

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITE MOULOUD MAMMARI de TIZI OUZOU



Mémoire de fin de cycle LMD académique  
Domaine Mathématique et informatique  
Filière Informatique  
Spécialité Système Informatique

En vue de l'obtention d'un Master académique en informatique

# Construction Automatique De Zones De Localisation Dans Les Réseaux Cellulaires 3G En Utilisant Le Clustering

Proposé et dirigé par :

Mr. DAOUI M.

Réalisé par :

Mlle AID Aicha

Devant le jury :

Promotion : 2010 / 2011



# Remerciements

*En préambule à ce mémoire, je tenais particulièrement à témoigner ma profonde gratitude à toute personne ayant contribué d'une façon ou d'une autre à sa réalisation.*

*Mes sincères remerciements s'adressent à mon promoteur Monsieur Daoui, qui malgré ses multiples responsabilités, s'est toujours montré à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi pour l'inspiration, l'aide et le temps qu'il a bien voulu me consacrer et sans qui ce mémoire n'aurait jamais vu le jour.*

*Mes remerciements sont également adresser aux membres du jury qui ont fait l'honneur d'examiner mon travail.*

De la part de Mlle AID Aicha

# Dédicace

A ceux sans qui je n'aurais jamais pu connaître le plaisir de dédicacer ce si humble mémoire de 100 et quelques pages, ceux qui ont fait de moi la personne que je suis devenue :

A mes chers parents et mes chers grands parents, dont l'amour et le soutien à chaque instant de ma misérable petite vie, m'ont donné le courage, la force et m'ont permis d'accomplir mes souhaits les plus fous. Mille fois merci pour tous vos sacrifices. Les mots sont très faibles pour vous l'affirmer : vous resteriez à toujours le nec plus ultra de ma pompe à sang.

A Starsky Phelps, mon unique frère, biologiquement parlant. Yooh, ne baisse jamais les bras ; Il viendra le jour ou tu pulvériseras ce record et t'enfileras la médaille d'or. Lochte ... tu l'auras un jour ... tu l'auras.

A mes amis et aux membres de la famille Layland : Rosa, la belle brune aux yeux ravageurs dont la folie et les bêtises m'ont toujours impressionné. J'espère que tu l'auras ton master toi aussi, ainsi que bina, sauf que le votre sera made in France. Ehh oui, on n'a pas la chance de naître tous égaux !! Hocine, à la conduite fast and furious, sans qui j'en serai encore au stade de : « Zut ! J'y comprend rien en STO », merci a khou. A3333e à miro, ma touche metaleuse dans ce monde de musique commerciale. C'est vrai que Hagrouna sadi9i mais t'en fait pas fréro, on n'ira au prochain Sonisphere avec The Big Four en tête d'affiche et on finira nos jours à engouffrer des cream cheese brownies en Alabama, their sweet home Alabama. Lynda thoura adla79agh, t'inquiète sœurlette l'humanité finira par te comprendra un jour ... ou pas. Sache que c'est la terre qui tourne autour du soleil, pas toi. Enfin, mes honorables citoyens et amis facebook, je vous poke tous.

Et je tiens surtout à remercier ma fainéantise légendaire et ma répulsion totale lors du dernier ramadan envers le verbe cuisiner, qui m'ont permis de me réfugier très souvent dans la phrase « peut pas venir, je révise ».

That's All Folks.

Dixit moi-même, Mille AID Aicha

**A**vec les progrès technologiques et l'évolution de la société, les télécommunications sont devenues un besoin crucial. Pour répondre à ce besoin, on doit concevoir des réseaux de communications mobiles permettant aux abonnés de communiquer et d'échanger des informations à l'aide des infrastructures sans fil tout en gardant une mobilité étendue à l'intérieur de ces réseaux.

Les réseaux de télécommunications mobiles sont constitués d'installations fixes et d'un grand nombre d'unités mobiles. Celles-ci peuvent se déplacer librement à l'intérieur du réseau sans subir d'interruption de communications. Chaque unité mobile rapporte périodiquement sa position dans le réseau selon un processus dit de mise à jour de localisation. Lorsqu'une unité mobile reçoit un appel, le réseau doit déterminer la position exacte de cette unité en exécutant une procédure dite de paging. Les coûts associés à la gestion de localisation des abonnés du réseau constituent la majeure partie des dépenses engendrées par l'exploitation d'un réseau de communication mobile. Pour cette raison, de nombreuses recherches sont orientées vers l'optimisation du processus de gestion de localisation des abonnés dans les systèmes cellulaires.

Le travail de recherche dont il est question ici consiste essentiellement à concevoir et à implanter un mécanisme de gestion de localisation des unités mobiles dans un réseau cellulaire de troisième génération. Nous avons proposé un algorithme pour optimiser les coûts associés à la gestion de la localisation. En effet, la proposition se constitue d'un algorithme de gestion de localisation utilisant le clustering, une technique du Data mining, pour créer automatiquement des zones de localisation et de paging individuelles pour chaque abonné. Ensuite, on a utilisé un modèle de mobilité servant à tester et à implémenter l'algorithme proposé. Ce modèle est basé sur les activités utilisateur déjà collecté lors d'une précédente enquête.

---

Des simulations ont été effectuées dans le but d'étudier les régularités de comportement de la fonction de coût de localisation. Cette proposition permet d'obtenir des économies significatives et de fournir une qualité de services satisfaisante aux abonnés. Pour en évaluer la performance, nous avons soumis l'implantation à une série de tests. Les résultats obtenus confirment l'efficacité de cette méthode.

*Mots Clés:*

Réseau cellulaire, 3G, Mobilité, Gestion de localisation, Zone de localisation, Mise à jour de localisation, Paging, Data Mining, Clustering.

**W**ith rapid technological advances and the evolution of society, communication became a crucial necessity. To meet this need, we have to design wireless cellular networks that let the subscribers communicate and exchange information using the wireless infrastructures and be free to travel within those networks.

Cellular networks consist of shared wired infrastructure and a number of mobile users carrying portable devices. Mobile users have access to this infrastructure independently of their physical location, and they are free to move within the wireless network area without interruption of services. Each mobile unit reports its location to the network according to a reporting process called location update. When an incoming call arrives for one mobile unit, the network has to locate this unit by executing the process called terminal paging. The location management costs constitute a major part of all expenses involved in operating a cellular network. For this reason, much research is turned toward optimization of location management techniques for wireless communication systems.

This report aims at designing and implementing location management algorithm for mobile units in 3G mobile networks. We proposed an algorithm to optimize the costs involved in location management. Indeed, the contribution is a proposal for a location management algorithm based on the use of one of the data mining techniques which is clustering, to automatically create personal location and paging areas. Then, we use a realistic mobility model to test the proposed location management algorithm. The model is based on activity selections of users, collected through a previous survey.

Simulations were performed to study the behavior of the location function's and their cost. The use of this algorithm enables us to get significant savings and to provide satisfactory quality of services to mobile subscribers. In order to evaluate the performance, we submitted our implantation to a range of tests. The results obtained confirm an efficiency of our method.

**Key Words:**

Cellular Network, 3G, Mobility, Location management, Location area, Location Update procedure, Paging procedure, Data Mining, Clustering.

---

# Table des matières v

---

Résumé	i
Abstract	iii
Table des matières	v
Liste des figures et des tableaux	vii
Liste acronymes	ix
Introduction générale	1
Objectif du travail	3
Organisation du mémoire	3
Chapitre I : <b>Généralité et Gestion de localisation dans les réseaux cellulaires 3G</b>	5
<b>I- Les réseaux 3G</b>	<b>8</b>
<b>1) Présentation</b>	<b>8</b>
<b>2) Le réseau UMTS</b>	<b>9</b>
➤ Caractéristiques générales	9
<b>3) Architecture du réseau UMTS</b>	<b>12</b>
A. Le réseau d'accès radio UTRAN	13
B. Le réseau cœur CN (Core Network)	16
C. Station Mobile (UE, User Equipment)	19
<b>4) Interface radio de l'UTRAN : Découpage en strates</b>	<b>20</b>
1. AS : Access Stratum	20
2. NAS : Non Access Stratum	28
<b>5) Gestion de la mobilité</b>	<b>28</b>
➤ Principes de base de la gestion de mobilité	28
➤	

<b>II- La gestion de Localisation dans l'UMTS</b> .....	<b>29</b>
➤ Concepts généraux .....	<b>29</b>
➤ Attachement au réseau UMTS .....	<b>33</b>
a) <i>La procédure IMSI attach</i> .....	<b>35</b>
b) La procédure GPRS attach .....	<b>37</b>
➤ La procédure de mise à jour de la localisation .....	<b>39</b>
a) Location updating procedure .....	<b>39</b>
b) Rounting Area updating procedure .....	<b>41</b>
➤ Le paging .....	<b>41</b>
Chapitre II : <b>la gestion de localisation</b> .....	<b>42</b>
<b>1) Principes de localisation</b> .....	<b>45</b>
✓ Stratégies de mise à jour .....	<b>46</b>
<b>2) Planification des zones de localisation</b> .....	<b>49</b>
✓ Éléments de la problématique .....	<b>49</b>
<b>3) Revue des stratégies et des travaux précédents sur la gestion de localisation</b>	
✓ Classification des stratégies de gestion de localisation .....	<b>52</b>
a) Propositions basées sur l'interrogation (Interrogation-Based Proposals) .....	<b>53</b>
b) Propositions basées sur la mise à jour et le paging (Location update and Paging proposals) .....	<b>54</b>
1. Variation dans l'implémentation des LAs .....	<b>54</b>

➤ Static LAs ..... 54

➤ Dynamic LAs ..... 54

2. Mise à jour avec des alternatives de déclenchement ..... 55

- Mise à jour basée sur le temps (Time – Based) ..... 56
- Mise à jour basée sur le mouvement (Movement-Based) 56
- Mise à jour basée sur la distance (Distance-Based) .... 56

4) Conclusion ..... 57

Chapitre III : **Le Data Mining** ..... 58

1) Définition du Data Mining ..... 59

✓ Intérêt du Data Mining ..... 59

2) Difficultés du data mining ..... 60

3) Domaines d’application du data mining ..... 61

4) Les méthodologies du Data Mining ..... 62

5) Etat de l’art ..... 65

➤ La Classification ..... 65

a) Classification supervisée de données ..... 66

b) Classification non supervisée de données (Clustering) .... 67

➤ Terminologie ..... 68

➤ Les algorithmes du clustering ..... 70

a) Les algorithmes de partitionnement ..... 71

✓ L’algorithme Des k-Means ..... 71

<b>6) Conclusion</b> .....	<b>73</b>
----------------------------	-----------

Chapitre IV : **L’algorithme de gestion de localisation proposé, basé sur le clustering**

<b>1) Motivation</b> .....	<b>75</b>
<b>2) Historique de la mobilité et profil utilisateur</b> .....	<b>76</b>
<b>3) Description de l’algorithme proposé</b> .....	<b>77</b>
a) Construction des zones de localisation et des zones de paging . . . .	<b>78</b>
b) La stratégie de mise à jour .....	<b>80</b>
c) La stratégie de paging .....	<b>81</b>

Chapitre IV : **Description du modèle de la mobilité et Evaluation des résultats. 82**

<b>1) Les éléments de la demande de voyage</b> .....	<b>84</b>
<b>2) Le modèle de mobilité proposé basé sur l’activité</b> .....	<b>85</b>
a) Exigences du modèle et disponibilité des données .....	<b>85</b>
b) Aperçu de modèle de mobilité .....	<b>86</b>
c) Les données d’entrée .....	<b>88</b>
d) La procédure de simulation .....	<b>91</b>
<b>3) Analyse des résultats</b> .....	<b>96</b>
1. Description des données et des procédures .....	<b>96</b>
2. Comparaison des coûts liés à la signalisation .....	<b>97</b>
a) Performance moyenne globale .....	<b>97</b>
b) Coût total de la gestion de localisation .....	<b>101</b>

<b>4) Conclusion</b>	.....	<b>105</b>
----------------------	-------	------------

Conclusion et perspectives	.....	<b>106</b>
Conclusion	.....	<b>106</b>
Evaluation et contribution	.....	<b>107</b>
Perspectives	.....	<b>109</b>
Références	.....	<b>110</b>

# Liste des figures et des tableaux x

---

- I-1:- Hiérarchie des cellules de l'UMTS
- I-2 :- Les services UMTS et leurs besoin en débit
- I-3 :- Vue d'ensemble d'un réseau UMTS
- I-4 :- Architecture du réseau d'accès
- I-5 :- Architecture du réseau cœur de l'UMTS
- I-6 :- Architecture globale du réseau UMTS
- I-7 :- Vue en couches de l'interface radio UTRAN
- I-8 :- Les différents canaux en UMTS
- I-9 :- Les différents états d'une station mobile
- I-10 :- Etablissement d'une connexion RRC
- I-11 :- Le découpage du réseau en zones géographiques
- I-12:- La procédure IMSI Attach
- I-13:- La procédure Combined GPRS/IMSI attach
- I-14 :- Mise à jour de localisation Inter MSC/VLR
- I-15:- Paging type 1
- I-16:- Paging type 2
  
- II-1 :- Classification des stratégies de gestion de localisation
  
- III-1 – Le Data Mining Dans Le Processus KDD
- III.2- : Les processus de la méthodologie CRISP-DM
- III-3: - Illustration de la détection de clusters par mesure de distances
- III-4 :- Exemple de centroïde
- III-5 :- Exemple de médoïde

III-6 : Partitionnement basé sur k-Means

III – 7 : - Dendrogramme

III – 8 : - Illustration du concept "accessible"

III – 9 : - Illustration du concept "connecté"

Tab V-1 :- Exemple de données de la matrice de transition d'activité

Tab V-2 :- Exemple de données de la matrice de durée d'activité

Tab V-3 : - loi de probabilité des appels entrants au cours d'une journée

V-1 :- Schéma de flux de données pour la génération de voyage dans le modèle de mobilité proposé.

V-2 :- Nombre moyen des messages de mise à jour des différents algorithmes

V-3 :- Nombre moyen des messages paging générés pour 3 appels entrants quotidiens

V-4 :- Nombre moyen des messages paging générés pour 9 appels entrants quotidiens

V-5 :- Nombre moyen des messages paging générés pour 15 appels entrants quotidiens

V-6 :- Coût total de la gestion de localisation ( $c=5$ ) pour 3 appels entrants

V-7 :- Coût total de la gestion de localisation ( $c=5$ ) pour 9 appels entrants

V-8 :- Coût total de la gestion de localisation ( $c=10$ ) pour 9 appels entrants

V-9 :- Coût total de la gestion de localisation ( $c=5$ ) pour 12 appels entrants

V-10 :- Coût total de la gestion de localisation ( $c=10$ ) pour 12 appels entrants

3G : Troisième Génération

UMTS : Universal Mobile Telecommunication Services

GSM: Global System for Mobile communication

AMPS: Advanced Mobile Phone System

TACS: Total Access Communication System

ETACS: Extended TACS

GPRS: General Packet Radio Service

EDGE: Enhanced Data rate for GSM Evolution

HSDPA: High-Speed Downlink Packet Access

W-CDMA: Wideband Code Division Multiple Access

CDMA: Code Division Multiple Access

TDMA: Time Division Multiple Access

FDMA: Frequency Division Multiple Access

RAN: Radio Access Network

UTRAN: UMTS Terrestrial RAN

CN: Core Network

3GPP: 3rd Generation Partnership Project

CS: Circuit Switched

PS: Packet Switched

UE: User Equipment

RNC: Radio Network Controller

RRC: Radio Resource Control

HLR: Home Location Register

AuC: Authentication Center

EIR: Equipment Identity Register

IMEI: International Mobile station Equipment Identity

MSC: Mobile-services Switching Center

GMSC: Gateway MSC

PSTN: Public Switched Telephone Network

VLR: Visitor Location Register

LA: Location Area

PA: Paging Area

RA: Routing Area

USIM: UMTS Subscriber Identity Module

IMSI : International Mobile Subscriber Identity

TMSI : Temporary Mobile Subscriber Identity

AS: Access Stratum

NAS: Non Access Stratum

MAC: Medium Access Control

RLC: Radio Link Control

PLMN: Public Land Mobile Network

LU: Location update

PG: Paging

# Introduction Générale

---

Le développement de techniques de télécommunication occupe une place importante dans les activités de recherche des pays industrialisés. Durant les dernières années, les réseaux de télécommunications mobiles ont connu une croissance phénoménale en nombre de générations développées (1G, 2G, 3G, 4G), de services offerts et de technologies utilisées. Les systèmes de téléphonie cellulaire, les téléavertisseurs et le téléphone mobile sont devenus très populaires et la demande ne cesse d'augmenter. Les services offerts dans la troisième génération de ces réseaux mobiles par exemple, incluent les transferts de données, la vidéophonie et les communications vocales, formant ainsi les services personnels de communication PCS (Personal Communication Services), qui peuvent être obtenus partout, en tout temps, grâce aux petites unités glissées dans nos poches. En effet, le principal avantage qu'offrent ces réseaux par rapport aux réseaux classiques de télécommunications est la liberté de mouvement offerte aux abonnés qui peuvent dès lors communiquer sans subir d'interruptions ou de détérioration de services.

Cela dit, l'introduction de la mobilité nécessite la définition de nouvelles fonctions par rapport aux réseaux fixes classiques. Par exemple, lorsqu'un abonné reçoit un appel, le réseau doit retrouver sa position en temps limité. Pour répondre à cette exigence, le réseau doit employer des stratégies efficaces de localisation des abonnés. En cause, la surface des réseaux mobiles est divisée en petites zones géographiques appelées cellules. Chaque cellule est desservie par un équipement de communication mobile, appelé Node B dans la 3G, dont le rôle principal consiste à allouer des canaux de communication aux abonnés. Afin d'éviter des interférences, on utilise des canaux de communication différents pour des cellules avoisinantes. Les cellules avoisinantes se chevauchent pour permettre aux abonnés mobiles de transiter entre elles sans subir d'interruption ou de détérioration de communications.

La gestion de localisation engendre ainsi un trafic de signalisation important sur l'interface radio et dans le réseau. Afin de minimiser des coûts associés à la localisation, plusieurs cellules sont regroupées ensemble pour former des zones de localisation. La localisation des abonnés consiste donc en deux actions principales : mise à jour de localisation et le paging

de l'unité mobile. La procédure de mise à jour s'effectue chaque fois qu'une unité mobile change sa zone de localisation. De cette façon, le réseau connaît en tout temps la position approximative de chaque unité. A l'arrivée d'un appel entrant, le réseau effectue la procédure de paging qui consiste à rechercher l'abonné. Suite à cette procédure, le réseau localise le Node B desservant l'unité mobile appelée afin d'établir la communication.

Pour supporter les communications des abonnés, les réseaux mobiles utilisent une bande de radiofréquences très restreinte et les coûts associés à l'utilisation de cette bande sont très élevés. La gestion de la mobilité des abonnés, à son tour, nécessite l'utilisation des mêmes radiofréquences. Avec l'augmentation du nombre d'abonnés surgit le problème de surcharge des réseaux dû aux procédures de mise à jour et de paging. Dans le but d'optimiser les ressources et les coûts associés à ces procédures, il faut trouver des stratégies efficaces de gestion de localisation des abonnés. Ces méthodes incluent le choix de la taille de zones de localisation, d'un modèle de regroupement des cellules en zones de localisation et d'une stratégie de paging.

Il existe plusieurs stratégies de regroupement de cellules en zones de localisation. Les deux méthodes extrêmes et les moins efficaces consistent, d'une part, à former des zones de localisation d'une seule cellule et, d'autre part, à regrouper toutes les cellules dans une seule zone de localisation. Ainsi, dans le premier cas, le réseau peut connaître la position exacte de chaque unité mobile et économiser sur les coûts reliés au paging. Cependant, à chaque changement de cellule les unités mobiles sont obligées d'effectuer une procédure de mise à jour, ce qui est très coûteux. Dans le deuxième cas, par contre, les unités mobiles n'exécutent pas de procédure de mise à jour. Toutefois, le réseau ne peut connaître la position exacte de chaque unité mobile et doit procéder à un paging sur tout l'ensemble des cellules du réseau, ce qui est très coûteux aussi.

Il faut donc trouver la taille optimale de la zone de localisation afin de minimiser le coût total constitué des coûts de mise à jour et de paging. En ce qui concerne le paging, la méthode la moins efficace consiste à effectuer la recherche simultanément dans toutes les cellules constituant la zone de localisation. Pour minimiser son coût on peut subdiviser la zone de localisation en plusieurs régions appelées zones de paging et effectuer ainsi la recherche de

l'unité mobile séquentiellement dans chaque région, selon un certain modèle, en respectant un certain délai préétabli.

