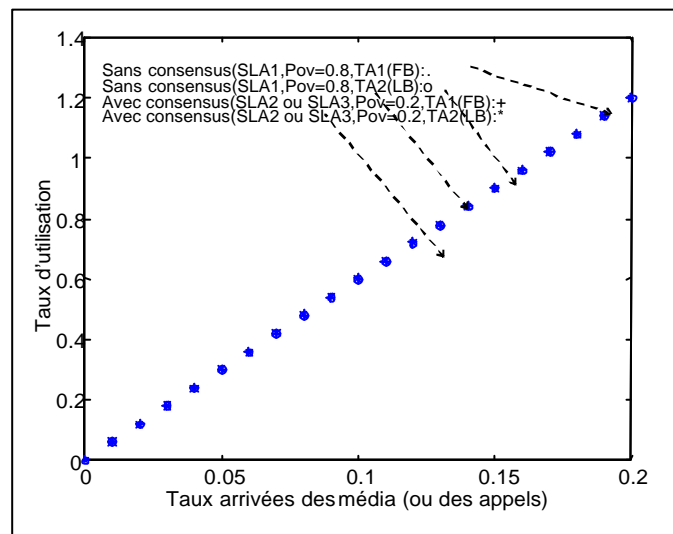


Figure 10.19 Taux d'utilisation en fonction de taux d'arrivée d'appel

iii. Markov – pilotage d'accessibilité aux média

Au cours d'un consensus et lors d'une session d'appel, un utilisateur mobile peut établir et maintenir plusieurs connexions simultanément. Il est également possible de modifier ces caractéristiques durant une connexion grâce à une procédure de re-négociation. Donc au cours de la recherche des ressources (i.e. négociation selon SLAmax, SLAmoy et SLAmin par exemple), le système est ainsi alimenté par un deuxième outil : la chaîne de Markov à deux dimensions (cf. figures 10.15 et 10.16), utilisée pour s'informer du point de vue surcharge, blocage et rejet sur l'état des domaines que la session pourrait parcourir. Les états dans la chaîne représentent le

nombre des sessions continues dans le domaine ainsi que le nombre de sessions arrivants des domaines voisins.

Dans la figure 10.20, nous remarquons que la probabilité de blocage augmente avec l'augmentation des taux d'arrivée des média nouveaux. En se basant simultanément sur le modèle de consensus, trois contrats négociés dans l'algorithme et dans les transitions des sessions (i.e. les états de la chaîne de Markov), les gestionnaires des ressources considérés continuent l'acceptation (i.e. la réservation des ressources) de différents types de média provenant soit d'une même session, soit de différentes sessions jusqu'à certains seuils (par exemple la probabilité des média large bande est égale à $1.2 \cdot 10^{-3}$ et la probabilité des média faible bande est égale à $0.2 \cdot 10^{-3}$) au delà desquels les média seront refusés. Les gestionnaires des ressources sont dans ce cas considérés comme surchargés (la probabilité de surcharge est maximale est égale à $0.8 \approx 1$). Ainsi, nous remarquons que l'accessibilité à de nouveaux média faible bande (probabilité de blocage faible varie entre $0.01 \cdot 10^{-3}$ et $0.09 \cdot 10^{-3}$) est beaucoup importante que celle des média large bande (probabilité de blocage importante varie entre $0.2 \cdot 10^{-3}$ et $1.2 \cdot 10^{-3}$) dans le cas où les gestionnaires des ressources sont considérés surchargés ou non.

Dans la figure 10.21, et en se basant sur les mêmes paramètres de simulation, le rejet des médias faible bande est pratiquement nul (accessibilité des nouveaux média faible bande est importante (voir figure 10.20)) par rapport à celui des média large bande (probabilité de rejet = $3 \cdot 10^{-3}$ pour une probabilité de surcharge = 0.8).

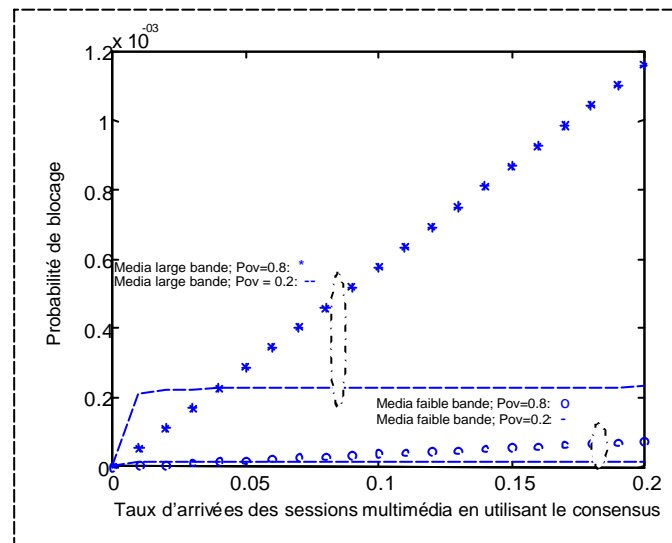


Figure 10.20 Probabilité de blocage en fonction de taux d'arrivée des sessions multimédia en utilisant le consensus