Un effort considérable de recherche a été investi dans la minimisation des coûts associés à la gestion de localisation des abonnés dans les réseaux de communications mobiles. En effet, plusieurs stratégies et algorithmes ont été proposés dans la littérature, utilisant des techniques aussi nombreuses que différentes, parfois combinés dans le seul but de soulager le réseau des signalisations, ne serait-ce qu'un minimum.

### o Objectif du mémoire

L'inexistence d'une méthode efficace pour la résolution du problème de gestion de localisation des abonnés dans les réseaux de communications mobiles rend nécessaire l'élaboration de nouvelles stratégies de localisation. Ces stratégies consistent à trouver un équilibre entre les ressources et les coûts engagés pour la mise à jour et le paging.

L'élaboration et l'implantation d'une de ces méthodes fait l'objet de ce mémoire. En effet, ce mémoire a pour principal objectif la mise au point d'une méthode efficace permettant de minimiser le coût total de la gestion de localisation, en choisissant les zones de localisation de façon judicieuse et diviser la zone de localisation en régions de paging de façon optimale.

En premier lieu, la proposition est constituée d'un algorithme de gestion de localisation utilisant le clustering, une technique du Data mining, pour créer automatiquement des zones de localisation et de paging individuelles pour chaque abonné. En effet, le but du clustering consiste à regrouper les cellules en cluster de manière judicieuse prenant en compte les déplacements et les activités entreprises par les abonnés à chaque instant de son quotidien. Les clusters représentent dans notre cas les zones de localisation et les zones de paging pour chaque abonné. En second lieu, l'évaluation de l'algorithme proposé par un modèle de mobilité réaliste servant à tester et à implémenter l'algorithme proposé. Ce modèle est basé sur les activités utilisateur déjà collectées lors d'une précédente enquête.

## o Structure du mémoire

Ce mémoire est organisé en 5 chapitres. Les deux premiers sont orientés fonctionnement des réseaux cellulaires et de la gestion de localisation dans les réseaux 3G. Le premier chapitre présente les concepts généraux de l'architecture du réseau principal de la 3G, à savoir l'UMTS. Ceci consiste à présenter les points importants du fonctionnement de cette norme qui serve de passerelle pour le chapitre suivant, notamment de la manière dont elle gère la mobilité de ses abonnés en termes de paging et de mise à jour. En effet, le chapitre II introduit le mécanisme de gestion de localisation de façon globale, ainsi que les différentes stratégies d'optimisation de cette dernière.

Le troisième chapitre traitera le data mining en général, et le clustering en particulier. On commentera à ce titre, les différentes techniques et algorithmes existants dans ce domaine. Les deux derniers chapitres seront consacrés à l'explication de notre solution proposée. L'algorithme de gestion de localisation développé est décrit dans le quatrième chapitre alors que le modèle de mobilité utilisé sera détaillé dans le chapitre IV. Enfin, nous achèverons le mémoire par une conclusion sur l'ensemble du travail et quelques perspectives.

# Chapitre I

## *Généralités Et Gestion De Localisation Dans Les Réseaux Cellulaires 3G*

**D**iverses inventions telles que le téléphone, le télégraphe, l'ordinateur, le téléphone cellulaire ont donné lieu à une forme nouvelle de communications appelée les **télécommunications**. Le besoin de transmettre la voix, les données, les images, et d'effectuer des traitements à distance est devenu une nécessité quotidienne. La concrétisation de ce phénomène, de plus en plus répandu, demeure l'apparition et l'évolution rapide **des réseaux de télécommunications mobiles**. Le principal avantage qu'offrent ces réseaux par rapport aux réseaux classiques de télécommunications est la liberté de mouvement quasi illimité offerte aux abonnés, qui peuvent dès lors communiquer partout et en tout temps sans subir d'interruptions ou de détérioration de services. Durant ces dernières années, ces réseaux de communications mobiles ont connu une croissance phénoménale en nombre de services offerts et de technologies utilisées. On distingue plusieurs générations et sous générations.

La première génération de téléphonie mobile (notée 1G) possédait un fonctionnement analogique et était constituée d'appareils relativement volumineux. Il s'agissait principalement des standards suivants : AMPS (Advanced Mobile Phone System), TACS (Total Access Communication System) qui est la version européenne du modèle AMPS et enfin une version améliorée du TACS nommée ETACS (Extended TACS).

La seconde génération de réseaux mobiles (notée 2G) est numérique, orienté mode circuit, avec le standard le plus répandu dans le monde : le GSM (Global System for Mobile communication), qui permet un transfert de données à faible vitesse. Il existe une génération dite deux et demi (2G+) avec le GPRS (*General Packet Radio Service*) qui n'est qu'une évolution du GSM, et qui permet le transfert de voix ou de données numériques de volume modéré avec une commutation de paquet. Une autre forme d'évolution des systèmes mobiles de seconde génération est incarnée par les systèmes EDGE (*Enhanced Data rate for GSM Evolution*) notés 2,75G, permettant le transfert simultanés de voix et de données numériques avec un débit théorique plus grand.

C'est donc dans le cadre d'une homogénéisation globale des standards qu'est née l'idée d'une troisième génération de ces systèmes. La 3G, souvent assimilée en Europe à l'UMTS (*Universal Mobile Telecommunication System*), a été principalement conçue et imaginée afin de satisfaire l'acheminement de données multimédia (accès à Internet, aux réseaux d'entreprise, à des services de messageries, de visioconférence, de jeux, ...). Au même titre que le GPRS et EDGE ont permis, sans modifier le principe de base des communications, à leurs générations de systèmes d'évoluer en terme d'efficacité. Le HSDPA (*High-Speed Downlink Packet Access*) est déjà considéré comme le standard 3.5G des télécommunications mobiles et le WIMAX comme le standard 4G.

Les réseaux mobiles issus de la 2G sont appelés **réseaux cellulaires**. Dans un système cellulaire, la zone géographique couverte est divisée en cellules. Une cellule a une forme circulaire mais dépend en réalité de la topographie de la zone qui est servie par l'antenne de la cellule. Pour plus de clarté, on peut les illustrer par des hexagones. La taille des cellules peut varier entre 0.5 et 35 km et dépend de la densité d'utilisateurs, de la puissance de l'émetteur/récepteur, et de la topographie. Les cellules sont regroupées en bloc (appelé motif ou cluster). On en distingue trois types de cellule :

Une pico-cellule qui permet des débits de l'ordre de 2 Mbits/s lors d'un déplacement avec une vitesse moyenne de 10 km/h (marche à pied, déplacement en intérieur, etc.).

Une micro-cellule qui permet des débits de l'ordre de 384 kbits/s lors d'un déplacement avec une vitesse moyenne de 120 km/h (véhicule, transports en commun, etc.).

Une macro-cellule qui permet des débits de l'ordre de 144 kbits/s lors d'un déplacement avec une vitesse moyenne de 500 km/h (Train à Grande Vitesse, etc.).

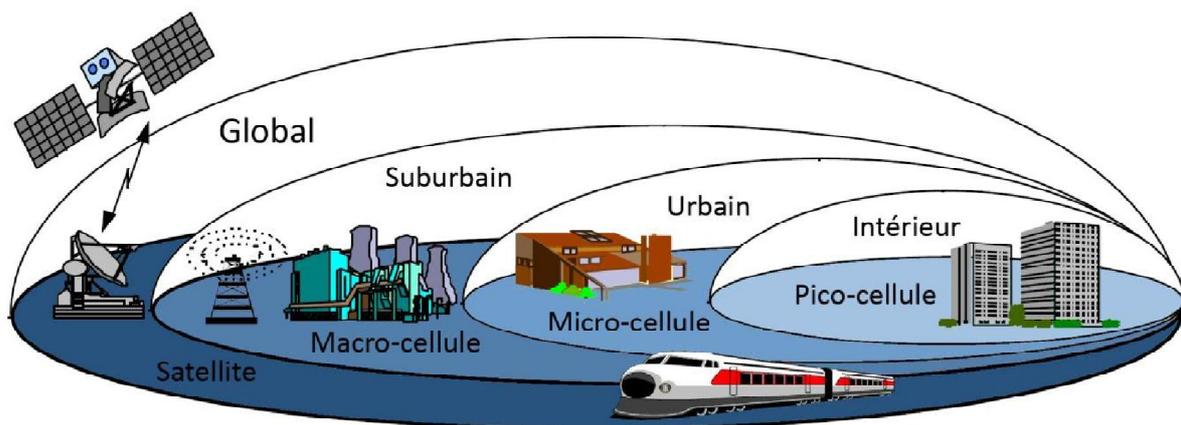


Fig I – 1 : – Hiérarchie des cellules de l'UMTS

Dans la partie suivante, nous présenterons l'architecture générale du réseau mobile de troisième génération UMTS, ses caractéristiques et ses configurations, les méthodes d'allocation des canaux, ainsi que la planification des zones de localisation et la gestion de localisation des abonnés.

# I. Les réseaux 3G

## 1. Présentation

Il existe plusieurs technologies 3G dans le monde. Ces normes permettent de transporter les données à haut débit sur la même connexion. Chacune d'elles suivent les recommandations IMT2000\*. Suivant les continents, la norme utilisée est différentes :

- Europe : UMTS (Universal Mobile Telecommunications System)
- Amérique : CDMA-2000
- Japon et Corée : W-CDMA (Wideband Code Division Multiple Access)
- Chine : TD-SCDMA

Cela dit, nous allons nous intéresser de plus près au réseau UMTS.

## 2. Le réseau UMTS

L'UMTS est donc la norme de télécommunications de troisième génération utilisée en Europe et est basée sur la technologie W-CDMA (utilisée au Japon et en Corée), une évolution de la CDMA. Il ne remplaçait pas les réseaux 2G existant puisque la coexistence entre ces deux réseaux est possible. Avec un débit supérieur aux technologies précédentes, ce standard permet à la fois la téléphonie mobile et le transport de données (images vidéo en direct, visioconférence mobile, etc.). Son déploiement, initialement prévu pour le début du siècle a été freiné en raison de son coût et de la mauvaise conjoncture économique du monde des télécommunications suite à l'éclatement de la bulle internet.

Le 1er décembre 2002, l'opérateur norvégien Telenor a annoncé le déploiement du premier réseau commercial UMTS. L'opérateur autrichien Mobilkom Austria a quant à lui lancé le premier service commercial UMTS le 25 septembre 2002. En France, SFR a lancé son offre commerciale le 10 novembre 2004 et Orange a fait de même le 9 décembre 2004. L'opérateur Bouygues Telecom a préféré se concentrer sur la technologie EDGE en 2005, pour offrir les mêmes types de services (excepté la visiophonie) avec un investissement moindre ; néanmoins, Bouygues Telecom dispose d'une licence UMTS et a été tenu, du fait

\*IMT-2000 est le sigle choisi par l'Union internationale des télécommunications (UIT) pour désigner les cinq technologies d'accès radio des systèmes cellulaires de la troisième génération qui sont retenues parmi les dix technologies proposées par les différents organismes de standardisation des membres de l'UIT.

### ➤ Caractéristiques générales

Les technologies développées autour de la norme UMTS conduisent à une amélioration significative des **débits**. Il a été décidé que l'UMTS serait conçu de manière à assurer les débits suivants : **144 kbit/s** en environnement rural extérieur, **384 kbit/s** en environnement urbain extérieur et pouvant atteindre les **2 Mbit/s** pour des faibles distances à l'intérieur d'un bâtiment couvert. De tels débits sont significativement supérieurs à ceux permis par les réseaux GSM actuels (9,6 kbit/s) ou par le GPRS.

Cette amélioration des débits est rendue possible par l'évolution des technologies radio qui autorise une meilleure efficacité spectrale et l'exploitation de bandes de spectre de fréquence supérieure à celles utilisées par la technologie GSM [2]. Alors que les réseaux GSM déployés au cours des dernières années reposaient sur l'utilisation de **bandes de fréquences** autour de 900 MHz et de 1800 MHz, la norme UMTS exploite de nouvelles zones du spectre (notamment les bandes **1920-1980 MHz** et **2110-2170 MHz**).

- ✓ **W-CDMA** : Comme indiqué précédemment, UMTS repose sur la technologie W-CDMA, une technique dite à étalement de spectre, alors que l'accès multiple pour le GSM se fait par une combinaison de division temporelle TDMA et de division fréquentielle FDMA. Le CDMA (Code Division Multiple Access) est basé sur la répartition par codes. En effet, chaque utilisateur est différencié du reste des utilisateurs par un code N qui lui a été alloué au début de sa communication. Dans ce cas, pour écouter l'utilisateur, le récepteur n'a qu'à multiplier le signal reçu par le code N associé à cet utilisateur.

Afin de couvrir l'ensemble des besoins présents et futurs des services envisagés pour l'UMTS, quatre classes ont été définies afin de regrouper les services en fonction de leurs contraintes respectives :

La classe A (Conversation) qui regroupe tous les services bidirectionnels impliquant deux interlocuteurs ou plus (phonie, visiophonie, jeux interactifs). Les services de cette classe sont caractérisés par un délai de transfert de 100 à 200 ms, une tolérance aux erreurs de

transmission qui permet d'offrir une qualité de service acceptable même en présence d'erreurs de transmission.

La classe B (Streaming) qui regroupe les services impliquant un utilisateur et une base de données (vidéo à la demande, diffusion radiophonique, transfert d'images). Les délais acceptables de transfert sont plus importants.

La classe C (Interactive) destinée à des échanges entre l'équipement usager et le réseau comme la navigation Internet qui engendre une requête et une réponse par le serveur distant.

La classe D (Background) qui affiche la plus faible priorité, permet des transferts de type traitements par lots qui ne demandent pas de temps réel et un minimum d'interactivité (envoi et réception de messages électroniques).

Les principales contraintes retenues pour la définition des classes de services de l'UMTS sont les suivantes :

- Le délai de transfert de l'information.
- La variation du délai de transfert des informations.
- La tolérance aux erreurs de transmission.

Les quatre classes de services définies dans le cadre de l'UMTS peuvent se répartir en deux groupes : Les classes A et B pour les applications à contrainte temps réel ; Les classes C et D pour les applications de données sensibles aux erreurs de transmission.

### ✓ Les services de l'UMTS

Le schéma ci-après présente les différents services que propose l'UMTS. Sur l'axe des ordonnées se trouve le débit demandé pour le service en question. Chacun des services est regroupé par leur type de connexion (bidirectionnel, unidirectionnel, diffusion point/multipoint).

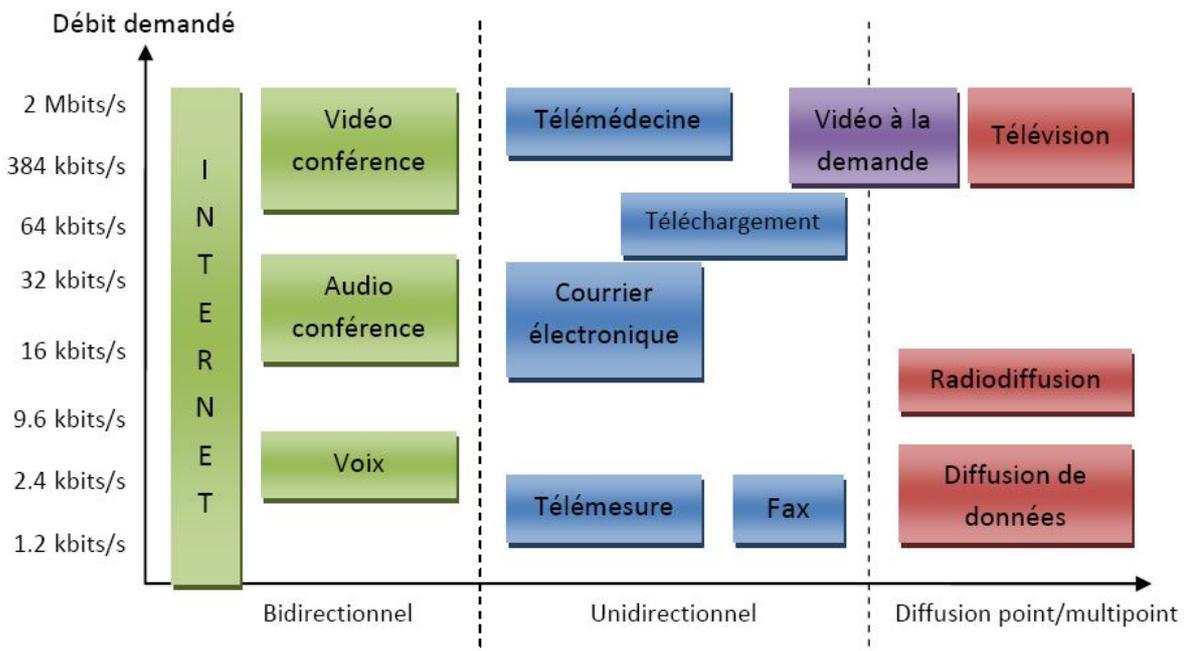


Fig I-2 :- Les services UMTS et leurs besoin en débit

### 3. Architecture du réseau UMTS

L'architecture du système UMTS est similaire à celle de la plupart des réseaux de deuxième génération. Le système UMTS est composé de différents éléments qui possèdent chacun leurs propres fonctionnalités. Il est possible de regrouper ces éléments de réseau en fonction de leurs fonctionnalités ou en fonction du sous réseau auquel ils appartiennent.

Les éléments du réseau du système UMTS sont répartis en deux groupes. Le premier groupe correspond au réseau d'accès radio (RAN, Radio Access Network ou UTRAN, UMTS

Terrestrial RAN) qui supporte toutes les fonctionnalités radio. Quant au deuxième groupe, il correspond au réseau cœur (CN, Core Network) qui est responsable de la commutation et du routage des communications (voix et données) vers les réseaux externes.

Les trois versions (releases) de l'architecture UMTS (R3, R4, R5) considèrent une même partie radio. Le réseau cœur est quant à lui différent d'une version à une autre. La Release 3 (Aussi appelée Release 99) des spécifications de l'UMTS, élaborée dans le cadre du projet de partenariat de 3ème génération (3GPP, 3rd Generation Partnership Project), a défini deux domaines pour la partie CN :

- ✓ Le domaine de commutation de circuits (CS, Circuit Switched),
- ✓ Le domaine de commutation de paquets (PS, Packet Switched).

L'architecture de référence du réseau CN est divisée en trois groupes. Le premier est celui du domaine CS comprenant les entités MSC, GMSC, et VLR. Le second est celui du domaine PS regroupant les entités SGSN et GGSN. Le dernier comprend les entités du réseau communes aux domaines PS et CS, à savoir, HLR, et AuC.

Pour compléter le système, on définit également le terminal utilisateur UE (User Equipment) qui se trouve entre l'utilisateur proprement dit et le réseau d'accès radio.

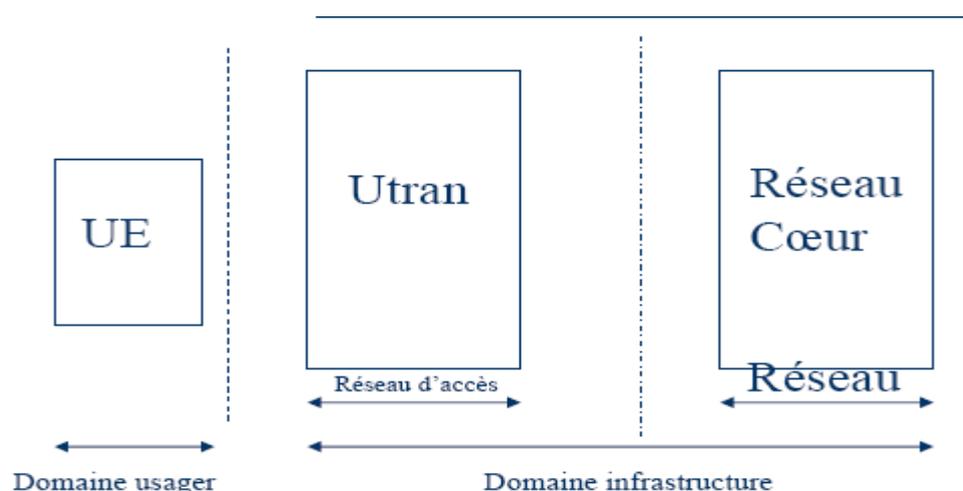


Fig I-3 : - Vue d'ensemble d'un réseau UMTS

**A. Le réseau d'accès radio UTRAN :** Sa fonction principale est de transférer les données générées par l'utilisateur. Il est une passerelle entre l'équipement usager et le réseau cœur via les interfaces Uu et Iu (explicitées plus bas). Cependant, il se charge de ces fonctions suivantes :

- Sécurité : Il permet la confidentialité et la protection des informations échangées par l'interface radio en utilisant des algorithmes de chiffrement et d'intégrité.
- Mobilité : Une estimation de la position géographique est possible à l'aide du réseau d'accès UTRAN.
- Gestion des ressources radio : Le réseau d'accès est chargé d'allouer et de maintenir des ressources radio nécessaires à la communication.
- Synchronisation : Il est aussi en charge du maintien de la base temps de référence des mobiles pour transmettre et recevoir des informations.

Le réseau d'accès UTRAN est composé de plusieurs éléments : une ou plusieurs stations de base (appelées NodeB), des contrôleurs radio RNC (Radio Network Controller) et des interfaces de communication entre les différents éléments du réseau UMTS.

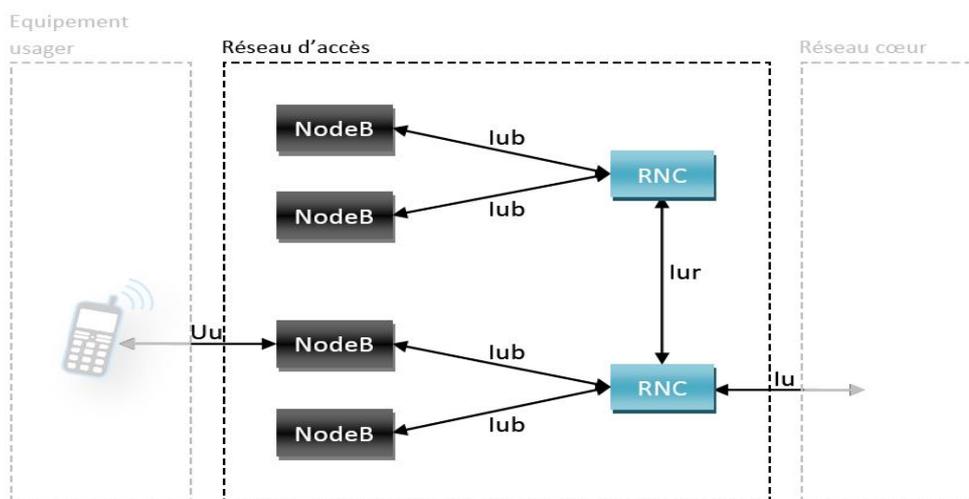


Fig I-4 : – Architecture du réseau d'accès

a) **Le Node B:** Le Node B est une antenne. Répartis géographiquement sur l'ensemble du territoire, les Nodes B sont au réseau UMTS ce que les BTS sont au réseau GSM. Le rôle principal du Node B est d'assurer les fonctions de réception et de transmission radio pour une ou plusieurs cellules du réseau d'accès de l'UMTS avec un équipement usager. Il travaille au niveau de la couche physique du modèle OSI (codage et décodage). Ils communiquent directement avec le mobile sous l'interface dénommée Uu. Il permet notamment le contrôle de puissance.

En effet, le node B va prélever quelques mesures sur le signal reçu et va envoyer une commande pour que l'UE adapte sa puissance. Cela permet une meilleure autonomie pour l'équipement usager, ainsi qu'une limitation des interférences dans une même cellule. Cette limitation des interférences est particulièrement importante car les performances du W-CDMA y sont directement liées.

b) **Le RNC (Radio Network Controller):** C'est la partie intelligente de l'UTRAN. Le rôle principal du RNC est de router les communications entre le Node B et le réseau cœur de l'UMTS. Il travaille au niveau des couches 2 et 3 du modèle OSI.

Le RNC assure les mécanismes de gestion de mobilité, à savoir le handover et la macro-diversité. Le handover est la capacité du réseau à maintenir une communication lorsqu'un mobile change de cellule. La macro-diversité est la phase pendant laquelle la station mobile maintient plusieurs liens radio avec des cellules différentes. Le W-CDMA utilise la macro-diversité pour obtenir un signal de meilleure qualité. Ainsi, lorsque la station mobile se situe à la limite des cellules, elle va sélectionner le signal de meilleure qualité parmi ceux reçus comme si elle n'avait qu'une seule connexion au réseau. L'avantage de la macro-diversité est que la transmission n'est pas interrompue lors du changement de cellule de l'utilisateur à la différence du handover.

Le RNC gère le handover et la macro-diversité à travers l'interface Iub (lorsqu'il s'agit d'un déplacement entre cellules de différents Node B sous le contrôle du même RNC), à travers l'interface Iur (lorsque les deux cellules sont contrôlés par des RNCs différents) ou à travers l'interface Iu (lorsque par exemple l'interface Iur est absente). Le mécanisme de macro-diversité est aussi appelé soft-handover alors que le handover est nommé hard-handover.

Deux rôles de RNC ont été introduits afin de gérer la macro-diversité et le handover : le Serving RNC et le Drift RNC (un RNC joue l'un ou l'autre des deux rôles pour une communication). Chaque communication met en œuvre un Serving RNC, et passe par 0, 1 ou plusieurs Drift RNC :

- Le Serving RNC, qui gère les connexions radios avec le mobile et sert de point de rattachement au réseau de base via l'interface lu. Il contrôle et exécute le handover.
- Le Drift RNC, sur ordre du Serving RNC, gère les ressources radios des Node B qui dépendent de lui. Il effectue la recombinaison des liens lorsque du fait de la macro-diversité plusieurs liens radios sont établis avec des Node B qui lui sont attachés. Il route les données utilisateur vers le Serving RNC dans le sens montant et vers les Node B dans le sens descendant.

**Exemple :** Lorsqu'une communication est établie par un équipement usager, une connexion de type RRC (Radio Resource Control) est établie entre celui-ci et un RNC du réseau d'accès UTRAN. Dans ce cas de figure, le RNC concerné est le SRNC. Si l'utilisateur se déplace dans le réseau, il est éventuellement amené à changer de cellule en cours de communication. Il est d'ailleurs possible que l'utilisateur change de NodeB vers un NodeB ne dépendant plus de son SRNC. Le RNC distant est appelé « drift RNC ». Il a pour fonction de router les données échangées entre le SRNC et l'équipement usager.

### c) Les interfaces de communication

Plusieurs types d'interfaces de communication coexistent au sein du réseau UMTS :

- **Uu** : Interface entre un équipement usager et le réseau d'accès UTRAN. Elle permet la communication avec l'UTRAN via la technologie CDMA.
- **lu** : Interface entre le réseau d'accès UTRAN et le réseau cœur de l'UMTS. Elle permet au contrôleur radio RNC de communiquer avec le SGSN.
- **lur** : Interface qui permet à deux contrôleurs radio RNC de communiquer.

- **Iub** : Interface qui permet la communication entre un Node B et un contrôleur radio RNC.

## B. Le réseau cœur CN (Core Network)

Le réseau cœur de l'UMTS est composé de trois parties dont deux domaines :

- Le domaine CS (Circuit Switched) qui permet la commutation de circuits.
- Le domaine PS (Packet Switched) qui permet la commutation de paquets.
- Les éléments communs aux domaines CS et PS.

Ces deux domaines permettent aux équipements usagers de pouvoir gérer simultanément une communication paquets et circuits. Ces domaines peuvent être considérés comme des domaines de service. Ce type d'architecture permet de pouvoir créer ultérieurement d'autres domaines de service. Le schéma représente l'architecture du réseau cœur de l'UMTS :

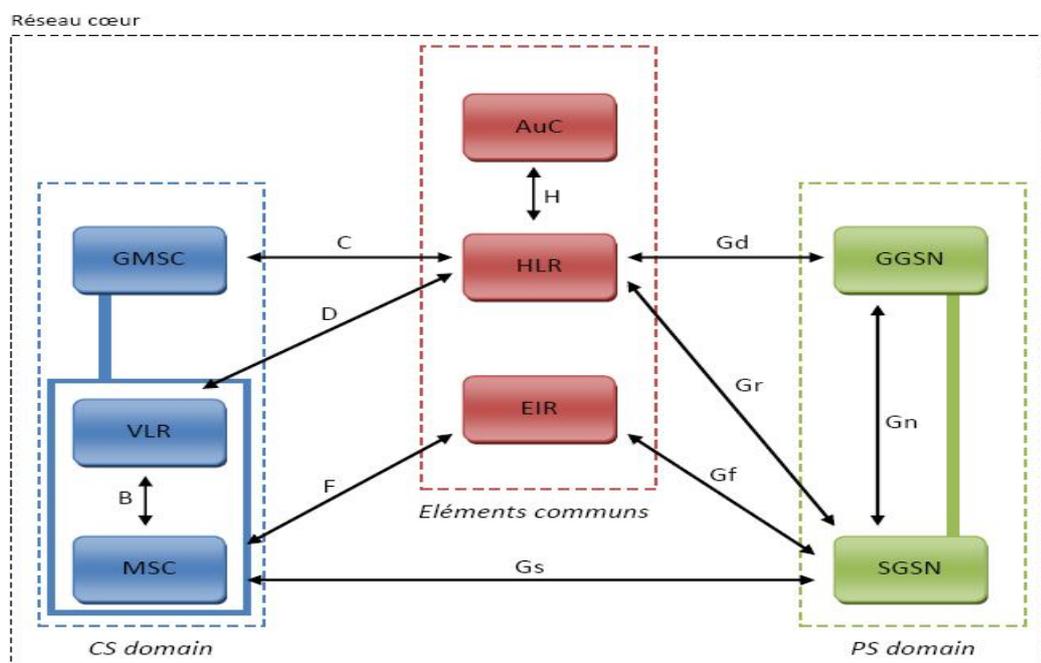


Fig I-5 : – Architecture du réseau cœur de l'UMTS

a) **Éléments communs**

Le groupe des éléments communs est composé de plusieurs modules :

- **Le HLR** (Home Location Register) représente une base de données des informations de l'utilisateur suivantes : L'identité de l'équipement usager, le numéro d'appel de l'utilisateur et les informations relatives à l'abonnement souscrit par l'utilisateur.
- **Le AuC** (Authentication Center) se charge de l'authentification de l'abonné, ainsi que du chiffrement de la communication. Si une de ces deux fonctions n'est pas respectée, la communication est rejetée. L'AuC se base sur le HLR afin de récupérer les informations relatives à l'utilisateur et pour ainsi créer une clé d'identification.
- **L'EIR** (Equipment Identity Register) se charge de la gestion des vols des équipements usagers. Il est en possession d'une liste des mobiles blacklistés par un numéro unique propre à chaque équipement usager, le numéro IMEI (International Mobile station Equipment Identity).

b) **Le domaine CS (Circuit Service Domain)** : Le domaine circuit permettra de gérer les services temps réels dédiés aux conversations téléphoniques (vidéo-téléphonie, jeux vidéo, applications multimédia). Ces applications nécessitent un temps de transfert rapide. Il est composé de plusieurs modules :

- **Le MSC** (Mobile-services Switching Center) se charge d'établir la communication avec l'équipement usager. Il a pour rôle de commuter les données.
- **Le GMSC** (Gateway MSC) est une passerelle entre le réseau UMTS et le réseau téléphonique commuté PSTN (Public Switched Telephone Network). Si un équipement usager contacte un autre équipement depuis un réseau extérieur au réseau UMTS, la communication passe par le GMSC qui interroge le HLR pour récupérer les informations de l'utilisateur. Ensuite, il route la communication vers le MSC dont dépend l'utilisateur destinataire.

- **Le VLR** (Visitor Location Register) est une base de données, assez similaire à celle du HLR, attachée à un ou plusieurs MSC. Le VLR garde en mémoire l'identité temporaire de l'équipement usager dans le but d'empêcher l'interception de l'identité d'un usager. Le VLR est en charge d'enregistrer les usagers dans une zone géographique LA (Location Area).

c) **Le domaine PS (Packet Service Domain):** Le domaine paquet permettra de gérer les services non temps réels. Il s'agit principalement de la navigation sur l'Internet, de la gestion de jeux en réseaux et de l'accès/utilisation des e-mails. Ces applications sont moins sensibles au temps de transfert, c'est la raison pour laquelle les données transiteront en mode paquet. Le débit du domaine paquet sera sept fois plus rapide que le mode circuit, environ 2Mbits/s. Il est composé de plusieurs modules :

- **Le SGSN** (Serving GPRS Support Node). Il maintient les informations identifiant l'abonné et les services utilisés. Il contrôle la localisation du mobile sur une Routing Area.
- **Le GGSN** (Gateway GPRS Support Node) est une passerelle vers les réseaux à commutation de paquets extérieurs tels que l'Internet.

### C. **Station Mobile (UE, User Equipment)**

L'utilisateur UMTS est équipé d'un UE (User Equipment) qui se compose du Mobile Equipment (ME) correspondant au combiné téléphonique (un terminal mobile) plus la carte USIM (UMTS Subscriber Identity Module). Le rôle de l'USIM est semblable à celui de la carte SIM en GSM. Elle enregistre les identités de l'abonné telles qu'IMSI (International Mobile Subscriber Identity), TMSI (Temporary Mobile Subscriber Identity), les données de souscription, la clé de sécurité (Ki) et les algorithmes d'authentification et de génération de clé de chiffrement. L'UE peut se rattacher simultanément aux domaines circuit et paquet et peut alors disposer simultanément d'un service UMTS et d'une communication téléphonique.

La figure ci-dessus représente l'architecture globale du réseau UMTS.

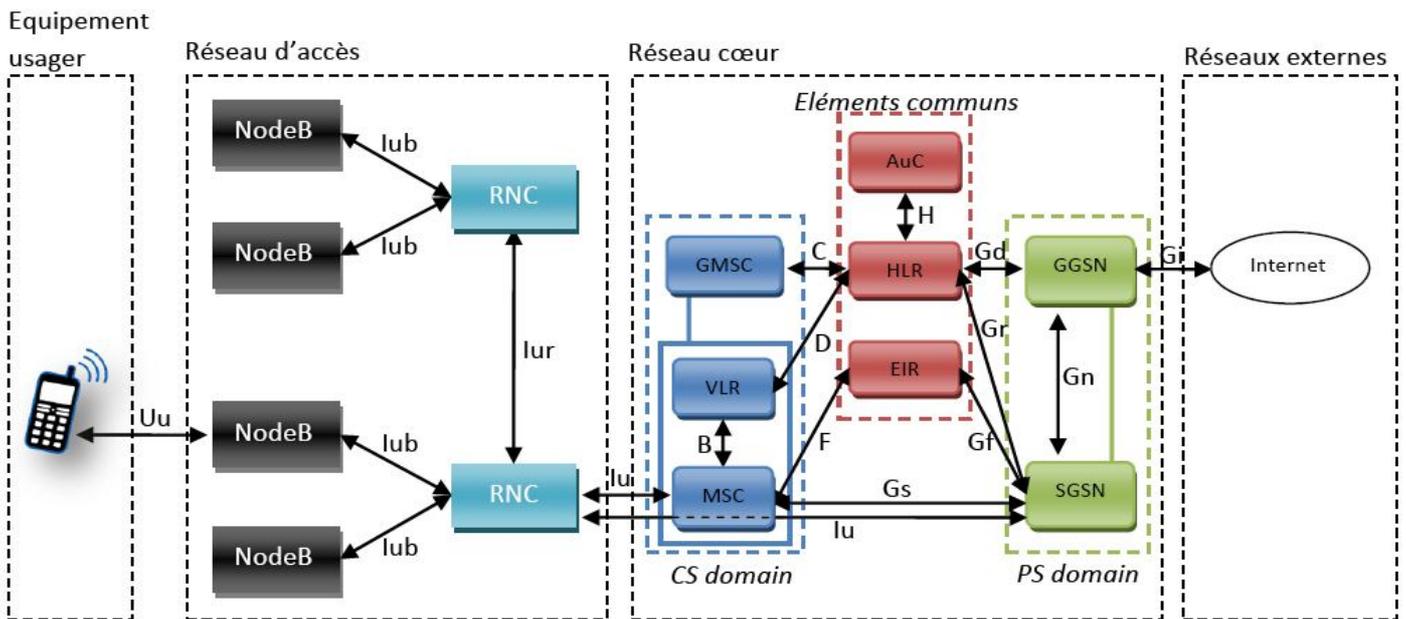


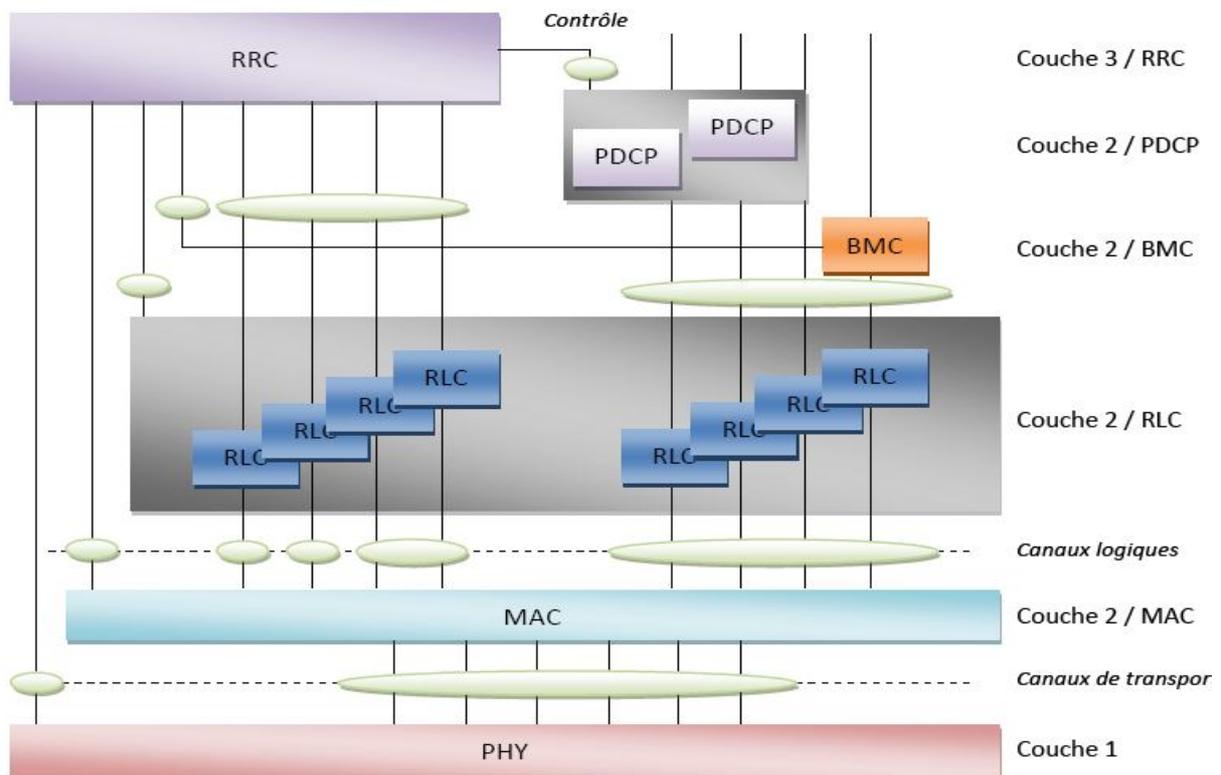
Fig I-6 : - Architecture globale du réseau UMTS

#### 4. Interface radio de l'UTRAN : Découpage en strates

Au cours de la modélisation du réseau UMTS, un découpage en strates (ou niveaux) a été introduit dans les spécifications du 3GPP. Ce découpage est conforme à celui du modèle OSI, permettant de séparer les niveaux de services indépendants. Ce découpage en niveaux correspond à une répartition logique des fonctions du réseau. Il y a deux grandes strates dans le réseau UMTS : AS (Strate d'accès) et NAS (Strate de non accès).

##### 1. AS : Access Stratum

L'Access stratum regroupe toutes les fonctions et les protocoles du réseau UMTS liées au réseau d'accès, dont, par exemple, les fonctions de gestion des ressources radio et de handover. Par définition, l'UTRAN est, en tant que réseau d'accès de l'UMTS, entièrement



inclus dans l'Access stratum. Par ailleurs, l'Access stratum comprend aussi une partie de l'équipement mobile (celle qui gère les protocoles de l'interface radio) ainsi qu'une partie du réseau coeur (correspondant à l'interface lu). Il est structurée en couches dont les protocoles se basent sur les 3 premières couches du modèle OSI (respectivement la couche physique, la couche liaison de données et la couche réseau).