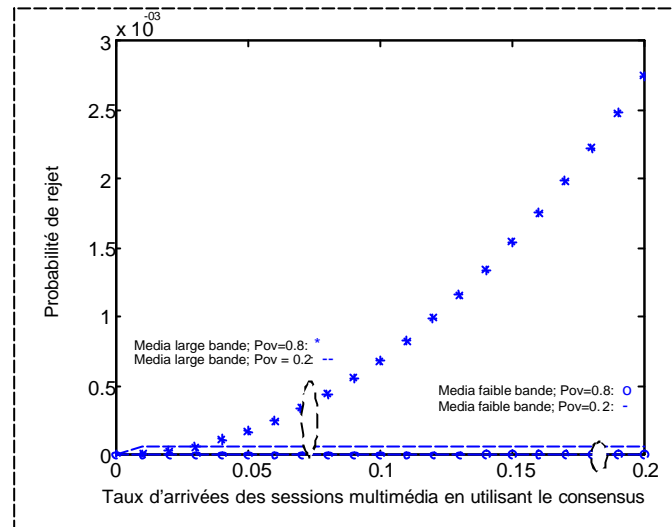


Figure 10.21 Probabilité de rejet en fonction de taux d'arrivée des sessions multimédia en utilisant le consensus

Le tableau 10.1 permet donc de piloter les choix de l'utilisateur en lui présentant en code couleur les autres media qu'il pourrait sélectionner dans le cadre de la session. Les probabilités de rejet, de blocage et de surcharge pourrait alors être utilisées combinées ensemble afin de générer ce code : soit il évite de sélectionner ce service, soit il paye le prix fort, soit l'opérateur le lui refuse.

L'intérêt des chaînes de Markov est donc de donner la probabilité de passer à un nouvel état à partir de l'état actuel (ex : demander l'image si on dispose déjà de la voix).

Type de média	Probabilité de surcharge minimale considérée $P_{ov} = 0.2$	Probabilité de surcharge maximale considérée $P_{ov} = 0.8$
Média faibles bande	Probabilité de blocage minimale : $P_{B \min} = 0.0110^{-3}$ Probabilité de rejet minimale : $P_{hd \min} = 0.00110^{-3}$	Probabilité de blocage Maximale : $P_{B \max} = 0.0910^{-3}$ Probabilité de rejet maximale : $P_{hd \max} = 0.00210^{-3}$
Média larges bande	Probabilité de blocage minimale : $P_{B \min} = 0.210^{-3}$ Probabilité de rejet minimale : $P_{hd \min} = 0.1510^{-3}$	Probabilité de blocage maximale : $P_{B \max} = 1.210^{-3}$ Probabilité de rejet maximale : $P_{hd \max} = 2.710^{-3}$

Tableau 10.1 Markov - pilotage d'accessibilité aux média

10.4 Conclusion

Notre travail permet de mettre en place l'infrastructure de la mise en œuvre d'un modèle de pré-réservation des ressources d'une manière consensuelle en tenant compte

des contraintes de chaque domaine, et des profils des utilisateurs. Ce modèle se base sur un mécanisme du contrôle d'admission d'appel de bout en bout pendant une session multimédia demandée.

Dans cette approche, le mécanisme d'admission d'appel se fait en deux étapes : Premièrement au niveau de la Capacité d'Acceptation de Nouveaux Appels par Cellule (CANAC) enregistrée au niveau du SLA radio mobile, là où la mobilité d'utilisateur est considérée comme une contrainte majeure pour l'acceptation d'un appel. La seconde étape a lieu au niveau du réseau DiffServ de l'opérateur. L'acceptation de la qualité de service est définie par le contrat SLA qui spécifie les paramètres et les différents niveaux d'acceptation. Au niveau global, l'acceptation d'un appel handover ou d'un nouvel appel se base sur une décision collégiale et consensuelle entre les différents réseaux participants. Un tel consensus permet une prise de décision cohérente pour réserver une ressource suivant un des niveaux de service prédéfini par le SLA du demandeur puis par la suite de permettre à l'appelant de renégocier le niveau de qualité de service du service demandé en fonction de la fluctuation de la charge de trafic ou négocier de nouvelles ressources pour de nouvelles applications demandées au cours de la session. Les autres réseaux participants peuvent également initier une renégociation des ressources allouées à cette session pour les mêmes raisons.

Conclusion générale

Bilan de travail

L'augmentation du nombre d'utilisateurs et l'évolution de leurs besoins vers des services multimédia couplés avec des exigences de qualité de service posent des défis aux opérateurs de réseaux radio mobiles dont les ressources sont limitées. Une approche judicieuse d'allocation des ressources a été développée alors dans cette thèse. Cette approche permet de limiter les rejets d'appels lors de handovers et améliorer la gestion la gestion des ressources radio en tenant compte des facteurs liés aux attentes en terme de qualité de service.

Le travail présent une approche de gestion de l'allocation des ressources basée principalement sur des paramètres de qualité de service et un mécanisme de contrôle d'admission d'appel.