Fig I-7 : - Vue en couches de l'interface radio UTRAN

➤ **Les canaux radio**

L'interface radio UTRAN doit être capable de transmettre des informations à haut débit et à faible temps de latence. En effet, les messages de signalisations UTRAN doivent être transmis le plus rapidement possible en utilisant le meilleur schéma de protection contre les erreurs, car ils sont critiques dans le cas d'une mobilité radio. Dans le but d'être flexible et de permettre différents schéma pour la transmission de données, les spécifications de l'UTRAN ont introduit plusieurs types de canaux :

- ✓ Les canaux **logiques** – ce qui est transmit.
- ✓ Les canaux de **transport** – comment est transmit.
- ✓ Les canaux **physiques**.

a. **Canaux logiques** : Les canaux logiques correspondent aux services de transfert de données offert par les protocoles des couches hautes de l'interface radio. Il y a uniquement deux types de canaux logiques : les canaux de contrôle, pour les transferts des informations du plan de contrôle et les canaux de trafic pour les transferts des données utilisateur du plan usager. Chacun des canaux de ces deux catégories correspond à un certain type de flux d'information.

Les canaux logiques de contrôles :

- **BCCH** (Broadcast Control Channel) : est un canal commun en voie descendante, utilisé par le réseau pour broadcaster les informations système de l'UTRAN à l'ensemble des terminaux présents dans une cellule radio. Ces informations sont utilisées par le terminal, par exemple pour connaître l'opérateur, pour avoir des informations sur la configuration des canaux commun de la cellule et comment accéder au réseau, etc.
- **PCCH** (Paging Control Channel) : est un canal commun en voie descendante qui transfert les informations de paging aux terminaux présente dans une cellule.
- **CCCH** (Common Control Channel) : est un canal commun bidirectionnel transportant les informations de signalisation. Il est utilisé lors de l'établissement d'une connexion RRC mais aussi lors des mises à jour de localisation. Ce canal est typiquement utilisé dans les premières phases de l'établissement de communication.
- **DCCH** (Dedicated control Channel) est un canal point-à-point bidirectionnel qui supporte les informations de contrôle entre un terminal donné et le réseau. Il supporte uniquement les signalisations RRC et NAS.

Les canaux logiques de trafic :

- DTCH (Dedicated Traffic Channel) : est un canal point-à-point bidirectionnel utilisé entre un terminal donné et le réseau. Il peut supporter la transmission des données utilisateurs après l'établissement d'une connexion RRC qui inclut les données elles mêmes et la signalisation de niveau application associée à ce flux de données.
  - MTCH (Multicast Traffic Channel) : est un canal de données point-à-multipoint pour la transmission de trafic de données du réseau à un ou plusieurs terminaux.
- b. Canaux de transport**: Les canaux de transport décrivent pourquoi et avec quelles données caractéristique sont transférés à travers l'interface radio. Par exemple, les canaux de transport décrivent comment les données sont protégées contre les erreurs de transmission, le type du canal, la protection CRC qui est utilisée, la taille des paquets de données envoyés sur l'interface radio, etc. Les canaux de transports sont classifiés en deux catégories :

Les canaux de transport communs :

- BCH (Broadcast Control Channel) permet en voie descendante le transport d'infos balise.
- PCH (Paging Chanel) utilisé pour le transport des messages de paging sur une ou plusieurs cellules.
- RACH (Random Access Channel): permet sur voie montante de demander un accès à une ressource radio mais aussi transporter de la signalisation en Uplink sans contrainte réel.
- CPCH : rôle proche du RACH utilisé qu'en mode RRC connecté il est aussi à accès aléatoire.
- FACH: Forward Access Channel: voie descendante permet le transport des signalisations de petites tailles, l'identité de l'utilisateur doit être indiqué.
- DSCH : Downlink Shared Channel canal utilisé en association avec un ou plusieurs canaux dédiés, **partagé** entre plusieurs utilisateurs il transporte des données de contrôle.

Le canal de transport dédié :

- DCH (Dedicated Channel) : Seul canal dédié, utilisé dans le sens montant ou descendant.

- c. **Canaux physique** : Les canaux physiques sont l'implémentation des canaux transport sur l'interface radio. Ils sont classifiés en deux catégories :

Les canaux physiques en uplink (voie montante):

- PRACH : canal physique supportant le RACH
- PCPCH : Physical Common Packet Channel, supporte le CPCH
- DPDCH : Transporte le DCH
- DPCCH : Canal associé à un ou plusieurs DPDCH

Les canaux physiques en downlink (voie descendante):

- DPCH : supporte le DCH, on trouve contrairement à la voie montante un seul canal pour les données utilisateurs et les données de contrôle.
- SCH : synchronization channel, il transporte les 2 codes (primary et secondary codes) qui permettent aux terminaux de se synchroniser en temps sur la cellule.
- CPICH : ce canal transmet des bits pilotes prédéfinis qui permettent ainsi d'estimer les conditions radios et de faire des relevés de mesures
- PCCPCH : (Primary Common Control Physical Channel) supporte le BCH il est unique par cellule.
- SCCPCH (Secondary Common Control Physical Channel) supporte un PCH et un ou plusieurs FACH.
- PICH : Paging Indicator Channel toujours associé à un SCCPCH il transporte les paging indicator.
- PDSCH, il supporte le canal DSCH toujours associé au DPCH.
- AICH Acquisition Indicator Channel, ce canal est associé au canal PRACH et transporte les informations d'acquisitions.
- AP/AICH canal identique à l'AICH il est toujours associé au PCPCH il est utilisé s'il faut s'assurer qu'un préambule a bien été reçu sur le canal RACH.
- CD/CA-ICH permet de détecter les problèmes de collisions sur les accès RACH.
- CSICH multiplexé avec les autres canaux de contrôle et d'accès, il transporte les informations de disponibilités des canaux PCPCH.

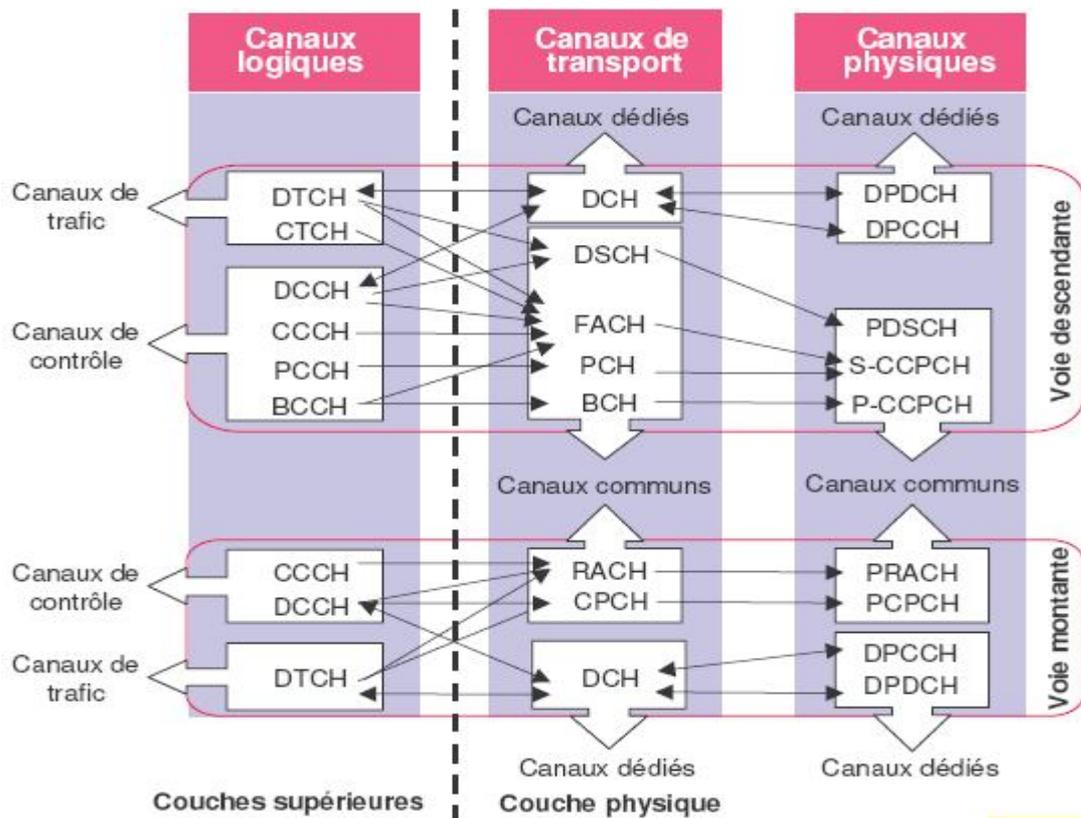


Fig I-8 :- Les différents canaux en UMTS

- **Couche 1** : Cette couche PHY représente la couche physique de l'interface radio qui réalise les fonctions de codage, décodage, modulation et d'entrelacement via W-CDMA, l'adaptation des débits, etc.
- **Couche 2** : Cette couche est divisée en plusieurs sous couches :
  - La sous-couche **MAC** (Medium Access Control) a pour rôle de multiplexer les données sur les canaux de transport radio.
  - La sous-couche **RLC** (Radio Link Control) permet la fiabilité du transport des données entre deux équipements du réseau. On distingue trois modes RLC :
    - Mode transparent : transport des données de manière transparente sans ajout d'informations de contrôle. Sa seule fonction est la segmentation/réassemblage.
    - Mode non acquitté : Utilisé dans le cas de streaming, il rajoute un contrôle d'erreur mais pas de retransmission.

- Mode acquitté : Ajout des processus de contrôle d'erreurs et de retransmission.

- La sous-couche **PDCP** (Packet Data Convergence Protocol) permet de compresser les données via des algorithmes de compression. Cela permet d'exploiter plus efficacement les ressources radio. De plus, cette sous-couche PDCP a pour rôle de rendre indépendant les protocoles radio du réseau d'accès UTRAN (sous-couches MAC et RLC) par rapport aux couches de transport réseau. Ce type d'architecture permettra l'évolution future des protocoles réseaux sans modifier les protocoles radio de l'UTRAN.

- La sous-couche **BMC** (Broadcast/Multicast Control) est en charge d'assurer les fonctions de diffusion de messages sur l'interface radio.

- **Couche 3** : Cette couche RRC (Radio Resource Control) gère la connexion de signalisation établie entre le réseau d'accès UTRAN et l'équipement usager, utilisée lors de l'établissement ou de la libération de la communication. Ces tâches principales sont :

- Gestion de la connexion RRC ;
- Gestion des états de service de RRC ;
- Gestion de la diffusion des informations systèmes générées par l'UTRAN ;
- Gestion de la localisation ;
- Sélection initiale/ resélection de cellule ;
- Gestion de la mobilité dans l'UTRAN ;

- Les états de services RRC : Les 2 modes opérationnels de base d'un UE sont le mode **non connecté (veille, idle)** et le mode **connecté**. Le mode connecté peut être divisé en états de service qui définissent quel type de canal physique est utilisé par l'UE.

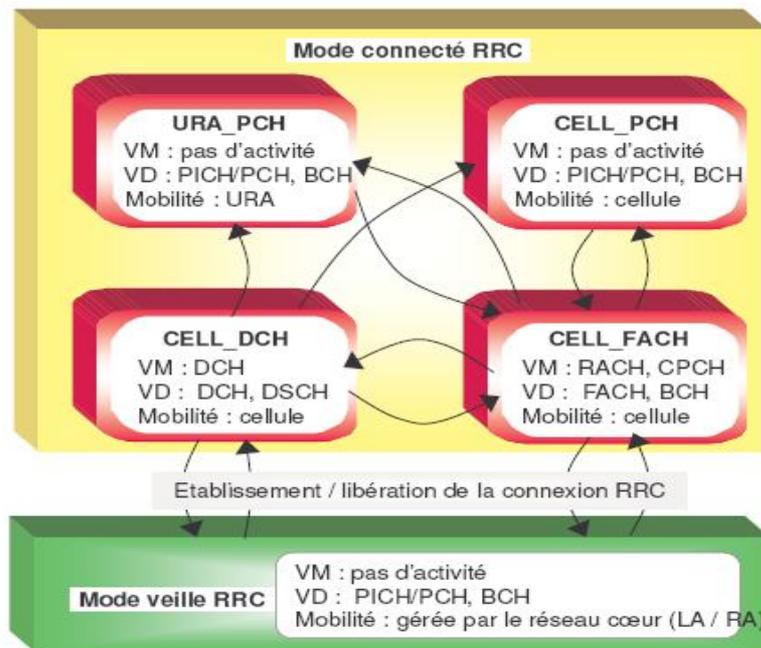


Fig I-9 :- Les différents états d'une station mobile

En mode veille, l'UE est réglé sur le canal de contrôle d'une cellule du réseau, en particulier une cellule couvrant l'UE. L'UE reste en mode veille jusqu'à ce qu'il transmette une requête d'établissement de connexion RRC. A ce moment là, il passe en mode connecté.

Dans l'état **Cell\_DCH**, un canal physique dédié (DCH et éventuellement DSCH) est alloué à l'UE et l'UE est connu par son SRNC.

Dans l'état **Cell\_FACH**, les canaux communs RACH et FACH sont utilisés pour transmettre des messages de signalisation et des données utilisateur. L'UE est également capable d'écouter le canal de diffusion (BCH) pour obtenir des informations système et le canal CPCH quand l'UTRAN demande une resélection de cellule.

Dans l'état **Cell\_PCH**, l'UE est toujours connu par son SRNC mais n'est joignable que via le canal d'appel (PCH). Si l'UE effectue une resélection de cellule, l'état de l'UE passe à Cell\_FACH pour exécuter la procédure « Cell Update » après quoi, l'UE repasse en état Cell\_PCH si aucune activité n'est survenue durant la procédure.

L'état **URA\_PCH** est similaire à l'état Cell\_PCH excepté que l'UE n'exécute pas la procédure « Cell Update » après chaque resélection de cellule mais lit son identité URA (UTRAN Registration Area). La procédure « URA Update » est similaire à « Cell Update ».

L'UE quitte le mode connecté et retourne en mode veille lorsque la connexion RRC est libérée ou lors d'un problème de connexion RRC.

- Etablissement d'une connexion RRC :

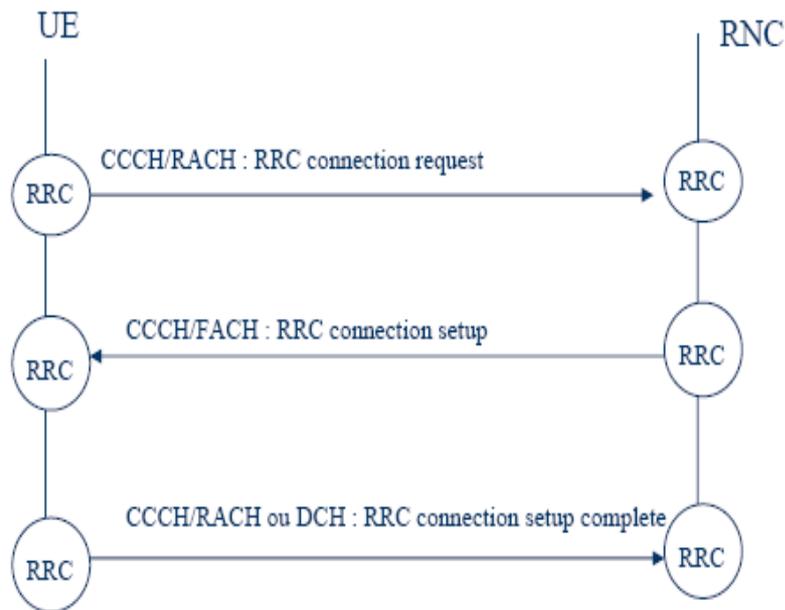


Fig I -10 :- Etablissement d'une connexion RRC

## 2. NAS : Non Access Stratum

Le niveau Non Access stratum regroupe toutes les autres fonctions et protocoles du réseau UMTS, qui permettent l'échange d'information entre l'équipement usager et le réseau cœur indépendamment du réseau d'accès radio utilisé, comme:

- o Les fonctions d'établissement d'appel : **CC (Call Control)** pour les appels circuit, et **SM (Session Management)** pour les appels paquet.
- o Les fonctions de gestion de la mobilité : **MM (Mobility Management)**
- o Les fonctions de gestion d'identité pour l'authentification et la sécurité : **ID management.**

## 5. Gestion de la mobilité

A la différence du réseau fixe, où le téléphone est toujours rattaché au même commutateur d'accès, l'UE de sa nature mobile, n'est pas toujours rattaché au même MSC. De plus, Les réseaux de communications mobiles sont constitués d'installations fixes et d'un grand nombre d'unités mobiles. Celles-ci peuvent se déplacer librement à l'intérieur de la surface, du réseau divisée en cellules, sans subir d'interruption de communications. C'est la raison pour laquelle le mobile doit régulièrement informer le réseau de sa localisation courante. Afin de réaliser cette action d'enregistrement, un mobile utilise un protocole de gestion de la mobilité (mobility management protocol, MM).

Principes de base de la gestion de mobilité : Dans la plupart des systèmes cellulaires, on distingue deux niveaux de gestion de la mobilité : la gestion de la mobilité radio et la gestion de mobilité réseau.

La gestion de la **mobilité radio** permet à une unité mobile de changer de cellule tout en maintenant sa communication avec le réseau (**handover**) ; elle permet aussi de localiser la cellule actuelle de l'unité mobile pour lui livrer un appel selon une opération dite le **paging** (la radiorecherche).

La gestion de la **mobilité réseau** permet à un abonné de pouvoir bénéficier, n'importe où et n'importe quand, des services auxquels il a souscrit (en particulier la réception et l'émission des appels) auprès de son réseau fournisseur.

La gestion de **mobilité réseau** intègre deux procédures de base : la **sélection/resélection** de cellules et la **gestion de localisation**. La sélection/resélection permet à l'unité mobile de recevoir les informations de la part du réseau et de se caler sur une cellule particulière, qui sera considérée comme la cellule d'accès en cas de nécessité d'établissement d'un lien radio avec le réseau. La gestion de localisation permet au système de connaître à tout moment la position courante de l'unité mobile.

Ces différents mécanismes constituent les principales fonctionnalités réservés aux systèmes de communications mobiles. Bien que la mobilité permet aux utilisateurs finaux d'être joignables partout et à tout moment, néanmoins, elle apporte des exigences strictes et complexes pour l'architecture de réseau cellulaire. Outre les conséquences dû à la mobilité

de l'UE sur le comportement des canaux de transmission, gardant l'appareil mobile relié au réseau et la gestion de sa mobilité est l'une des fonctions les plus difficiles, mais essentielles, de tout réseau cellulaire. Pour répondre à ces exigences, le réseau doit employer des stratégies efficaces de **localisation des unités mobiles**.

## II. La gestion de Localisation dans l'UMTS

Le rôle principal d'un mécanisme de gestion de la localisation est de permettre au système de connaître à tout instant la position d'un mobile. Cette fonction est nécessaire pour que le système puisse joindre un abonné. Dans la gestion de la localisation des mobiles, deux mécanismes de base interviennent :

- La **mise à jour de localisation (Location Update)** qui permet au mobile d'indiquer sa position dans le réseau zone à un moment donné;
- La recherche de l'abonné (ou **paging**) qui consiste à émettre des messages d'avis de recherche dans les cellules où le système a précédemment localisé l'abonné.

Ces deux mécanismes sont antagonistes dans la mesure où, lorsque la position du mobile est connue avec précision, le coût de la localisation est important alors que le coût d'une recherche éventuelle sera faible. A contrario, une connaissance imprécise de la position du mobile entraîne un coût de recherche élevé alors que le coût de la localisation aura été faible.

L'une des préoccupations des concepteurs de systèmes radio mobiles est de minimiser le coût des méthodes de gestion de localisation. En effet, elles engendrent beaucoup de signalisation entraînant une charge supplémentaire qui peut minimiser la qualité de service du réseau.

Les ressources de réseau, surtout les canaux radio, doivent être utilisées d'une façon optimale pour qu'elles desservent le plus grand nombre possible d'utilisateurs et ainsi augmenter la qualité de service du réseau. Pour optimiser l'utilisation des ressources, la plupart des réseaux mobiles adoptent une stratégie qui maintient le réseau informé des

mouvements des usagers. Cette stratégie consiste à regrouper plusieurs cellules du réseau dans ce qu'on appelle une **zone de localisation (LA : Location Area)** à l'intérieur duquel le mobile peut se déplacer sans se signaler au réseau. Lorsque le mobile entre dans une nouvelle zone de localisation, il le signale au réseau.

Ainsi, le système connaît la dernière zone de localisation dans laquelle l'abonné s'est signalé mais ignore la cellule précise où se trouve l'abonné. En cas de réception d'un appel, le système va rechercher l'abonné dans cette zone de localisation en émettant des avis de recherche (ou messages paging) dans les cellules de cette zone. Cette opération induit une réduction de la consommation des ressources. Cette technique de localisation nécessite automatiquement une mise à jour des informations de localisation des abonnés.

Cela dit, dès qu'une unité mobile entre dans une nouvelle zone de localisation, elle doit s'inscrire auprès de la base de données **VLR** propre à cette zone et rapporter au réseau sa localisation courante. Une base de données de visiteurs stocke les données sur tous les abonnés enregistrés dans la zone de localisation qui dépendent d'elle. Ces données consistent en une recopie partielle des éléments contenus dans la base de données HLR des abonnés et elles sont importées, soit directement à partir de la base de données HLR, soit à partir de l'ancienne base de données de visiteurs de l'unité mobile.

Les techniques de gestion de localisation consistent à attribuer à chaque zone de localisation une adresse unique ou, en d'autre terme, un numéro d'identification unique **LAI (Location Area Identifier)**. Les **Node B** des différentes cellules qui forment la zone diffusent cette adresse périodiquement aux unités mobiles qui se trouvent dans l'espace de couverture de la zone. L'unité mobile, qui garde normalement l'adresse de sa zone actuelle, compare le numéro reçu du Node B avec le numéro de sa zone actuelle. Si les deux numéros sont différents, c'est qu'elle a franchit la frontière de sa zone de localisation et se trouve dans une nouvelle zone. A ce moment, l'unité mobile doit faire connaître sa nouvelle zone de localisation au réseau qui pourra déterminer sa position courante en cas d'appel.

Une zone de localisation très grande garantit que les utilisateurs ne sont que rarement allés au-delà la limite de la zone de localisation. Cela réduit le besoin de signalisation des mises à jour l'emplacement. D'autre part, le nombre de cellules dans lesquelles les utilisateurs sont

recherchés par le paging durant une connexion est très élevé. Ainsi, les ressources radio sont nécessaires dans de nombreuses cellules, donc elles ne sont pas utilisées de manière productive.

Ces facteurs ont entraîné la nécessité de définir des zones distinctes pour les services de commutation de circuits et ceux de la commutation de paquets dans les zones de localisation. Contrairement aux services de commutation circuits, les utilisateurs des services de commutation de paquets reçoivent fréquemment des courts paquets de données. Par conséquent, un effort est nécessaire pour définir des zones de taille petite pour les services à commutation de paquets, alors que les grandes zones sont suffisantes pour les services à commutation de circuit.

Pour pallier ce problème, le concept de la zone de routage (**RA : Rounting Area**) a été introduit dans le cadre de transmission des données à commutation de paquets dans le GPRS puis dans l'UMTS. Le principe est le même que celui des zones de localisation, la seule différence étant que les zones de routage sont utilisés pour les Services du **domaine PS**. Dès qu'une station mobile quitte la zone de routage, elle exécute une **mise à jour de zone de routage** avec le **SGSN**. Ainsi, le SGSN est en mesure de localiser la station mobile avec plus de précision dans le cas des paquets entrants.

Les figures ci-dessus précises la division du réseau en zones de routage et en zones de localisation. La plus petite unité considérée ici est l'**UTRAN Registration Area (URA)**. Une URA est constituée d'une ou de plusieurs cellules qui sont réunies afin d'éviter un surplus de signalisation lors du changement de cellules sur les frontières de celles-ci. Une zone de routage se compose d'un ou plusieurs URA qui à son tour se compose d'un ou de plusieurs cellules.

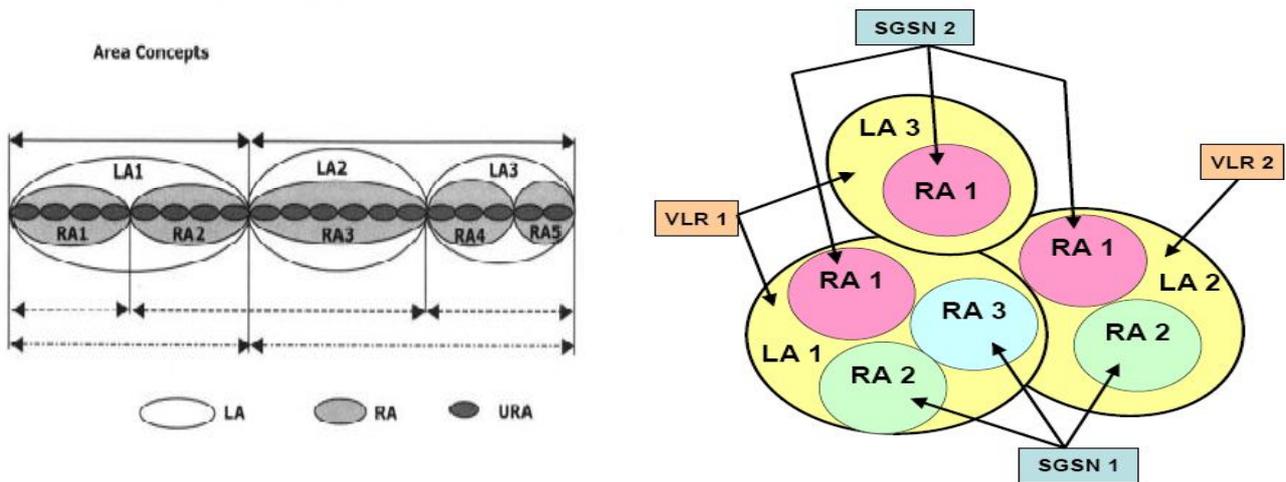


Fig I -11 :- Le découpage du réseau en zones géographiques

### ➤ Concepts généraux

**TMSI (Temporary Mobile Station Identify)** : A l'intérieur d'une zone gérée par un VLR, un abonné dispose d'une identité temporaire, le TMSI, attribué au mobile de façon locale, c'est à dire uniquement pour la zone gérée par le VLR courant du mobile. Le TMSI n'est connu que sur la partie MS-MSC/VLR et le HLR n'en a jamais connaissance. Le TMSI est utilisé pour identifier le mobile appelé ou appelant lors d'un établissement de communication.

Plusieurs mobiles dépendants de VLR différents peuvent avoir le même TMSI. A chaque changement de VLR, un nouveau TMSI doit être attribué. La structure du TMSI est laissée libre à l'opérateur. Il est codé sur 4 octets. Sa structure plus courte que l'IMSI permet de réduire la taille des messages d'appel sur la voie radio.

**La notion de PLMN :** (Public Land Mobile Network) : C'est un réseau de télécommunication constitué d'un réseau coeur et d'un réseau d'accès, installé et géré par un opérateur. Chaque PLMN dispose d'une identité, composée de deux champs, présent notamment dans l'IMSI :

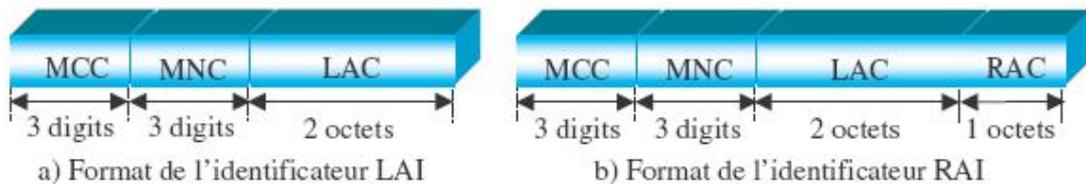
- **MCC** (Mobile Country code) : de 0 à 999 : est le code du pays du PLMN.
- **MNC** (Mobile Network Code) : de 0 à 99 permet de différencier les PLMN d'un même pays.

Pour répondre aux besoins des réseaux de troisième génération en termes d'indépendance entre opérateur de réseau et fournisseur de services, la notion de « PLMN équivalent » a été introduite.

La carte USIM de l'utilisateur n'est pas modifiée et reste attachée à l'opérateur auprès duquel l'abonnement a été souscrit. Par contre, le réseau a la possibilité d'indiquer au mobile lors de la procédure d'inscription au réseau, ou encore lors d'un changement de localisation, les PLMN que le mobile doit considérer comme équivalent lors de la sélection de cellule en mode veille. Ainsi, le mobile a la possibilité de passer d'un PLMN à l'autre comme s'il s'agissait d'un réseau unique.

La liste des PLMN équivalent est stockée dans la mémoire du terminal et est effacée lors de la mise sous tension.

#### LAI (Location Area Identifier) et RAI (Routing Area Identifier)



Avec :

LAC : Location Area Code : code de la zone de localisation librement affecté par l'opérateur (jusqu'à 2 octets au maximum).

RAC : Routing Area Code : code de la zone de routage (1 octet au maximum).

#### ➤ Attachement au réseau UMTS

L'UE peut se trouver dans 3 états de services différents : détaché, connecté ou inactif. Il est en état détaché quand il est éteint. Aucune communication avec le réseau n'est alors possible. L'UE ne peut rien envoyer ni recevoir. Pour pouvoir communiquer avec le réseau, l'UE doit s'attacher au réseau en l'allumant, en sélectionnant une cellule à laquelle s'attacher, et s'attacher à cette cellule. Quand l'UE est attachée au réseau, il passe en état connecté et peut débuter une communication, ou passer en état inactif s'il ne demande pas de communication. Par conséquent, l'attachement, qui est effectué uniquement si une carte

USIM valide est active dans l'UE, consiste en l'enregistrement de l'abonné (la carte USIM) au réseau afin d'accéder aux services fournis par le réseau auquel il est souscrit.

Le processus d'attachement au réseau débute quand l'utilisateur allume l'UE. Il doit alors entrer un code PIN (un numéro d'identification personnel) pour s'authentifier à l'USIM. Si l'authentification USIM est validée, l'UE cherche un Node B (dans une cellule ou secteur de cellule) auquel s'attacher. La procédure d'attachement est toujours initiée par l'UE. Quand l'UE trouve un Node B auquel s'attacher, il se synchronise avec lui, et tente ensuite de s'y attacher en envoyant une requête d'attachement au réseau, en particulier au RNC. Le réseau répond en envoyant l'identification USIM de l'UE au HLR pour l'informer de la requête d'attachement de l'UE.

Le HLR et l'USIM partagent une clé secrète de 128bits, que le HLR applique à un nombre aléatoire. Le résultat et le nombre aléatoire sont alors envoyés au réseau. Le réseau envoie le nombre aléatoire à l'UE. De façon semblable, l'USIM applique la clé secrète au nombre aléatoire, ensuite l'UE renvoie le résultat au réseau. Si le résultat envoyé par l'UE est identique au résultat envoyé par le HLR, le réseau accepte l'UE et l'attache au réseau. Enfin, le réseau récupère le profil de l'utilisateur depuis le HLR et le transmet au VLR associé au Node B d'attachement.

L'attachement est effectué en utilisant les deux procédures de signalisation du NAS : *IMSI attach* pour le domaine CS et *GPRS attach* pour le domaine PS.

### c) **La procédure IMSI attach**

La procédure IMSI attach est gérée par le protocole de gestion de mobilité MM de la station mobile. Ceci commence par envoyer le message LOCATION UPDATING REQUEST contenant, parmi d'autres, les paramètres suivants :

- Le type de la mise à jour de localisation ayant pour valeur dans ce cas *IMSI attach*.
- Le LAI (Location Area Identifier) déjà stocké dans l'USIM.
- L'identificateur de l'UE : IMSI ou TMSI
- Le paramètre *follow on request* utilisé pour demander au réseau de ne pas terminer la

connexion à la fin de la procédure lorsque il y'a d'autre signaux ou messages qui arrivent.

La figure ci-dessus illustre un exemple de la procédure d'attachement au réseau dans le CS domain. Dès que la VLR reçoit le message *LOCATION UPDATING REQUEST* avec le type de la mise à jour (IMSI attach), elle exécute les procédures d'authentification et d'intégrité. La VLR peut éventuellement procéder à l'identification de l'UE par son IMEI.

Dans le cas de succès de cette première étape, la VLR indiquera à la HLR la nouvelle localisation de l'UE avec l'IMSI et l'identificateur de la VLR.

La HLR envoie les données de l'abonné vers la VLR et reconnaît la requête de mise à jour. La VLR attribue ensuite un identifiant temporaire à l'UE et lui envoie un message *LOCATION UPDATING ACCEPT* qui pourrait contenir, en plus d'un TMSI et du LAI, le paramètre *follow on proceed* utilisé pour indiquer à l'UE que la connexion est maintenue et le paramètre *equivalent PLMN* contenant une liste des PLMNs équivalent.

Ensuite, la station mobile envoie à la VLR un message *TMSI REALLOCATION COMPLETE* afin de terminer la procédure.

Note: MAP (*Mobile Application Part*) protocole groupant les procédures de gestion des d'appel et de mobilité.

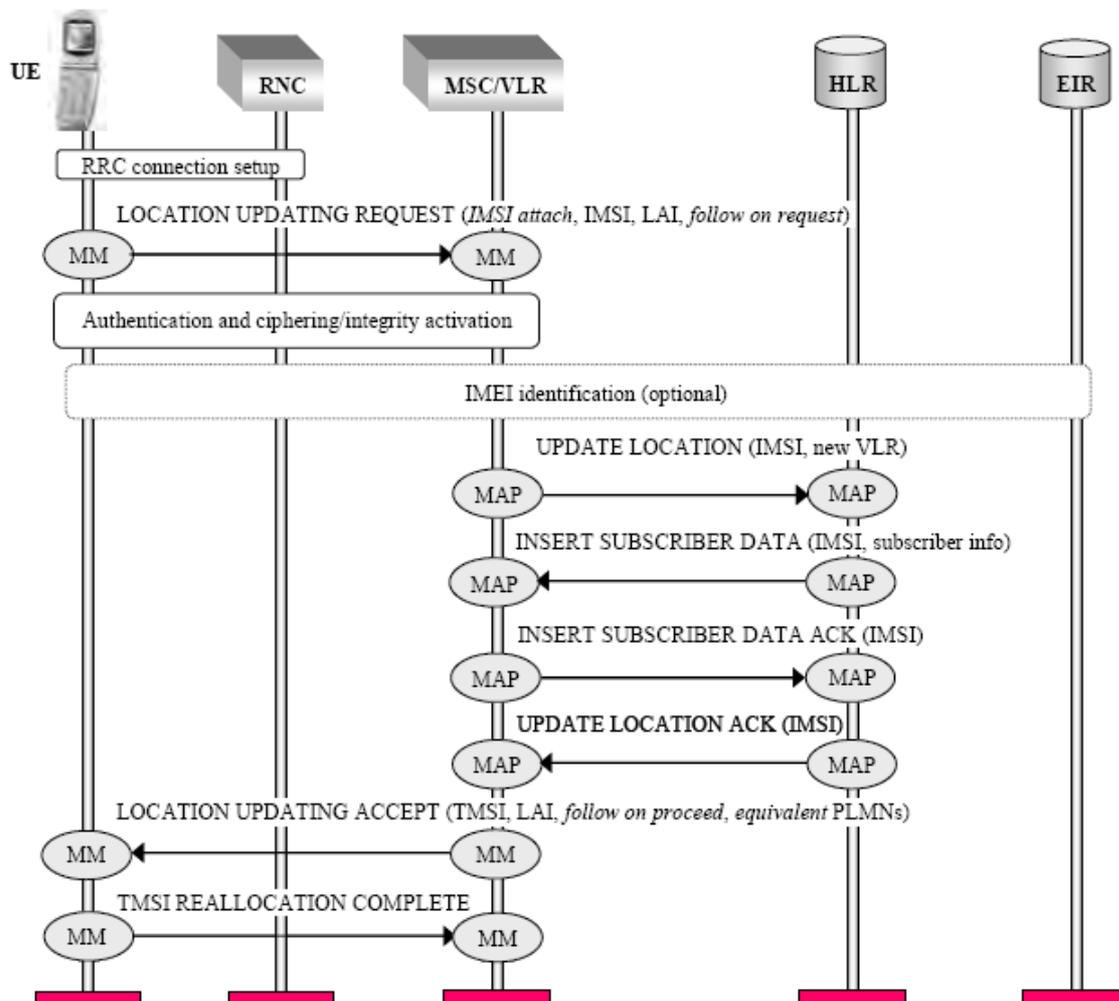


Fig I - 12: - La procédure IMSI Attach

#### d) La procédure GPRS attach

Cette procédure est utilisée pour s'attacher au domaine PS seulement (GPRS attach procedure for GPRS services) ou bien pour s'attacher aux deux domaines simultanément (Combined GPRS attach for GPRS and non-GPRS services). Le dernier cas est possible si l'interface Gs entre le SGSN et le MSC /VLR est implémentée dans le CN. La procédure est gérée et initiée par le protocole GMM (GPRS Mobility Management) de l'UE en envoyant au réseau un message ATTACH REQUEST avec les paramètres suivants :

- Le type de la requête d'attachement qui pourra prendre pour valeurs *GPRS attach*, pour un simple enregistrement au domaine PS, ou *GPRS attach while IMSI attached or combined GPRS/IMSI attach* pour un enregistrement au domaine PS alors que l'UE est

déjà enregistré au domaine CS ou bien pour un enregistrement simultané au deux domaine ;

- L'identificateur de l'UE qui pourrait être l'IMSI ou bien un P-TMSI (Packet-TMSI) ;
- La signature, si elle existe, associée au P-TMSI dans le cas où ce dernier est utilisé comme identificateur ;
- Le RAI associé au P-TMSI ;
- Le *TMSI status* utilisé dans le cas d'un attachement combiné pour indiquer qu'un TMSI valide n'est pas disponible ;
- Le paramètre *follow on request* ;

La figure ci-dessus illustre un cas générique d'un attachement combiné GPRS, prenant en compte un changement du SGSN et du MSC /VLR depuis le dernier détachement du réseau.

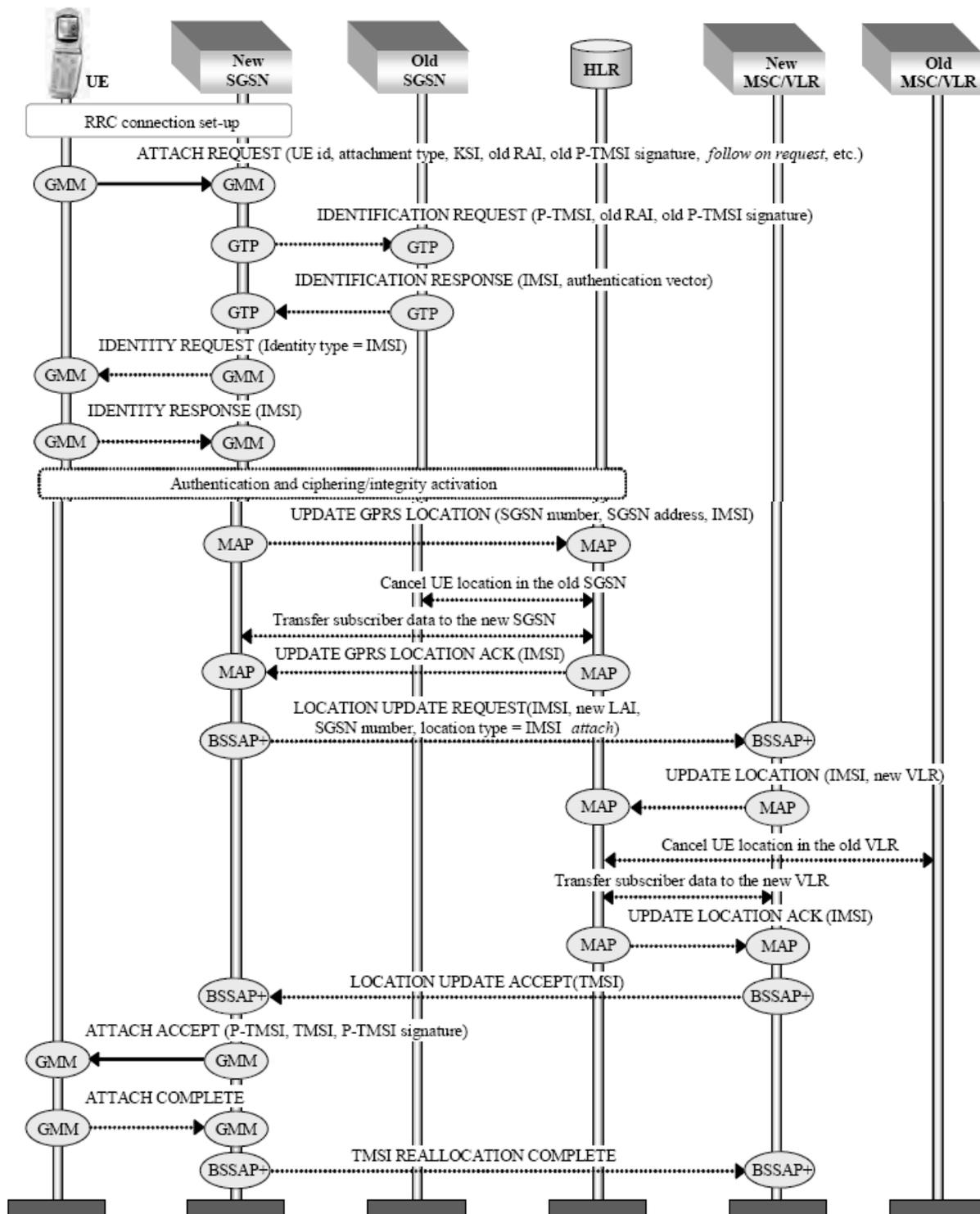


Fig I – 13 :- La procedure Combined GPRS/IMSI attach

➤ **La procédure de mise à jour de la localisation:**

Après l'attachement et l'inscription au réseau, incluant la mise à jour de localisation initiale, la station mobile doit signaler chaque changement de localisation au réseau afin qu'elle puisse être joignable à chaque instant. Ceci est possible grâce aux deux procédures de mise à jour : *location updating* (CS domain) et *routing area updating* (PS Domain).

c) Procédure de mise à jour de localisation

On distingue deux types de mise à jour de localisation dans le domaine à commutation de circuit : normale et périodique. Ces deux procédures, tout comme la procédure IMSI attach, sont initiés par l'UE en utilisant le message LOCATION UPDATING REQUEST ; l'élément location updating type est utilisé pour les différencier les uns les autres.