Cette approche a pour objectif d'augmenter le niveau d'acceptation d'appels au sein du domaine tout en garantissant la qualité de service requise. Le mécanisme qui sous-tend notre approche, permet d'adapter le comportement du trafic aux caractéristiques du réseau tout en restant conformes aux engagements de qualité de service déclinés selon les paramètres tels que le taux de handovers, la charge de trafic, la probabilité de rejet d'appel, le niveau de service (GoS), etc. Cette approche est développée spécifiquement pour les réseaux cellulaires radio mobiles pour lesquels la mobilité des usagers est considérée comme étant une contrainte majeure pour l'acceptation d'un appel.

Au cours de cette approche nous avons présenté le réseau radio UMTS, ses spécificités, ses caractéristiques et son infrastructure. Ceci est important pour la mise en œuvre de notre modèle de mobilité qui a été développé au niveau du chapitre 7. Ce modèle de mobilité nous a permis de déterminer au niveau d'une cellule aussi bien qu'au niveau du système entier, les paramètres seuils, au-delà desquels un appel peut être refusé ou non. Ces paramètres ont été utilisés par la suite pour paramétrer notre modèle d'allocation de ressources de bout en bout qui a été développé principalement au niveau des chapitres 8, 9 et 10.

Stratégie de travail

- La collecte des données (cartes géographiques, compteurs OMC, etc.). Des indicateurs clés ajustés et optimisés sont pris de deux configurations urbaines (Rapports techniques : Tunisie Télécoms 2001/2002) de deux réseaux GSM existants, placées l'une à Tunis Hached avec un rayon global de 1,7 km et couverte approximativement par 50 cellules et l'autre au centre ville de Sfax avec un rayon de 2 km, couverte avec 30 cellules. Dans chacune des deux configurations, nous considérons une BTS/cellule, des grands axes routiers, des avenues, des petites rues, etc. Ainsi le trafic total est considéré important surtout durant les heures de pointe.
 - Traitement et analyse des données, puis l'interprétation des résultats.
 - Utilisation des langages de programmation et d'exploitation : C, Matlab et Excel.
- Un modèle de mobilité considéré dans le dimensionnement de réseau et dans quatre modèles de trafic étudiés (voix, vidéo-téléphonie, www, ftp), se base principalement sur l'étude des flux d'utilisateurs dans les rues. Il est calibré par des données récupérées à partir des sites de configurations réelles (Rapports techniques : Tunisie Télécoms 2001/2002) : des stations de base GSM existantes afin de bénéficier de dimensionnement et de configurations réels des sites et des liens de transmissions pour ainsi les extrapoler dans le cadre des réseaux 3G.

Ce modèle de mobilité nous a permis de déterminer au niveau d'une cellule, mais également au niveau du système entier les paramètres seuils au delà desquels un appel peut être refusé. Il nous a ainsi permis de connaître les limites de réservation d'une ressource : le taux de handovers, la probabilité de rejet d'appels, le trafic offert, le grade de service, etc.

- L'élaboration d'un mécanisme de gestion de la qualité de service proposé lors d'allocation des ressources afin d'adapter et d'assurer un service multimédia mobile de qualité tout en optimisant l'usage des ressources de l'opérateur. Ce mécanisme est basé essentiellement, sur l'état du réseau où il est implanté et plus particulièrement sur la charge, sur la probabilité de rejet d'appel handovers et sur le grade de service comme critères pour une garantie de qualité de service. Le mécanisme est implanté au niveau du gestionnaire des ressources du domaine. Il tient compte des informations contractuelles au sein du SLA et est paramétré avec les données seuils afin de gérer les niveaux de qualité de service possibles (haut, moyen, bas). Son objectif est ainsi de maximiser le taux d'acceptation des appels et

l'utilisation des ressources tout en contrôlant la charge du réseau de telle sorte que les différents niveaux de garantie de services demandés soient satisfaits.

- L'élaboration d'un modèle de contrôle d'admission d'appel multimédia proposé qui permet de mettre en place un mécanisme de négociation et de renégociation des ressources selon les contrats SLA et leurs niveaux de qualité de service. Ce modèle se base sur des paramètres de qualité de service seuils issus de mesures sur le terrain et de simulation qui permettent de tenir compte du dimensionnement des domaines et des modèles de trafic et de mobilité. Le modèle adopte un algorithme de décision afin d'accorder des ressources selon une approche consensuelle. Cette approche permet d'éviter les réservations incohérentes. La même approche est utilisée pour rendre la réservation des ressources plus flexible en libérant en cas de besoin par la dégradation de la qualité de service au niveau SLA inférieur.
- La mise en place d'une politique de pré-réservation des ressources bout en bout des services télécoms pour l'UMTS.

Un service est la combinaison d'une politique de conditionnement du trafic et d'un comportement des nœuds qui l'achemine. La politique proposée suppose que chaque nœud a le contrôle des ressources. Dans le cas des liens point à point, cette hypothèse est vraie. Par contre, dans le cas des liens multipoints le contrôle de la bande passante est partagé et distribué. Dans ces conditions, le trafic d'un nœud est perturbé par les conditions de trafic lien.