**La mise à jour normale** est effectuée à chaque détection de changement de zone de localisation de la part de la station mobile dans le BCH de la cellule courante, ou bien quand elle est informée par le réseau en réponse à une requête d'établissement d'une connexion de gestion de mobilité.

**La mise à jour périodique** consiste à envoyer suivant une période définie, la localisation de l'abonné. Cette opération est automatiquement effectuée par l'UE. La période est fixée par la valeur du compteur fournit par le réseau dans l'élément *CS domain specific NAS system information* transmis dans des messages propres à la connexion RRC. Le compteur peut prendre des valeurs de 1 à 25,5 heures par des pas de 6 minutes. La valeur 0 est aussi utilisée pour indiquer que la mise à jour périodique n'est pas utilisée.

La figure suivante montre un exemple de mise à jour de localisation inter MSC/VLR où les deux zones de localisation, à savoir l'ancienne et la nouvelle, sont contrôlées par des VLRs différentes. Comparée à l'exemple de la procédure IMSI attach, la procédure de mise à jour de localisation ajoute l'annulation de la localisation dans l'ancien MSC/VLR.

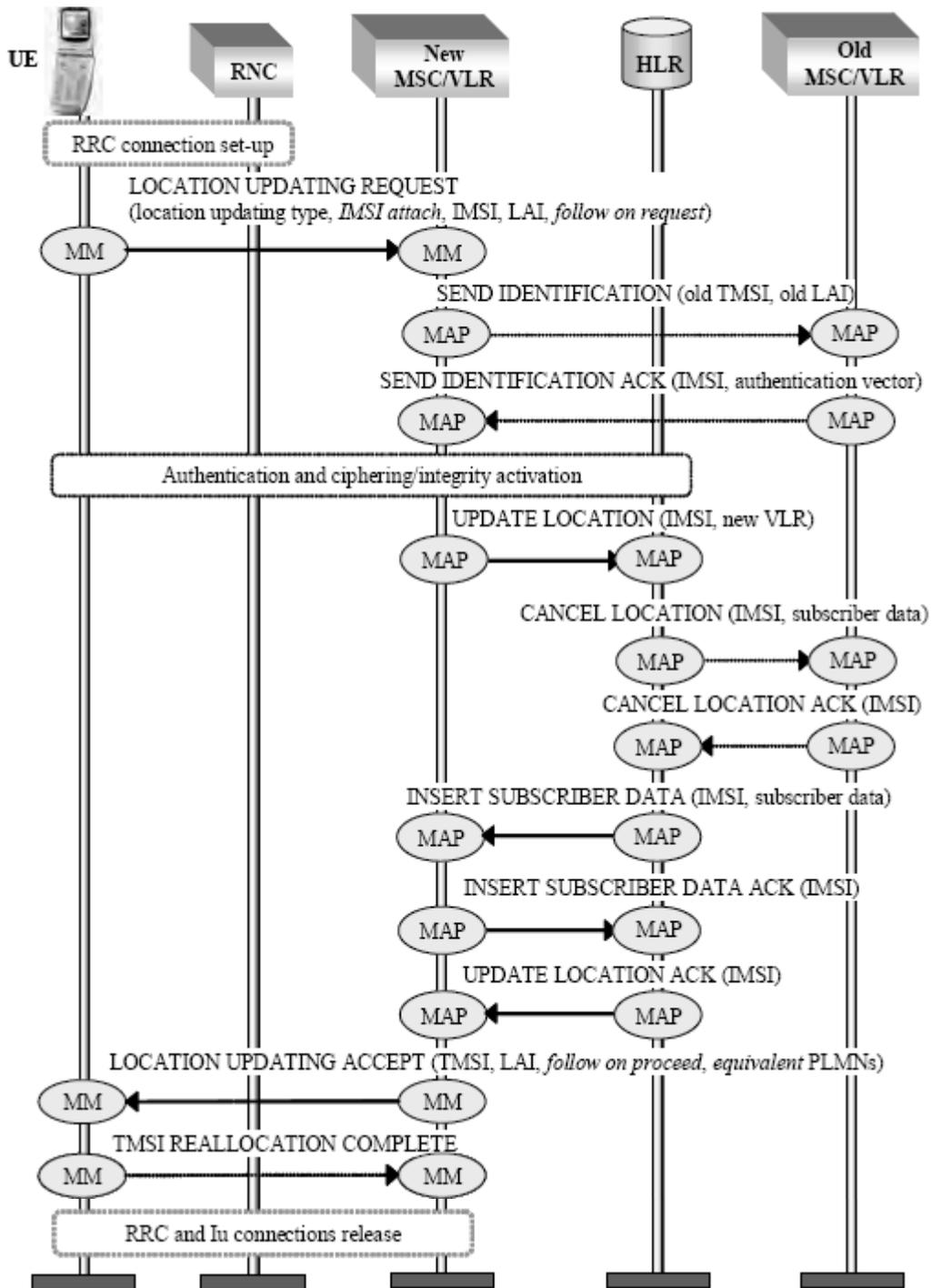


Fig I – 14 :- Mise à jour de localisation Inter MSC/VLR

d) Procédure de mise à jour de zone de routage

La procédure de mise à jour de zone de routage est utilisée pour mettre à jour la localisation de l'UE dans le domaine PS (RA Updating) mais aussi dans les deux domaines CS et PS (combined RA and LA updating). On peut distinguer plusieurs variantes de ces procédures :

- **La mise à jour normale** de la zone de routage initiée par l'UE à chaque détection de changement zone de routage RA, ou bien de rétablir la connexion dans le domaine PS quand la connexion RRC est libérée.
- **La mise à jour périodique** utilisée par l'UE afin de signaler régulièrement au réseau sa présence dans une zone de routage. Sa période est déterminée par la valeur du compteur fourni par le réseau dans les messages ATTACH ACCEPT ou ROUTING AREA UPDATE ACCEPT ;
- **La mise à jour combinée RA/LA** utilisée dans le cas où l'UE devrait effectuer une mise à jour de localisation alors qu'il est déjà inscrit dans les deux domaines CS et PS.
- **La mise à jour RA/LA combinée avec IMSI attach** utilisée dans le cas où l'UE, qui est attaché au domaine PS seulement, doit effectuer simultanément une mise à jour de zone de routage et un attachement au domaine CS (IMSI attach procedure).

#### ➤ **Le paging**

Le paging permet de rechercher un UE dans une zone de localisation afin de l'alerter d'un événement quelconque. Il fait partie du plan de contrôle radio du réseau et est livré à travers l'interface lu en utilisant le message RANAP PAGING. Ce message contient deux paramètres obligatoires, le CN qui a lancé la requête et l'IMSI.

Contrairement au GSM, l'UMTS manipule deux types de procédure de paging, PAGING TYPE 1 et PAGING TYPE 2. Ainsi, quand le réseau veut alerter la station mobile, il lui envoie via la connexion RRC un message RRC paging qui peut être de type 1 (message paging type 1) ou de type 2 (message paging type 2). Ceci dépend de l'état de la connexion RRC courante.

Le message *paging type 1*, transmis sur le canal logique PCCH, est utilisé dans le mode non connecté et les états CELL\_PCH et URA\_PCH du mode connecté pour alerter l'UE d'un appel entrant ou de l'informer sur les modifications systèmes de la cellule. Ce type de paging est toutefois utilisé par l'UTRAN quand il reçoit des données utilisateur sur voie descendante

adressées au UE se trouvant dans l'un des états CELL\_PCH ou URA\_PCH, comme indication de passer à l'état CELL\_FACH.

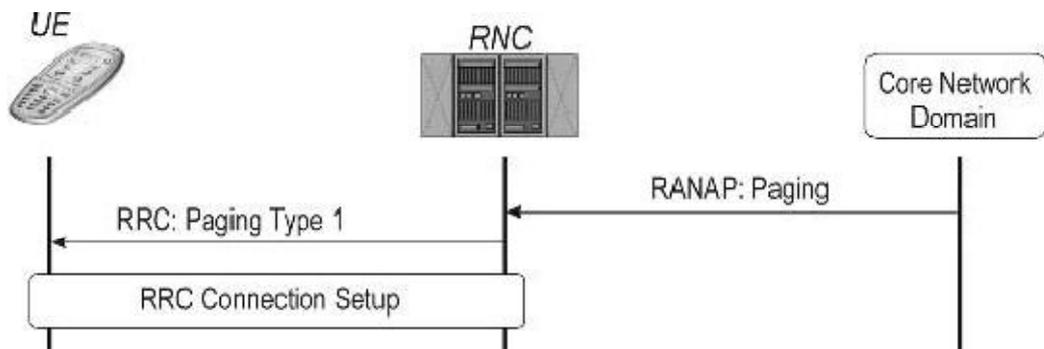


Fig I – 15 :- Paging type 1

Généralement, la réception d'un message Paging type 1 par l'UE mène à un établissement d'une connexion RRC nécessaire à la continuation de la procédure.

Le message Paging type 2, appelé aussi Dedicated Paging, est utilisé dans les deux états du mode connecté CELL\_FACH et CELL\_DCH pour alerter la station mobile d'un appel entrant. Il est transmis sur le canal logique dédié supporté par le canal de transport de type FACH ou DCH.

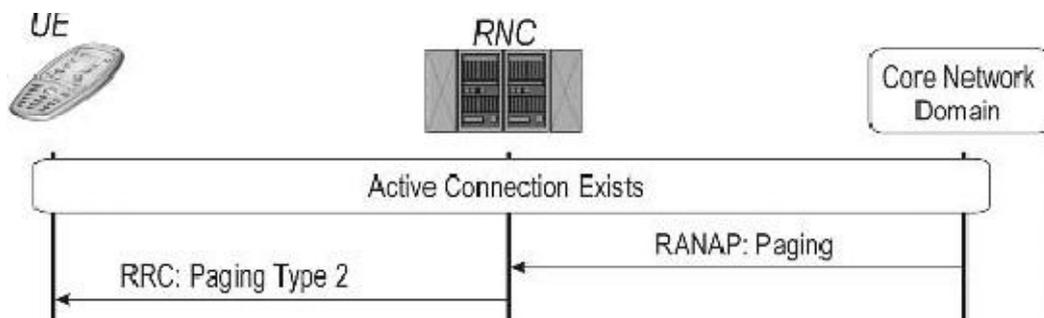


Fig I – 16 :- Paging type 2

Les terminaux 3G sont produits pour être en mesure de manipuler plusieurs connexion en même temps. Quand l'UE est déjà attaché à un domaine du CN et une autre connexion est établit avec ce même domaine, le CN envois alors un message Paging. Dans ce cas, le message paging envoyé à l'UE est de type 2.

Quand la couche RRC du UE reçoit le message paging :

- Si le paging a été enclenché par le sous réseau CN, RRC informe la NAS et le domaine concerné ;

- Si le paging avait pour raison d'alerter l'UE du changement des informations système, ce dernier lit le contenu du BCH histoire de mettre à jour ses valeurs ;
- Si le paging à pour but d'indiquer au UE de passer à l'état CELL\_FACH, la procédure de mise à jour de cellule est lancée.

Tout comme le GSM, l'approche DRX (Discontinuous Reception) est utilisée afin de minimiser la consommation d'énergie et prolonger la durée de vie de la batterie de l'UE. En effet, l'UE n'écoute qu'au canal PCH à travers ses groupes DRX et le réseau va dans ce cas là chercher (paging) la station dans ces groupes de canaux seulement.

# Chapitre II

## *La Gestion De Localisation*

**D**ans les systèmes de télécommunications mobiles, les abonnés entament et reçoivent les appels à l'aide de liaisons sans fil. Les unités mobiles se déplacent dans les zones définies par le système et qui correspondent aux localisations géographiquement délimitées. Quand un abonné entame un appel, l'infrastructure du réseau mobile doit effectuer de nombreuses tâches, parmi elles, la recherche de l'appelé et l'acheminement de l'appel au Node B qui dessert la zone géographique de l'appelé. Quand l'abonné A appelle l'abonné B, le problème de localisation consiste à rechercher la position de B dans un délai raisonnable, afin d'initialiser l'appel entre A et B. C'est ce qu'on appelle la localisation de l'abonné.

Dans ce chapitre, nous énonçons dans un premier temps les principes de la localisation des abonnés et la planification de zones de localisation. Par la suite, nous passons en revue les stratégies de mise à jour de localisation de l'unité mobile dans le réseau. Nous complétons le chapitre par une revue des schémas et des travaux proposés dans la littérature sur la gestion de localisation.

## 1) Principes de localisation

Comme c'est déjà indiqué au chapitre précédent, les schémas de gestion de localisation des usagers mobiles dans un réseau sans fil effectuent deux opérations principales : la **mise à jour** de localisation de l'unité mobile dans le réseau (update location) et la recherche de l'abonné mobile (**paging**). Les deux opérations permettent au réseau de suivre le déplacement de l'utilisateur afin de connaître sa position en tout temps. Ce suivi de déplacement par le réseau s'appelle donc la localisation des abonnés.

Il existe différentes stratégies de localisation des unités mobiles dans un réseau cellulaire. Elles sont divisées en deux groupes : les stratégies passives et les stratégies actives. Les stratégies passives combinent l'enregistrement de la position de l'abonné dans le réseau lors des mises à jour et le paging de l'unité lors d'un appel compte tenu de sa dernière position enregistrée. Les stratégies actives essaient de prévoir le mouvement de l'abonné selon l'historique de déplacement sauvegardé dans son profil d'utilisateur. Les deux types de stratégie peuvent être subdivisés en deux groupes : les stratégies basées sur l'utilisation des serveurs personnels, et les stratégies indépendantes du serveur personnel et permettant aux usagers d'avoir le numéro unique à vie sans tenir compte du lieu de son enregistrement dans le réseau.

La position actuelle du profil de chaque abonné est conservée dans sa base de données personnelle : La HLR. Quand l'abonné A appelle l'abonné B, l'infrastructure du réseau mobile envoie une requête à la base de données personnelle (HRL) de l'abonné B afin de déterminer sa position et d'initialiser l'appel ; Ceci est appelé : **interrogation**.

L'inconvénient de ce système est la lenteur du processus de consultation de la base de données éloignée. Pour pallier le grand délai de recherche de l'abonné appelé, le réseau maintient un registre de position de visiteur VLR (Visitor Location Register) dans chaque zone. Les VLR conservent les copies de profils des abonnés qui se trouvent actuellement dans les zones extérieures à leur zone personnelle. La recherche d'un abonné dans ce schéma HLR/VLR consiste à vérifier d'abord les registres VLR avant de consulter le HLR pour localiser l'appelé.

## ✓ Stratégies de mise à jour

Les stratégies de mise à jour de localisation de l'unité mobile dans le réseau peuvent être subdivisées en stratégies statiques et stratégies dynamiques [13]. Les stratégies statiques sont celles selon lesquelles le réseau mobile décide à quel moment l'unité mobile doit effectuer une mise à jour. Les stratégies dynamiques sont celles selon lesquelles l'utilisateur décide du moment de la mise à jour.

**Exemples de stratégies statiques** : Parmi les stratégies statiques, on distingue : la mise à jour basée sur le concept des zones de localisation, la mise à jour basée sur le concept de zones de localisation chevauchantes, et la mise à jour basée sur le concept des centres de notification.

**1. Mise à jour basée sur le concept des zones de localisation** : Selon cette stratégie, toutes les cellules du réseau sont combinées en groupes disjoints, appelés zones de localisation. L'unité mobile effectue une mise à jour de sa position seulement lorsqu'elle franchit la frontière d'une zone de localisation. Ainsi, la recherche de l'utilisateur s'effectue à l'intérieur de la dernière zone de localisation rapportée.

**2. Mise à jour basée sur le concept des zones de localisation chevauchantes** : Cette stratégie est semblable à la précédente, à l'exception que les cellules sont combinées dans des groupes non disjoints. Ceci permet d'éviter la surcharge du réseau lors des mises à jour effectuées pendant le mouvement proche des frontières des zones de localisation.

**3. Mise à jour basée sur le concept des centres de notification (reporting center)** : Selon cette stratégie, un sous-ensemble de cellules est désigné comme étant des centres de notification. L'utilisateur rapporte sa position dans le réseau seulement s'il entre dans un des centres de notification. La recherche de l'utilisateur s'effectue uniquement dans les cellules faisant partie du voisinage du centre de notification. Le voisinage du centre de notification est un ensemble de cellules situées autour du centre que l'utilisateur peut parcourir sans passer par un autre centre de notification.

Selon les trois stratégies statiques décrites, le réseau mobile décide à quel moment et à quel endroit les unités mobiles doivent effectuer une mise à jour. Par conséquent, toutes les unités mobiles effectuent la mise à jour de leur localisation au même endroit. La première stratégie basée sur les zones de localisation est la plus simple stratégie de mise à jour statique. Cependant, lorsqu'une unité mobile se déplace à proximité de la frontière entre les zones de localisation, cette stratégie présente un inconvénient majeur, car l'unité mobile risque d'inonder le réseau de messages de mise à jour à chaque fois que la frontière entre les zones de localisation est franchie. La stratégie basée sur les zones de localisation chevauchantes permet de remédier à ce problème. Selon cette stratégie, l'unité mobile peut appartenir en même temps à plusieurs zones de localisation, ce qui nécessite une stratégie efficace d'affectation de zones de localisation aux unités mobiles. La stratégie basée sur le concept de centres de notification est très semblable à la stratégie basée sur les zones de localisation chevauchantes. On peut considérer la stratégie basée sur les zones de localisation chevauchantes comme un cas particulier de la stratégie basée sur les centres de notification où les cellules constituant la frontière de la zone de localisation deviennent des centres de notification.

**Exemples de stratégies dynamiques :** Parmi les stratégies dynamiques, on distingue : la mise à jour basée sur le temps écoulé depuis la dernière mise à jour, la mise à jour basée sur le nombre de mouvements effectués, et la mise à jour basée sur la distance parcourue.

**1. Mise à jour basée sur le temps écoulé depuis la dernière mise à jour :** Selon cette stratégie, l'unité mobile rapporte sa position seulement après l'écoulement d'un certain laps de temps  $T$  depuis sa dernière mise à jour. La recherche de l'utilisateur s'effectue seulement dans l'ensemble de cellules que l'utilisateur peut atteindre dans le temps  $T$ , à partir du lieu de sa dernière mise à jour.

**2. Mise à jour basée sur le nombre de mouvements effectués :** Selon cette stratégie, chaque utilisateur mobile met à jour sa position dans le réseau seulement après avoir effectué un certain nombre de mouvements  $M$  entre les cellules. La recherche s'effectue dans l'ensemble des cellules que l'unité Mobile peut atteindre en  $M$  mouvements depuis sa dernière mise à jour.