Dans une architecture DiffServ, les éléments de bordure effectuent la classification et le conditionnement du trafic en fonction des accords de service SLA et des spécifications du niveau de services SLS entre le domaine du client et ses voisins. Un SLA est un accord bilatéral négocié du type client - fournisseur. Le SLS contient les paramètres tels que la capacité de transmission requise, le débit, le temps de transfert, etc. L'action du conditionnement porte sur l'admission des flux d'utilisateurs en garantissant une qualité de service de bout en bout. Les éléments du cœur du réseau se chargent du transit des flux d'utilisateurs.

Le contrôle de SLA introduit un régulateur au niveau de chaque nœud afin de contrôler les demandes d'accès au média alloué. La régulation est effectuée en fonction de l'état du nœud et de l'état du réseau. Le régulateur distribue des votes aux nœuds. Ces votes représentent le SLA de chaque classe de trafic pouvant être émise. L'activité des sources et le trafic généré sont dynamiques ; la solution du contrôle de la capacité est donc dynamique. La localisation du contrôle est distribuée. Les méthodes d'accès sur les média alloués ont un contrôle distribué. Ainsi, la complexité est partagée par l'ensemble des nœuds et le système devient plus robuste. Pour la mettre en œuvre, la coordination peut se faire via le protocole COPS.

La coordination dynamique des demandes réparties est une tâche délicate qui doit tenir compte de deux objectifs antinomiques : actualité des informations d'état du trafic servant à l'allocation des ressources et faible sur débit induit par la diffusion de ces informations. Les propriétés attendues d'un mécanisme de contrôle sont la réactivité au changement des conditions de trafic, la précision dans les allocations

ainsi que la robustesse face aux pannes, aux arrêts et démarrages des nœuds et aux pertes de messages de coordination.

Pour contrôler le SLA, un protocole de consensus est mis en œuvre. Chaque nœud vote sur la pré-réservation selon la connaissance qu'il a de l'utilisation des ressources. Le trafic de coordination dépend également du nombre de nœuds et de la variabilité du trafic.

En effet, si un nœud se connecte au réseau de manière sauvage, sans être informé par le coordinateur, il accapare de la bande passante et fausse les calculs effectués du SLA disponible. Pour atteindre l'objectif du support de qualité de service, tous les nœuds doivent donc collaborer.

Prolongement et perspectives

A partir de notre travail, plusieurs perspectives de recherches et prolongements des travaux peuvent être considérés.

Une première perspective consiste à intégrer la mobilité des utilisateurs dans les différentes approches d'adaptation de services en UMTS-NGN. En effet, dans ces réseaux radio mobiles, les ressources physiques subissent des fluctuations fréquentes et importantes. Ces fluctuations sont produites par la mobilité de l'utilisateur des variations des paramètres de qualité de service sur les réseaux transport. Il nous paraît donc intéressant de considérer le volet rattaché à la mobilité de l'utilisateur, pour faire évoluer les approches d'adaptation de services télécoms multimédia.

L'introduction des notions de contrôle d'admission d'appel (CAC) dans nos travaux peut aussi améliorer les approches d'adaptation de services. Elles consistent à une procédure de contrôle avec une dégradation de service. Cette procédure permet de mettre en place un mécanisme de négociation et de renégociation des ressources selon les contrats SLA et leurs niveaux de qualité de service.

Dans le but d'assurer la gestion de la qualité de service dans les réseaux ad hoc, une architecture de gestion par les politiques et un modèle d'approvisionnement de qualité de service de type DiffServ ont été proposés dans nos travaux. En effet, dans un réseau *ad hoc* [WUI01], le modèle fonctionnel de gestion par politique ne peut être appliqué directement car on ne dispose par définition d'aucun moyen de gestion centralisée. Si une station de gestion est définie et qu'un terminal, qui était préalablement accessible par celle-ci, peut ne pas être à sa portée durant le temps et devient par conséquent incontrôlable.

Nos travaux préparent alors la mise en œuvre de l'infrastructure qui permet la mise en place de politiques de réservation de ressources de manière consensuelle en tenant compte des contraintes de chaque domaine et des profils et préférences des utilisateurs. Nous offrirons alors la possibilité d'appliquer les politiques et les stratégies d'allocation des ressources des opérateurs, leur permettant ainsi d'optimiser l'usage de leurs réseaux tout en satisfaisant les utilisateurs.

L'étude et la mise en œuvre de l'algorithme de routage pour assurer la connexion des réseaux *ad hoc*, sont un problème qui se pose. L'environnement est dynamique et évolue donc au cours du temps, la topologie du réseau peut changer fréquemment. Il nous semble donc important d'utiliser notre approche consensuelle. Elle est basée sur un algorithme asynchrone distribué temps réel (*Losange S*) qui se repose sur les votes pour des contrat SLA conclus. Elle pourra être utilisé pour résoudre les problèmes rencontrés dans la conception de protocole de routage tels que :

- La minimisation de la charge du réseau : l'optimisation des ressources du réseau renferme deux autres problèmes qui sont l'évitement des boucles de routage, et l'empêchement de la concentration du trafic autour de certain nœuds ou lien.
- Assurer un routage optimal : la stratégie de routage doit créer des chemins optimaux et pouvoir prendre en compte différentes métriques de coût (bande passante, nombre des liens, ressources du réseau, délai de bout en bout, etc.)
- Le temps de latence : la qualité des temps de latence et de chemins doivent augmenter dans le cas où la connectivité du réseau augmente.