**3. Mise à jour basée sur la distance parcourue** : Selon cette stratégie, l'unité mobile effectue la mise à jour de sa position dans le réseau après avoir parcouru une certaine distance  $D$  depuis sa dernière mise à jour. La recherche de l'unité mobile s'effectue dans l'ensemble des cellules situées à la distance  $D$  du dernier centre de notification.

Selon les trois stratégies dynamiques décrites précédemment, l'unité mobile doit décider elle-même où et quand effectuer une mise à jour de sa position dans le réseau.

En conséquence, les différentes unités mobiles peuvent effectuer une mise à jour à des endroits différents du réseau.

Le schéma de mise à jour basé sur le temps écoulé depuis la dernière mise à jour comporte deux désavantages majeurs :

1. Selon ce schéma, l'unité mobile est forcée de rapporter sa position dans le réseau, même si l'utilisateur n'a effectué aucun mouvement depuis sa dernière mise à jour. Ceci a pour effet de gaspiller inutilement les ressources du réseau.

2. Compte tenu que la vitesse de déplacement de l'unité mobile peut varier dans le temps, il devient difficile de cerner l'ensemble des cellules qui peuvent être atteints en un laps de temps donné. Il faut alors un algorithme capable d'estimer l'étendue de cellules que l'unité mobile peut rejoindre en un intervalle de temps donné, ce qui a pour effet d'augmenter la complexité de la procédure de paging.

Le schéma de mise à jour basé sur le nombre de mouvements effectués peut aussi engendrer une surcharge inutile du réseau dans le cas d'un mouvement répétitif entre deux zones avoisinantes ou d'un mouvement circulaire. En réalité, l'unité mobile ne change pas sa position effective dans le réseau, car elle revient toujours à son point de départ, mais elle est « forcée » d'aviser le réseau de son déplacement, créant ainsi un gaspillage de ressources.

Parmi tous les schémas dynamiques, le moins coûteux et le plus facile à gérer s'avère le schéma de mise à jour basé sur la distance parcourue. Selon ce schéma, l'unité mobile ne se rapporte qu'après avoir parcouru une certaine distance  $D$  depuis sa dernière mise à jour.

Ainsi, si l'unité mobile, après avoir effectué des mouvements consécutifs, demeure à une distance inférieure à  $D$ , elle n'a pas besoin d'engager des ressources du réseau pour effectuer la mise à jour de sa position. En ce qui concerne le paging, sachant qu'on est certain de retrouver l'unité mobile dans une étendue de cellules de rayon  $D$  depuis sa dernière mise à jour.

## **2) Planification des zones de localisation**

Dans les systèmes de téléphonie ordinaire fixes, le numéro de téléphone de l'abonné sert d'adresse de routage de son point d'attache au réseau global. Ce principe n'est pas applicable dans les réseaux mobiles, car les unités mobiles sont libres de se déplacer à travers le réseau, ce qui implique que leur point d'attache au réseau mobile change continuellement. Pour retrouver la localisation d'une unité mobile, on applique ce qu'on appelle une procédure de localisation qui produit à la sortie le numéro du Node B de la cellule dans laquelle se trouve l'unité mobile [13].

**Éléments de la problématique** : Dans les systèmes de télécommunications mobiles, la position de chaque unité mobile en communication est préservée en spécifiant sa zone de localisation. Après l'opération de localisation, le numéro d'identification de la zone de localisation de l'unité mobile est transmis à travers tout le réseau. Aussitôt que l'unité quitte une zone de localisation, on enclenche la procédure de mise à jour de localisation de l'unité. Cette procédure accède à la base de données distribuée à travers le réseau et comportant les informations sur la position de l'unité. À l'initiation d'un appel, cette information est retrouvée dans la base de données et le réseau déclenche une opération de paging qui vise à retrouver l'unité. Durant cette opération, les Node B installées dans la zone de localisation envoient, un à un, des signaux afin de déterminer le Node B qui dessert la cellule abritant l'unité mobile.

Il est évident que l'opération de paging surcharge le réseau proportionnellement à la taille de la zone de localisation. C'est pourquoi, dans les nouvelles générations des réseaux mobiles, chaque zone de localisation est composée d'un petit nombre de cellules. Cela ne cause aucun inconvénient tant que le nombre d'abonnés est restreint. Aussitôt que le nombre d'abonnés grandit considérablement, l'opération de mise à jour devient beaucoup

trop importante à cause du nombre d'unités qui se déplacent dans le réseau. Les stratégies de planification de la configuration des zones de localisation sont donc très importantes pour les réseaux mobiles de troisième génération.

La solution idéale consiste à trouver la configuration des zones de localisation qui tend à diminuer les deux surcharges du réseau : celle causée par le paging et celle due à la mise à jour de la zone de localisation. Il faut en fait trouver un équilibre entre les zones de localisation de grande taille qui provoquent la surcharge du réseau en effectuant le paging d'une unité, et les zones de localisation de petite taille qui causent la surcharge du réseau en effectuant les mises à jour lors du déplacement de l'unité mobile en communication entre les zones de localisation.

On peut donc noter deux configurations extrêmes :

- Toute l'étendue du réseau mobile équivaut à une zone de localisation. Aussitôt que l'unité mobile est appelée, la recherche se déclenche à travers tout le réseau, tandis qu'il n'y a pas de mise à jour à effectuer lors de son déplacement.
- La zone de localisation est formée d'une cellule du réseau mobile. Il n'y a pas besoin de localiser une unité mobile lorsqu'il est appelé, car sa position est connue. Dans ce cas, les mises à jour sont fréquentes, car le changement de cellule signifie en même temps le changement de zone de localisation.

Les bases de données distribuées jouent le rôle de dépôts de données sur les abonnés du réseau mobile. On distribue les données afin de décharger le poids de chaque nœud et pour rapprocher les données aux abonnés dans le but d'accélérer les transactions. Il existe donc un lien direct entre la configuration des zones de localisation et la structure des bases de données distribuées qui les desservent. On peut noter cinq modèles de planification de zones de localisation qui tiennent compte de l'optimisation du nombre de mises à jour par rapport au paging [13] :

**1. Modélisation basée sur les algorithmes heuristiques** : cette méthode vise à se rapprocher de la configuration optimale. En effet, l'approche heuristique effectue la collecte des

résultats de différents essais de configuration des zones de localisation du réseau mobile pour ensuite en retrouver une qui est proche de l'optimum. Le degré d'approximation de l'optimum dépend du nombre d'essais effectués. En général, une telle approche tend à minimiser les deux contraintes : le nombre de mises à jour et le nombre de paging.

2. Modélisation basée sur la topologie des zones géographiques et des autoroutes : les zones géographiques sont définies en fonction des critères géographiques (exemple : centre de la ville) et l'approche se base sur la distribution géographique de la population et sur les déplacements de la population à travers les autoroutes.

3. Modélisation basée sur le chevauchement des frontières des zones de localisation : cette méthode permet de diminuer le nombre de mises à jour causées par le mouvement des unités proches de la frontière entre les zones.

4. Modélisation basée sur l'heure de la journée : la configuration vise à faire concorder la mobilité des abonnés aux conditions de la route, variables dans le temps, avec la configuration des zones de localisation, qui devient alors dépendante de l'heure de la journée.

5. Modélisation basée sur la classification des abonnés mobiles : cette méthode considère le comportement de mobilité des abonnés et vise à regrouper les abonnés en fonction de leur mobilité. Ainsi, pour chaque groupe, l'on définit la configuration de zone de localisation.

### 3) Revue des stratégies et des travaux précédents sur la gestion de localisation

Grace à son importance, la gestion de localisation a réussi à faire tourner les regards des chercheurs et des scientifiques sur elle. Certaines de leurs propositions sont commentées dans cette partie. Une classification des différentes stratégies de gestion de localisation est donnée notamment. Les propositions sont évaluées en termes d'objectifs globaux de la gestion de localisation, à savoir:

- ✓ Minimiser le taux de signalisation, à la fois sur l'interface radio et à travers le réseau ;
- ✓ Minimiser le délai d'établissement d'appel (notamment en raison de procédure de paging), afin de maintenir qualité de service acceptable ;

- ✓ Minimiser la quantité de données stockées pour chaque abonné, afin réduire les coûts d'équipements ;
- ✓ Minimiser la complexité des algorithmes, afin de réduire les coûts de matériel, économiser la batterie, et d'améliorer les temps de réponse.

En général, ces objectifs sont contradictoires, et des compromis doivent être faits. Typiquement, les propositions se concentrent sur l'amélioration d'un aspect particulier, en compromettant l'autre. Pour les systèmes pratiques, un équilibre doit être trouvé, comme, par exemple, augmenter la complexité des calculs de stockage de données pour diminuer le taux de signalisation, et ainsi augmenter la capacité du système.

### **Classification des stratégies de gestion de localisation**

Dans la littérature, plusieurs propositions existent en ayant l'objectif d'améliorer certains aspects de la gestion de localisation. Certaines d'entre elles se concentrent sur l'amélioration de l'efficacité de signalisation et d'interrogation sur le réseau alors que les autres se focalisent sur la réduction de signalisation de paging.

Ainsi, les propositions de la gestion de localisation peuvent être classifiées selon deux catégories majeures : celles qui se focalisent sur l'aspect d'interrogation des bases de données et celles qui se focalisent sur les procédures de mise à jour et de paging [12]. Les deux catégories ne sont pas nécessairement exclusives et peuvent être combinées pour ne faire qu'une stratégie globale. La figure (II-1) résume cette classification des différentes stratégies de gestion de localisation.

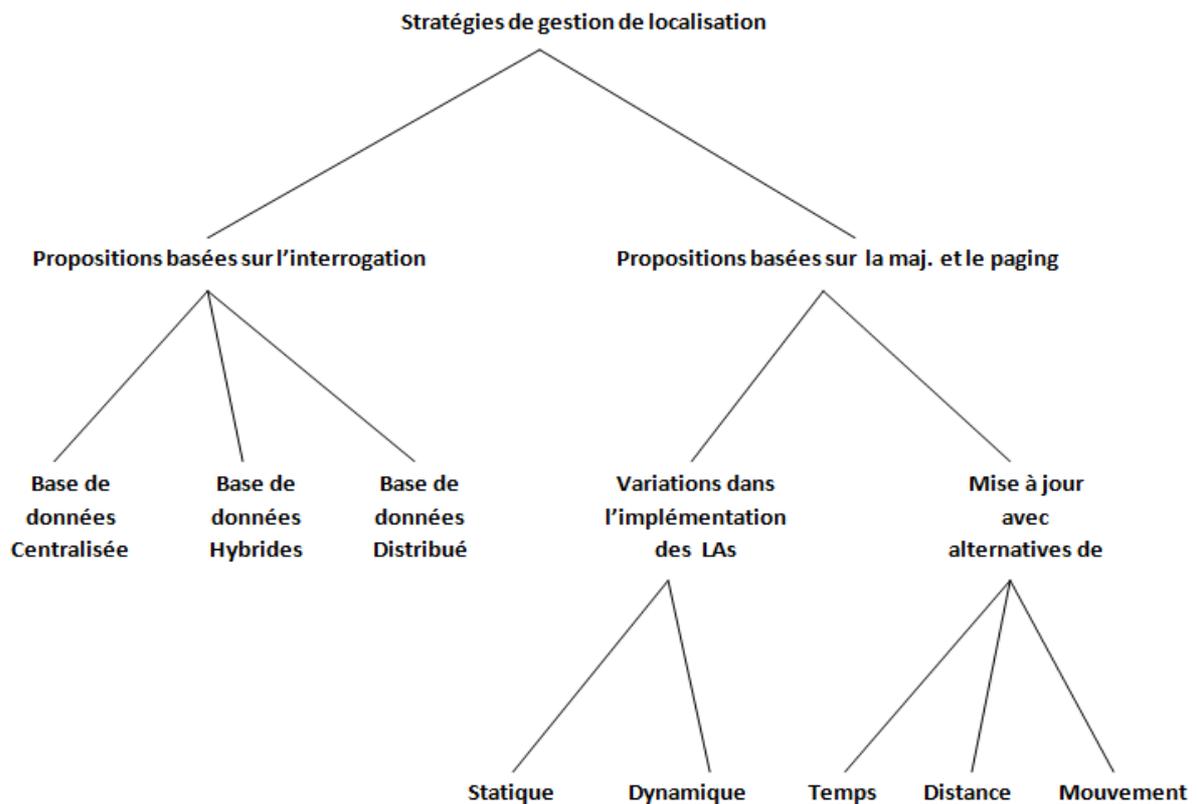


Fig II-1 : - Classification des stratégies de gestion de localisation

a) **Propositions basées sur l'interrogation (Interrogation-Based Proposals)**: Plusieurs propositions ont tenté d'améliorer l'efficacité de signalisation et d'interrogation à travers le réseau lors d'interaction avec les bases de données VLR et HLR, avec un rôle secondaire de réduire les coûts de signalisation dû au paging coté interface radio. La principale exigence est d'éliminer le goulot d'étranglement dû à une HLR centralisée, qui est interrogée à chaque appel entrant ou à chaque mise à jour de position associé à une unité mobile.

Pour répondre à ce problème, certaines propositions ont allé jusqu'à la modification de l'architecture des bases de données, pour ainsi la rendre distribuée ou encore un mélange des deux, à savoir, une architecture hybride. Cependant, un très grand nombre d'algorithmes ont été proposés et développés par plusieurs chercheurs utilisant des techniques aussi efficaces que nombrables, parfois combinés entre eux pour avoir un maximum de gain.

b) **Propositions basées sur la mise à jour et le paging (Location update and Paging proposals)**: La seconde approche servant à minimiser les coûts de la gestion de localisation consiste à réduire le nombre de messages émis lors des procédures de mise à jour et de paging via l'interface air.

Chaque unité mobile effectue une mise à jour de sa position lors de chaque changement de zone de localisation (LA). Quand un appel entrant arrive, le réseau tente de la localiser en envoyant un message paging simultanément à toutes les cellules de sa zone de localisation courante.

### 1. **Variation dans l'implémentation des LAs**

- **Static LAs**: Le schéma de mise à jour de localisation et de paging basé sur la zone de localisation est un schéma statique car il ne peut pas être ajusté périodiquement en fonction des paramètres de l'unité mobile. Autrement dit, les zones de localisation et de routage sont statiques, ce qui est un problème. La détermination de l'appartenance d'une cellule donnée à une zone donnée est faite par une procédure temporellement statique.
- **Dynamic LAs**: Des recherches récentes portent principalement sur des mécanismes dynamiques de mise à jour, qui effectue une mise à jour en se basant sur la mobilité de l'utilisateur et la fréquence des appels entrants. Plusieurs propositions suggèrent l'utilisation des profils utilisateur afin d'individualiser les zones de localisation. Une zone de localisation individuelle et personnelle, ancrée autour de la cellule où la mise à jour a survenu, consiste en les cellules les plus probables que l'abonné peut traverser. Les informations nécessaires à sa création sont tirées du profil utilisateur, qui contient un enregistrement des précédents chemins empruntés de cellule en cellule par l'abonné, ainsi que le temps passé dans chaque cellule. D'autre proposition regroupent les abonnés ayant le même comportement en termes de mobilité, alors que d'autres utilisent plusieurs niveaux de zones de localisation pour traquer les déplacements de l'unité mobile. La taille des zones à chaque niveau est différente. Chaque unité mobile est enregistrée à un niveau adapté, et ce niveau change dynamiquement selon le passé et la mobilité de l'abonné.

## 2. Mise à jour avec des alternatives de déclenchement

Toutes les stratégies précédemment décrites s'appuient sur le même mécanisme de mise à jour, à savoir, une procédure de mise à jour est appelée lors d'un chaque changement de zone de localisation ; le paging est toujours effectué à l'intérieur de la zone de localisation. D'autres schémas de mise à jour ont été proposés et qui n'exigent pas le découpage du réseau en zones de localisation. La seule information d'emplacement que possède le réseau sur l'abonné est son dernier point de contact avec lui, soit par le biais d'une mise à jour, d'une connexion ou d'une interaction avec le réseau. Ces méthodes de mise à jour non orienté zone de localisation sont essentiellement individualisées.

- Mise à jour basée sur le temps (Time – Based) : Comme expliquer plus haut, dans cette stratégie, la mise à jour s'effectue après l'écoulement d'un temps  $T$ . Peu importe que l'unité mobile soit recherchée par le système ou pas, le réseau tente de le localiser en premier lieu dans sa dernière position communiqué lors de la dernière mise à jour effectuée, disant la cellule  $i$ . S'il il n'est pas trouver, alors il est recherché dans la cellule  $(i+j)$  ou  $(i-j)$  pour  $j : [1,2 \dots N/2]$  tel que  $N =$  nombre total de cellules, commençant de  $j=1$ .
- Mise à jour basée sur le mouvement (Movement-Based) : Selon cette stratégie, chaque unité mobile compte le nombre de fois ou elle traverse une nouvelle cellule et effectue donc une mise à jour en arrivant à un certain nombre de mouvement déjà prédéfini par le réseau, disant  $M$ . Quand l'abonné est recherché par le paging, il pourrait être retrouvé à un voisinage de  $M$  cellules plus loin. Dans ce cas, chaque unité mobile doit garder le nombre de changement de cellule effectué après la dernière mise à jour.
- Mise à jour basée sur la distance (Distance-Based) : Dans cette troisième stratégie, l'unité mobile accomplit une mise à jour après avoir parcourue une distance prédéfinie  $D$ , partant de dernier l'emplacement mit à jour. Ce type d'algorithme a été utilisé dans les travaux menés par *Upamanyu Madhow, Michael L. Honig et Ken*

*Steiglitz* [28]. Les auteurs ont introduit un algorithme itératif qui génère un seuil optimal de la distance  $D$  pour un coût minimal de gestion de localisation. Le temps mis pour localiser un abonné est proportionnel à la distance parcourue depuis la dernière mise à jour. Les résultats ont démontré qu'en fonction de la mobilité et des paramètres d'appels entrants, le seuil optimal de mouvement varie largement.

**Comparaison** : Une étude menée dans [29] a comparé les trois différentes stratégies de mise à jour en ce qui concerne les coûts de mise à jour, de paging et le coût total de la gestion de localisation en développant un environnement orienté objet en langage java. Dans leur comparaison, les auteurs déclarent que la méthode basée sur la distance donne le moins de coût de mise à jour et de paging combiné.

La méthode basée sur le temps cause des mises à jour périodiques indépendamment de la mobilité de l'utilisateur, par conséquent, le coût de la mise à jour dans cette stratégie est plus stable que les deux autres stratégies. Cependant, le coût de paging s'accroît rapidement et proportionnellement avec le taux de mobilité, qui est définie comme le nombre moyen de cellules traversées par l'utilisateur en unité de temps.

Dans le cas des stratégies basées sur le mouvement et sur la distance, pour un taux minimal de mobilité, le coût de mise à jour dans ces stratégies est plus petit et le coût de paging est plus grand que ceux de la stratégie basée sur le temps. Alors que pour un taux maximale de mobilité, le coût de mise à jour est plus grand et le coût de paging est plus petit que ceux de la stratégie Time-Base.

#### **4) Conclusion**

Localiser un mobile dans un système cellulaire exige des services et des stratégies appropriés. Les communications radio mobile soulèvent deux problèmes majeurs. En premier lieu, l'amélioration sans cesse de l'efficacité des signalisations et des interrogations sur les bases de données. En second lieu, la réduction des coûts de la gestion de localisation en termes de mise à jour et de paging. Dans ce chapitre, on s'est intéressé au second point en commentant les différentes stratégies et algorithmes déjà proposés pour pallier à ce problème.

Comme déjà mentionné, chaque solution a ses avantages et ses inconvénients. Quelques unes sont relativement simples à implémenter, alors que les autres peuvent être intéressantes du côté théorique et conceptuel seulement. La plupart des propositions décrites dans ce chapitre ont été évaluées en se basant sur un modèle de mobilité aléatoire (A Random mobility model). Ce dernier ne prend pas en compte certains aspects d'un déplacement, comme le fait qu'il obéit à des règles en relation avec le profil de l'individu et le but du déplacement.

Par conséquent, il y'a comme un besoin d'analyser les performances des propositions en utilisant un modèle plus réaliste décrivant avec précision le mouvement et le comportement de l'utilisateur. Dans le chapitre 5, quelques problèmes associés à ce type de modèle sont décrits et le modèle de mobilité proposé par *Kunz* et *Scourias* est commenté et utilisé pour évaluer l'algorithme de gestion de localisation développé. Ce dernier est basé sur une technique appelée le Clustering ; une technique utilisée dans le Data Mining. A ce titre, le chapitre suivant a pour objectif d'expliquer et de mettre au clair ces deux mots.