Annexes

Annexe A

Liste des publications

- Sonia Ben Rejeb, Sami Tabbane, Zièd Choukair, «*Mobility model for multimedia mobile radio network*», article présenté à la conférence **MMNS'03** (parrainée par IEEE et IFIP), Queen University, Belfast North Irland, 11 Décembre 2003.
- Sonia Ben Rejeb, Sami Tabbane, Zièd Choukair, «*Mobility Model used for QoS Management for Multimedia Mobile Radio Network* », article présenté à la conférence **JS'03**, Ecole Borj El Amri, Tunis -Tunisie, 22 Mai 2003.
- Sonia Ben Rejeb, Zièd Choukair, Sami Tabbane, «*Resource allocation and QoS management for cellular multimedia network*», article présenté à la conférence **GRES'03** Fortaleza, Brésil, 22 Février 2003.
- Sonia Ben Rejeb, Zièd Choukair, Sami Tabbane, «*Modeling end-to-end QoS management and resource reservation for multimedia mobile radio network*», article présenté à la conférence **AICCSA'03** (parrainée par IEEE), Tunis - Tunisie, 18 Juillet 2003.
- Sonia Ben Rejeb, Zièd Choukair, Sami Tabbane, «*Combining Markovian Chains and Real Time Agreement protocols to Allocate Resources in Wireless Systems*», article présenté à la conférence **AINA'04** (parrainée par IEEE), Japan, 23 Mars 2004.

- Sonia Ben Rejeb, Zièd Choukair, Sami Tabbane «*Real Time Agreement Protocol and Resource Pre-reservation for Wireless Systems* », article présenté à la conférence **RRM-WCN/IPCCC'05** (First IEEE International Workshop on Radio Resource Management for Wireless Cellular Networks/24th IEEE International Performance Computing and Communications Conference), Phonix, Arizona, USA, 09 Avril 2005.
- Sonia Ben Rejeb, Zièd Choukair, Sami Tabbane «*End to end QoS management and Real Time Agreement Protocol for Resource Reservation for Multimedia Mobile Radio Network*», article présenté à la conférence **ICECS'05** (12th IEEE International Conference on Electronics, Circuits and Systems (ICECS2005), Gammarth, Tunisie, 11 Decembre 2005.
- Sonia ben rejeb, Zièd Choukair, Sami Tabbane «*QoS management and Real Time Agreement Protocol for Resource Reservation for Multimedia Mobile Radio Network*», article accepté et qui sera présenté à la conférence **ACCSA'06** (12th IEEE International Conference on Computer Systems and Application) Dobai/Sharjah, UAE, Mars 2006
- Sonia Ben Rejeb, Sami Tabbane, Zièd Choukair «*Mobility model used for QoS Management for Wireless Multimedia Network*», article accepté et qui sera publié dans le journal **International Journal of Wireless and Mobile Computing (IJWMC) - Special Issue on "Towards Ubiquitous Wireless Communication : The Integration of 3G/WLAN Networks"**, Juin 2005.
- Sonia ben rejeb, Zièd Choukair, Sami Tabbane, Nidal Nasser «*End-to-end QoS for Resource Reservation for Multimedia Mobile Radio Network*», article accepté et qui sera publié dans le **journal GESTS (International Transactions Journal on Computer Science and Engineering** , Janvier 2006

Annexe B

Glossaire (1)

- L : périmètre de la zone de service.
- l_j : périmètre de la sous-zone de service.
- Cont.1 : contour de la zone de service est représenté par les coordonnées $M_i(x_i, y_i, z_i)$ avec $0 < i < L$.
- Cont.2 : le contour de la sous-zone est représenté par les coordonnées $N_i(x'_i, y'_i, z'_i)$ avec $0 < i < l_j$.
- S : surface de la zone de service.
- C : capacité de la zone de service.
- l_j : périmètre de la sous-zone de service.
- S_j : surface de la sous-zone de service.
- C_j : capacité de la sous-zone de service.
- l_{tk} : longueur de la rue n_k .
- v_{tk} : vitesse des usagers dans la rue n_k .
- f_{tk} : flux des usagers dans la rue n_k .
- R : rayon d'une pico-cellule.
- R' : rayon d'une micro-cellule.
- R'' : rayon d'une macro-cellule.
- M : nombre des cellules qui couvrent la rue n_k .

Annexe C

Glossaire (2)

C :	Capacité d'une cellule (<i>capacity of cell</i>).
I :	Taux des arrivés des appels nouveaux a une cellule s^{-1} (<i>arrival rate for handoff calls to cell</i>).
I^H :	Taux des handovers dans une cellule (appel s^{-1}) (<i>handovers rate in a cell</i>).
m :	Taux de service moyen des appels nouveaux (s^{-1}) (<i>average channel holding time for new calls</i>)
$\frac{1}{m}$:	Durée de service moyen des appels nouveaux (s) (<i>average channel holding time for new calls</i>)
f :	Flux d'utilisateurs dans une rue (<i>users' flows in a rouad</i>)
a :	Taux de mobilité (transitions) d'usagers dans une cellule (<i>mobility (transition) rates in a cell</i>)
Θ :	Flux de transitions entre cellules (<i>transitions' flows beteen cells</i>)
S :	Densité d'usagers dans une rue (<i>users's density in a rouad</i>)
P_d :	Probabilité de rejet des appels handovers (dropping handovers calls)
P_b :	Probabilité de blocage des nouveaux appels (<i>blocking new calls</i>)
P^H :	Probabilité de handovers dans une cellule (<i>handovers probability</i>)
r :	Charge de trafic (Erl) (<i>load of traffic</i>)
GoS:	Grade de service (<i>grade of service</i>)
BTS:	Station de base (<i>base transceiver station</i>)

N:	Nombre des canaux dans une cellule (<i>number of channels in a cell</i>)
QoS :	Qualité de service (<i>quality of service</i>)
BB :	Bande passante counter (<i>bandwidth broker</i>)
SLA :	Niveau de service agreement (<i>service level agreement</i>)
CAC :	Contrôle d'admission d'appel (<i>call control admission</i>)
LARC :	Contrôle d'admission local (<i>local control admission</i>)
DiffServ :	<i>Differentiated service</i>
IntServ :	<i>Integrated service</i>
IP :	Protocole d'Internet (<i>Internet protocol</i>)
IETF :	<i>Internet Engineering Task Force</i>
UMTS :	<i>Universal Mobile Telecommunication System</i>
UTRAN :	<i>UMTS terrestrial radio access network</i>
CANAC :	Capacité de nouveaux appels /cellule (<i>new call capacity/cell</i>)
COPS :	<i>Common Open Policy Service protocol</i>
RSVP :	<i>Resource Reservation Protocol</i>
PEP :	<i>Policy Enforcement Point</i>
PDP :	<i>Policy Decision Point</i>
REQ :	<i>Request</i>
SLS :	<i>Client SI object</i>
TCP :	<i>Transmission Control</i>
FTP :	<i>File Transfert Protocol</i>
MMPP :	<i>Markov Modulated Process Poisson</i>
GGSN :	<i>Germanic Generative Synthax Network</i>
HTTP :	<i>Hyper Text Transfert Protocol</i>
MPEG :	<i>Moving Pictures Expert Group</i>

Annexe D

Paramètres de QoS

<i>Application</i>	<i>Customer demand</i>	<i>Amount of traffic</i>	<i>Data rate</i>	<i>Amount of data</i>	<i>Key performance parameters and target values</i>			
					<i>One-way delay</i>	<i>Delay variation</i>	<i>Information loss</i>	<i>Other</i>
voice	<i>Low</i>	<i>Low</i>	4-13 kb/s		< 150 msec preferred < 400 msec limite	< 1msec	< 3% FER	
Video-phone	<i>Low</i>	<i>Low</i>	32-384 kb/s		< 150 msec preferred < 400 msec limite	< 1msec	< 3% FER	Lip-synch :< 100msec
Video-audio	<i>Low</i>	<i>Low</i>	32-128 kb/s		< 10 sec	< 1 msec	< 1% FER	
e-mail (<i>server access</i>)	<i>High</i>	<i>High</i>		< 10KB	< 4 sec	N.A	Zero	
e-mail (<i>server to server</i>)	<i>High</i>	<i>High</i>		< 10KB	<i>Can be several hours</i>	N.A	Zero	
Web-browsing-HTML(www)	<i>High</i>	<i>High</i>		10KB	< 4 sec/page	N.A	Zero	
Bulk data Transfert/retrieval	<i>High</i>	<i>Medium</i>		10-10MB	< 10 sec	N.A	Zero	
image	<i>Medium</i>	<i>Medium</i>		< 1MB	< 10 sec	N.A	Zero	

Tableau 1. Les paramètres de la QoS

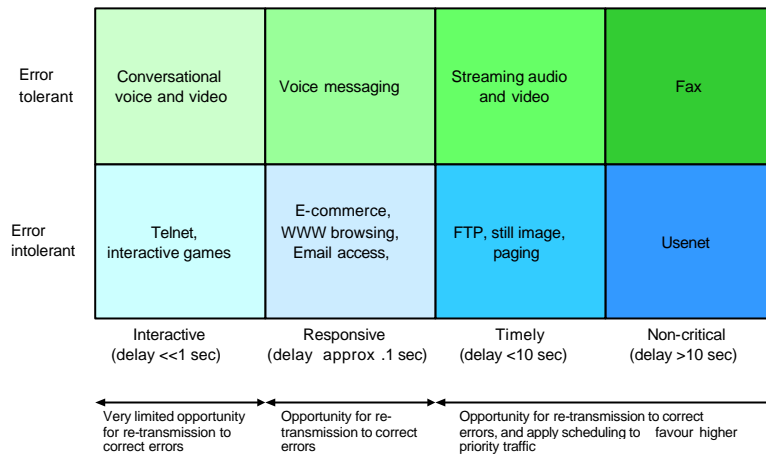


Tableau 2. Performances de la QoS

Pramètres	Distribution	Valeur moyennes
Inter-arrivée des appels « <i>conversational</i> »	Poissonien	0-120 s
Inter-arrivée des appels « <i>interactive</i> »	Bernouilli	0-120 s
Inter-arrivée des appels « <i>Background</i> »	Bernouilli	0-20 min
Inter-arrivée des appels « <i>Best effort</i> »	Bernouilli	0-120 s
Temps maximal d'attente des paquets dans le <i>buffer</i> avant transmission	Constant	2 s
Taille moyenne des emails	Géométrique	18.5 koctets
Taille moyenne des images JPEG	Géométrique	4 koctets

Tableau 3. Classes de services

Annexe E

Indicateurs majeurs (délivrés par le compteur OMC) pour la supervision et la planification du réseau cellulaire radio mobile

Libelle	Description	Indicateurs (AMCF)
<i>Ruptures des communications</i>		
TBLOCTC	Taux de blocage TCH	$(C12+C51) / (C16+C12+C51+C15)$
TCOUP	Taux de coupure de communications	$(C14c+C136+C139+C21) / (C17+C18-C46-C56-C62)$
ALLOCTC	Nombre d'allocation TCH	$C15+C16$
<i>Indicateurs de qualité des HO</i>		
ALLORTC	Nombre d'allocation TCH réussie	$C17+C18$
TEHO	Taux d'échec global du HO sortant (sur SDCCH et TCH)	$1 - [(C96+C86+C46+C56) / (C95+C85+C45+C55)]$
TFFHO	Taux d'efficacité du HO sortant (sur SDCCH et TCH)	$(C96+C86+C56) / (C96+C97+C98+C99+C56+C57+C58+C59+C86+C87+C88+C89+C46+C47+C48+C49)$
TEHO/entrant	Taux d'échec global du HO entrant (SDCCH et TCH)	$1 - (C92+C82+C42+C52) / (C91+C92+C93+C94+C51+C52+C53+C54+C81+C82+C83+C84+C41+C42+C43+C44)$
TFFHO/entrant	Taux d'efficacité du HO entrant (SDCCH et TCH)	$(C92+C82+C42+C52) / (C92+C93+C94+C52+C53+C54+C82+C83+C84+C42+C43+C44)$

<i>Nombre des HO inter-cellulaire</i>		
HOINDDE/inter	Nombre de HO demandé (SDCCH et TCH)	$(C91+C92+C03+C94+C51+C52+C53+C54+C81+C82+C83+C84+C41+C42+C43+C44)$
HODDE	Nombre de HO sortants demandés (sur SDCCH et TCH)	$(C95+C85+C45+C55)$
HOINEXE	Nombre de HO entrant exécutés (SDCCH et TCH)	$(C92+C93+C94+C52+C53+C54+C82+C83+C84+C42+C43+C44)$
HOEXE	Nombre de HO sortants exécutés (sur SDCCH et TCH)	$(C96+C97+C98+C99+C56+C57+C58+C59+C86+C87+C88+C89+C49+C47+C48+C49)$
HOINREU	Nombre de HO entrant réussis (SDCCH et TCH)	$(C92+C82+C42+C52)$
HOREU	Nombre de HO sortants réussis (sur SDCCH et TCH)	$(C96+C56+C86+C46)$
<i>Nombre des HO rapportes au nombre d'appels</i>		
NMHODA	Nombre de besoins de HO sortants demandés / appels initialisés sur SDCCH et TCH	$(C95+C85+C45+C55) / (C18)$
NMHOOA	Nombre de besoins de HO sortant réussis demandés / appels initialisés dans la cellule sur SDCCH et TCH	$(C96+C86+C46+C56) / (C18)$
NMHOIA	Nombre de besoins de HO entrants réussis / appels établi sur SDCCH et TCH	$(C92+C82+C42+C52) / (C18)$
NMHIDA	Nombre de besoins de HO entrants demandés / appels établi sur SDCCH et TCH	$(C91+C92+C93+C94+C51+C53+C54+C81+C82+C83+C84+C41+C42+C43+C44) / (C18)$
<i>Indicateurs de qualité de SDCCH et TCH</i>		
TRUPTTC	Taux de rupture TCH total	$(C14c+C136+C139+C21) / (C17+C18)$
TBLOCTC	Taux de blocage TCH	$(C12+C51) / (C16+C12+C51+C15)$
TRUPTSD	Taux de rupture SDCCH total	$(C137+C138+C07) / (C01+C02+C10)$

Annexe F

Protocoles de consensus

Procedure Consensus DWD (S_{ij}, T_r, D_d)

begin

(1) *broadcast propose vote (S_{ij}, T_r, D_d) by coordinator to all resource managers (including itself);*

(2) *let Cond-1 be at least x messages proposing “SLAmax” have been received by RM_i ;*

X denotes the set of objects from which messages proposing “SLAmax” have been received;

(3) *let Cond-1 be at least y messages proposing “SLAmoy” have been received by RM_i ;*

Y denotes the set of objects from which messages proposing “SLAmoy” have been received ;

(4) *let Cond-3 be deadline T_d occurrence ;*

(6) *wait (Cond-1 or Cond-2 or Cond-3) ;*

(7) *case Cond-1 ->*

let T be the greatest starting time T_r among objects $RM_j \in X$;

let D be the lowest deadline of objects $RM_j \in X$;

(7') if $D - T \geq d$ then “decision_i” : SLAmax else decision_i : SLAmin end ;

(8) *Cond-2 ->*

let T be the greatest starting time T_r among objects $RM_j \in Y$;

let D be the lowest deadline of objects $RM_j \in Y$;

(8') if $D - T \geq d$ then decision_i : SLAmoy else decision_i : SLAmin end ;

(9) *Cond-3 -> decision_i : SLAmin ;*

(10) *end case ;*

end.

VU :

VU :

Le Directeur de Thèse

Le Responsable de l'école Doctorale

VU pour autorisation de soutenance

Tunis, le

Références Bibliographiques

[PER98]Perkins C.E., “*Mobile networking in the internet*,” Mobile Networks and Applications, ACM/Baltzer, Vol.3 1998. pp. 19-334.

[WUI01]Wui K.et. Harmes J, “*QoS support in mobile ad hoc networks*,” Crossing Boundaries-an interdisciplinary Journal, Vol.1, No 1- Fall 2001.

[WAN]<http://perso.wanadoo.fr/michel.hubin/>.

[RA02][http://www.inria.fr/rapportsactivite/RA2002/hipercom/module 14.html](http://www.inria.fr/rapportsactivite/RA2002/hipercom/module%2014.html).

Références Bibliographiques

[AHM03]Ahmed, T. and al., “*QoS dynamic management of video flow in IP DiffServ architector,*” Proc. GRES’03, 2003, pp. 53-65.

[BEAI04]Ben Rejeb, S., Tabbane, S. and Choukair, Z., “*Combinig markovian chains and real time agreement protocols to allocate resources in wireless systems,*” AINA’04, 23 March 2004.

[BERR05]Ben Rejeb, S., Tabbane, S. and Choukair, Z., “*Real time agreement protocol and resource pre-reservation for wireless systems,*” RRM-WCN/IPCCC’05(First IEEE International Workshop on Radio Resource Management for Wireless Cellular Networks/24th IEEE International Performance Computing and Communications Conference), Phonix, Arizona, USA, 09 April 2005.

[CHAH8]Chahdra, P. and al., “*Network Support for application-oriented QoS,*” Proc. Of Sixth International Workshop on Quality of Service (IWQoS’98), pp. 187-195.

[CHA00]Chang, F. and al., “*Automatic configuration and run-time adaptation of distributed application,*” ninth IEEE Intl. Symposium on High Performance Distributed Computing (HPDC), August 2000, pp. 11-20.

[CHE00]Cheng, Y. and Zhuang, W., “*DiffServ resources allocation for fast handoff in wireless mobile internet,*” IEEE Communications Magazine May 2000.

[CHEN02]Cheng, Yu., Weihua, Z., “*DiffServ resources allocation for fast handoff in wireless mobile Internet,*” IEEE Communications Magazine May 2002, pp. 130-6.

[HAO02]Hao, F. and al., “*QoS routing for anycast communications: motivation and an architecture for diffServ networks,*” IEEE Communication Magazine, June 2002, pp.4 8-56.

[KHA02]Khalil, I. and Braun, T., “*Edge provisioning and fairness in VPN-Diffserv networks,*” the 9th International Conference on computer communication and network (ICCCN 2000), 2002, Vol.2, pp. 1220-1224.

[JIA01]Jia, Y. and al., “*A new architecture of providing end-to-end quality-of-service for differentiated service network,*” IEEE 2001.

[LAI98]Lai, F.S., Mistic, J. and al., “*Complete sharing versus partitioning: Quality of service management for wireless multimedia networks,*” In Proc. of the 7th int. Conf. On Computer Communications and Networks (IC3N’98), Lafayette, Oct.1998.

[MISIC97]Mistic, J. and al., “*Event based resource estimation in admission control for wireless networks with heterogeneous traffic,*” In Mobile Computing and Communications Review, Vol.1, No.4, 1997.

[MISIC98]Mistic, J. and al., “*Quality of Service Management for Wireless Networks with Heterogeneous Traffic,*” IEEE 1998.

[MOS99]Mostefaoui, A and Raynal, M., “*Solving Consensus Using Chandra-Toueg’s Unreliable Failure Detectors: a Generic Quorum-Based Approach,*” Proc. 13th Int. Symposium on Distributed Computing (DISC’99), (Formerly WDAG), Springer-Verlag LNCS (P. Jayanti Ed.), pp. 49-64, Bratislava (Slovakia), 1999

[RAY97]Raynal, M., “*Real time dependable decision in timed asynchronous distributed systems,*” Institue of Electical and Electronics Engineer 1997.