# Chapitre III

## *Le Data Mining*

Les nouvelles technologies de l'information, facilitent le déplacement et le stockage des informations, et contribuent à leur croissance exponentielle. Les informations sont en nombre disproportionnées par rapport aux moyens humains pour les traiter, elles constituent des mines d'informations stratégiques aussi bien pour les décideurs que pour les chercheurs et les ingénieurs. Cependant, devant la masse croissante d'information, les organismes ont besoin de systèmes d'aide à l'analyse de plus en plus performants. Les technologies du **data mining** permettent, grâce à des techniques spécifiques tirées de plusieurs domaines, de traiter ces masses monstrueuses de données afin d'en extraire l'information cruciale, celle qui sera déterminante pour une prise de décision efficace.

L'extraction des connaissances à partir des données (ECD) : « Knowledge Discovery in Databases (KDD) », est un processus complexe désignant tout le cycle de découverte d'informations ou de connaissances dans les bases de données. Il regroupe donc toutes les opérations à effectuer pour extraire de l'information de ces données. Il existe une distinction précise entre le concept de KDD et celui de data mining. En effet, ce dernier n'est que l'une des étapes du processus de découverte de connaissances correspondant à l'extraction des connaissances à partir des données. Cette étape consiste à appliquer les analyses et les algorithmes de découverte aux données précédemment nettoyées.

Le data mining a aujourd'hui une grande importance économique du fait qu'elle permet d'optimiser la gestion des ressources (humaines et matérielles), ce qui fait d'elle une technique très appréciable pour les entreprises ayant à gérer une masse volumineuse de données. A ce titre, elle offre plusieurs techniques de fouille. Le choix de la technique appropriée dépend du domaine d'application, de la tâche à accomplir et des données disponibles pour l'analyse.

Dans ce qui suit, nous allons donner les généralités et les concepts fondamentaux du data mining ainsi qu'un aperçu sur ses processus d'analyse en mettant l'accent sur le clustering ; la technique qui nous intéresse le plus.

## 1) Définition du Data Mining

Le data mining (la fouille de données) est un ensemble de techniques et de méthodes du domaine des statistiques, des mathématiques et de l'informatique permettant l'obtention de connaissances exploitables à partir d'un important volume de données brutes.

Il s'agit de "fouilles" visant à découvrir "de l'information cachée" que les données renferment et que l'on découvre en recherchant des règles, des associations, des tendances nouvelles et imprévisibles, et des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données.

➤ **Intérêt du Data Mining** : Le Data Mining permet de réaliser des modèles rendant compte des relations liant la description d'une situation (sous forme d'un certain nombre de *variables descriptives*) à un résultat (ou un jugement) concernant cette description (la *variable prédite*). Ceci ramène parfois des trésors à la surface et l'avantage de posséder un tel modèle est double :

- ✓ Le modèle est descriptif et explicatif : il éclaire la situation en vous disant où regarder, et quelles sont les variables descriptives qui sont importantes pour déterminer le résultat.

- ✓ Le modèle est prédictif : si vous êtes en face d'une situation, décrite par les variables sur lesquelles vous avez basé votre modèle, le modèle prédictif vise à extrapoler et à extraire de nouvelles informations à partir des informations présentes.

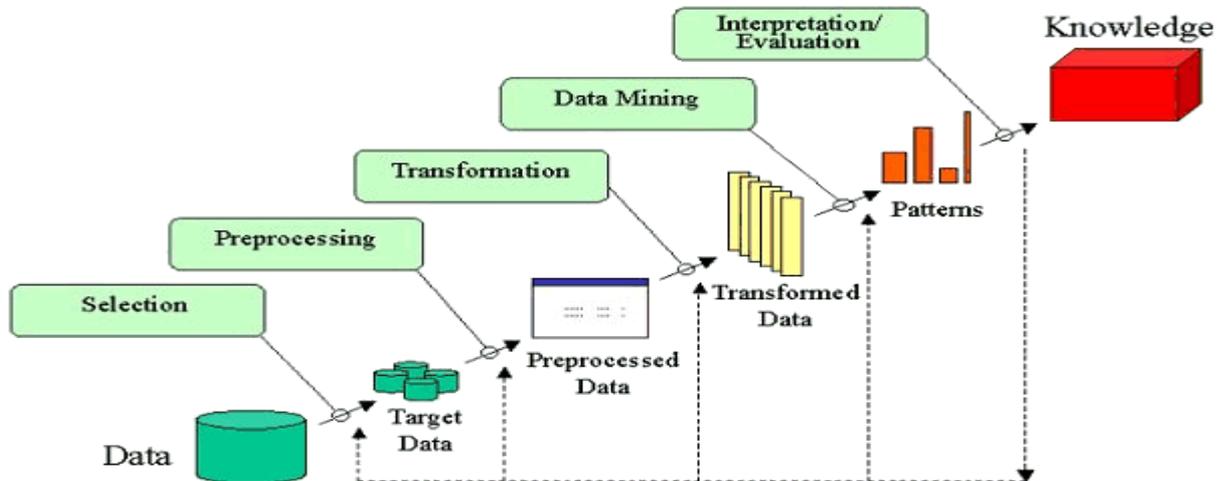


Fig III-1 – Le Data Mining Dans Le Processus KDD

## 2) Difficultés du data mining

La mise en œuvre du data mining rencontre trois difficultés principales :

- 1) **Qualité des données** : Les statistiques ont montré que 60% à 70% du temps de travail dans un projet de data mining est consacré au prétraitement des données (sélection, correction, transcodage, chargement...), ce qui montre que le temps de préparation est un inconvénient majeur qui influe sur le temps global du projet.
- 2) **Choix des algorithmes et de l'itinéraire du travail** : Pour pouvoir répondre aux questions qui se posent, les algorithmes doivent être choisis en fonction du problème traité.
- 3) **Evaluation des résultats** : Avant de procéder au déploiement final du modèle, il est important de l'évaluer plus complètement et de passer en revue toutes les différentes étapes exécutées pour construire ce modèle. Ceci permettra d'être certain qu'il permet d'atteindre les objectifs fixés. A la fin de la phase, une décision sur l'utilisation des résultats fournis par les outils de data mining devrait être prise.

### 3) Domaines d'application du data mining

Le data mining regroupe un ensemble de théories et d'algorithmes ouverts à tout domaine susceptible de drainer une masse importante de données. Les données peuvent être stockées dans des entrepôts (Data Warehouse), dans des bases de données distribuées ou sur Internet : web mining. Le data mining ne se limite pas au traitement des données structurées sous forme de tables numériques ; il offre des moyens pour aborder les corpus en langage naturel (text mining), les images (image mining), le son (sound mining) ou la vidéo et dans ce cas, on parle alors plus généralement de multimédia mining. Cela dit, on peut trier les domaines d'application par différents critères (secteur d'activité, type d'application, .....)

- Les assurances : Dans les assurances, il est surtout utilisé pour la détection de fraude et leur anticipation, après identification de ces dernières par un expert. Mais le data mining permet également d'extraire des caractéristiques à partir des déclarations (type d'accident, de blessures, etc.) permettant de définir des profils typés de clients et ainsi diminuer les risques dans les investissements.
- Secteur bancaire : L'objectif premier des banques est de réduire le risque de prêt bancaires, la création de modèle à partir de caractéristiques des clients permet de discriminer les clients à risque.
- Secteur industriel : L'analyse des habitudes du consommateur permet de prédire ses futurs achats, et de réorganiser les rayonnages de telle sorte à attirer l'attention du client sur d'autre produit susceptible de le séduire.
- Secteur médical : Dans le secteur médical, le datamining est naturellement répandu, tant dans les applications descriptives que prédictives. On trouve la détermination de groupes de patients susceptibles d'être soumis à des protocoles thérapeutiques déterminés. On a aussi les études sur les associations de médicaments, en vue notamment de détecter des anomalies de prescription.

- Télécommunications, eau et énergie : simulation des tarifs, détection du problème du churn des abonnés, détection des formes de consommation frauduleuses, classification des clients selon la forme de l'utilisation des services proposés.
- Secteur de l'éducation : Le data mining est employé dans les établissements scolaires pour améliorer la qualité d'enseignement. Par exemple, répartir les élèves ayant une grande capacité d'assimilation dans la même classe.

#### 4) Les méthodologies du Data Mining

La mise en œuvre de tout projet de Data Mining nécessite de suivre une méthodologie rigoureuse. De bonnes pratiques ont émergé au fil du temps pour améliorer la qualité de ces projets. Parmi celles-ci, les méthodologies aident les équipes à organiser les projets en processus. Au nombre des méthodes les plus utilisées on trouve la méthodologie SEMMA, la méthodologie DMAIC et la CRISP-DM qui est la méthode la plus employée.

a) La Méthodologie CRISP-DM : La méthodologie CRISP-DM (Cross Industry Standard Process for Data Mining) a été développée par un regroupement d'industriels européens, issus de divers marchés. Elle décrit le data mining en un processus itératif complet constitué de 6 principales phases.

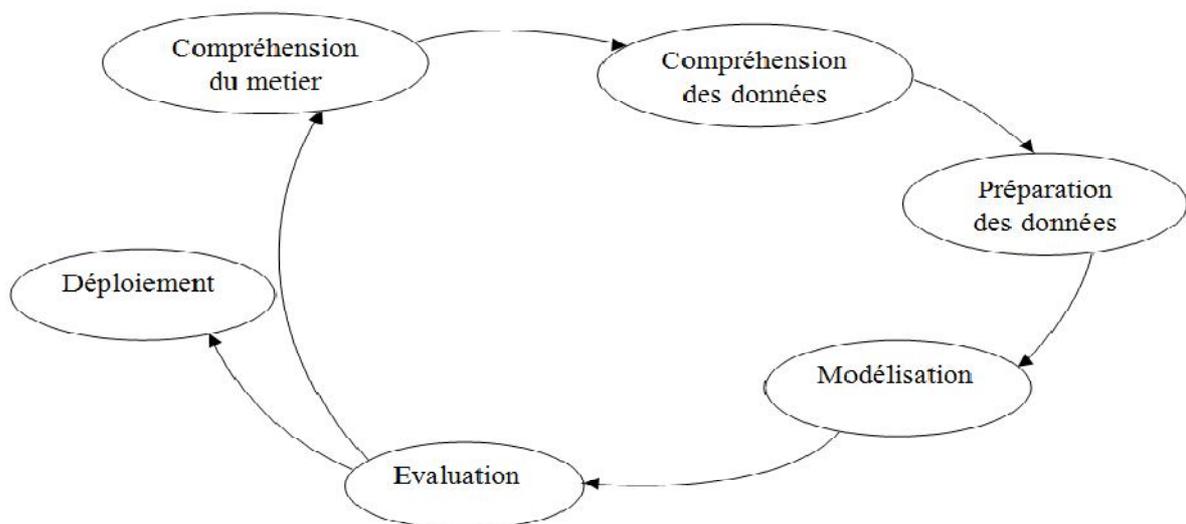


Fig III.2- : Les processus de la méthodologie CRISP-DM

- Comprendre le métier: C'est définir le problème et comprendre les enjeux : le chef de projet doit comprendre les objectifs du métier en matière de Data Mining et connaître les critères de réussite du projet.
- Comprendre les données : Dès que l'équipe de projet sait ce qu'il faut faire, elle doit se mettre en quête des données, des textes, et tout le matériel qui lui permettra de répondre au problème. Il lui faut ensuite en évaluer la qualité, et découvrir les premiers schémas apparents pour émettre des hypothèses sur les modèles cachés.
- Préparer : Le data mining se propose d'utiliser un ensemble d'algorithmes issus de disciplines scientifiques diverses (statistiques, intelligence artificielle, informatique) pour construire des modèles à partir des données, c'est-à-dire trouver des schémas intéressants ou motifs (des patterns) selon des critères fixés au préalable, et d'en extraire un maximum de connaissances utiles à l'entreprise. Les données que l'équipe de projet a collectées sont hétérogènes. Elles doivent être préparées en fonction des algorithmes utilisés, en supprimant les valeurs aberrantes - valeurs extrêmes -, en complétant les données non renseignées (par la moyenne par exemple), en supprimant les doublons, les variables invariantes et celles ayant trop de valeurs manquantes.
- Modéliser : Une fois que les données sont prêtes à être traitées il faut les explorer. L'activité de modélisation regroupe des classes de **tâches** pouvant être utilisées seules ou en complément avec les autres dans un but descriptif et/ou prédictif :

**La segmentation** - est la tâche consistant à découvrir des groupes et des structures au sein des données qui sont d'une certaine façon similaires, sans utiliser des structures connues à priori dans les données.

**La classification** - La classification consiste à examiner des caractéristiques d'un objet nouvellement présenté afin de l'affecter à une classe d'un ensemble prédéfini. Les algorithmes généralement utilisés incluent les arbres de décision, les k plus proches voisins, les réseaux de neurone, ...etc.

**La régression** - qui tente de trouver une fonction modélisant les données continues - non discrètes - avec le plus petit taux d'erreur, afin d'en prédire les valeurs futures.

**L'association** - C'est une tâche qui permet de découvrir les rapports de lien qui peuvent exister dans une base de données. Ces liens sont généralement exprimés sous la forme " A → B " qui signifie que la présence de A implique la présence de B (avec une certaine probabilité).

- **Évaluer** : Il s'agit d'évaluer les résultats obtenus en fonction des critères de succès du métier, d'évaluer le processus lui-même pour faire apparaître les manques et les étapes négligées. A la suite de ceci, il doit être décidé soit de déployer, soit d'itérer le processus en améliorant ce qui a été mal ou pas effectué.
- **Déployer** : C'est la phase de livraison et de bilan de fin de projet. On établit les plans de contrôle et de maintenance et on rédige le rapport de fin de projet.

**b) La Méthodologie SEMMA** : (Sample then Explore, Modify, Model, Assess pour), cette méthodologie est offerte par le logiciel SAS Enterprise Miner. Comme son nom l'indique, elle est constituée de cinq processus : Échantillonner, puis Explorer, Modifier, Modéliser, et enfin Évaluer.

**c) La Méthodologie DMAIC six Sigma** : (Define, Measure, Analyse, Improve, Control) conduite pour éliminer des défauts, pertes, ou problèmes de contrôle de qualité de toutes les sortes. Elle est constituée de cinq processus : Définir, Mesurer, Analyser, Améliorer et Contrôler.

Le cœur du data mining est constitué par la modélisation: toute la préparation est effectuée en fonction du modèle que l'analyste envisage de produire, les tâches effectuées ensuite valident le modèle choisi, le complètent et le déploient. La tâche la plus lourde de conséquences dans la modélisation consiste à déterminer le ou les algorithmes qui produiront le modèle attendu. La question importante est donc celle des critères qui permettent de choisir cet ou ces algorithmes.

## 5) Etat de l'art

Le data mining offre une très grande variété de techniques et d'algorithmes de fouille de données. Ces algorithmes ont des origines diverses ; Certains sont issus des statistiques (régression...), d'autres de l'intelligence artificielle (arbres de décision, ...), certains encore s'inspirent de la théorie de l'évolution (algorithmes génétiques...) ou de l'éthologie. Aucune de ces méthodes n'est optimale, mais elles sont complémentaires, et en les combinant intelligemment, il est possible d'obtenir de forts gains de performance. En effet, cette combinaison de technologies facilite la résolution, la compréhension, la modélisation et l'anticipation des problèmes. Le choix d'un algorithme particulier dépend du domaine d'analyse et d'application du processus du data mining mais aussi de la tâche choisie.

L'une des tâches du data mining la plus étudiée et la plus utilisée de nos jours est la classification. Qu'elle soit supervisée ou non, elle permet d'avoir une vue synthétisée des données traitées. Elle comporte un recueil d'algorithmes. Dans ce qui suit nous allons tenter d'avoir une vue sur l'ensemble de ces méthodes de classification, plus précisément aux méthodes dites non supervisées, en étant le plus représentatif possible afin de bien cerner les avantages et les inconvénients de chaque approche.

### ➤ La Classification

La classification est une tâche appliquée dans la vie courante. Elle est utilisée pour expliquer les nouveaux phénomènes rencontrés en les comparant avec des concepts et des phénomènes connus et en essayant de rapprocher les caractéristiques le plus possible. C'est un sujet de recherche actif qui se place au cœur de data mining, ce qui justifie pleinement l'intérêt qui lui est porté.

La classification est une opération de découpage d'un ensemble de données en un certain nombre de classes. Les méthodes de classification ont pour but de regrouper les éléments d'un ensemble  $X$ , de nature quelconque, en un nombre restreint de classes. La qualité de la classification peut être jugée sur la base des deux critères suivants :

- les classes générées doivent être les plus différentes possibles les unes des autres visa-

vis de certaines caractéristiques.

- chaque classe doit être la plus homogène possible vis-à-vis de ces caractéristiques.

De manière générale, les problèmes de classification s'attachent à déterminer des procédures permettant d'associer une classe à un objet (individu). Ces problèmes se déclinent essentiellement en deux variantes : la classification dite « **supervisée** » et la classification dite « **non supervisée** ».

1) **Classification supervisée de données** : Dans l'approche de la classification supervisée, les classes existent au préalable. Plus précisément, on dispose en données du problème d'un ensemble de classes et d'objets, chacun d'eux étant déjà placé dans une classe qui lui convient au mieux. On dit alors que les objets sont étiquetés. Cet ensemble de classes d'objets est souvent nommé « ensemble d'apprentissage ».

Le principe de la classification supervisée est de collecter le plus d'informations possibles à partir des groupements connus. Ceci vise à assurer une bonne connaissance des classes et à comprendre, en quelque sorte, pourquoi deux objets distincts sont séparés dans des groupes différents, ou au contraire rassemblés dans un même groupe. Le but ultime de cette démarche est de pouvoir allouer une classe à un nouvel objet (non initialement compris dans le jeu de données), en restant le plus cohérent possible avec la structure initiale en classes. Afin de s'assurer de ceci, différentes mesures de qualité de la classification sont disponibles. Indépendamment de la qualité obtenue pour un classement, on comprend assez facilement que cette classification est à revoir lorsque le nombre d'objets devient important. En effet, imaginons simplement que le nombre de nouveaux objets dépasse largement celui de l'ensemble d'apprentissage initial, il est évident que de nouvelles informations, plus proches de la réalité actuelle, pourraient être obtenues par une nouvelle étude des classes. Ceci exprime la nécessité d'actualisation de la classification, afin d'assurer sa cohérence. Parmi les algorithmes de classification supervisée les plus populaires, on trouve : les arbres de décision, et les réseaux de neurones.

## 2) Classification non supervisée de données (Clustering)

Dans l'approche du clustering, l'objectif principal est d'obtenir une représentation simplifiée de données initiales en les regroupant en classe homogène (groupe, cluster). Au départ, on dispose d'un ensemble de données qui reprend une collection d'objets non étiquetés. Les classes sont encore inexistantes. L'objectif est alors d'obtenir des clusters d'objets homogènes, en favorisant l'hétérogénéité entre ces différents groupes. Au final, l'utilisateur ne sait pas quelles classes, groupes ou catégories il va obtenir après classification.

La plupart des méthodes de clustering se basent sur une mesure de distance entre deux objets. Ceux-ci étant caractérisés par les attributs, cette notion de distance devra se baser sur des distinctions entre les valeurs prises par les attributs pour les différents objets.

On peut donc dire que toutes les techniques de clustering suivent un même principe général qui consiste à minimiser la distance entre deux objets d'un même cluster et à maximiser la distance entre deux objets de clusters distincts.

La figure ci-dessus illustre ces propos de manière schématique dans le cadre d'un problème bidimensionnel, c'est-à-dire pour des objets décrits par deux attributs. On détecte la présence de trois clusters.

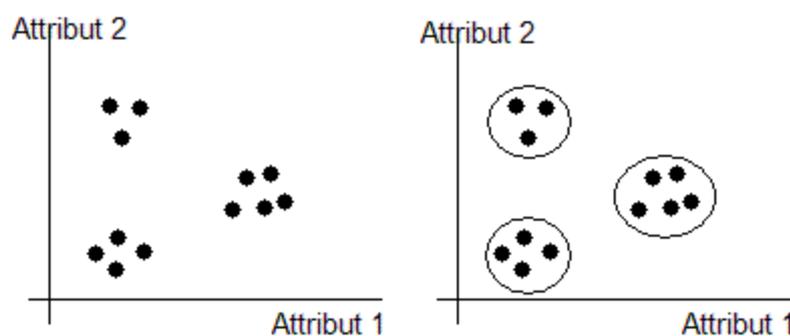


Fig III-3: - Illustration de la détection de clusters par mesure de distances

Le clustering est un sujet abordé par plusieurs métiers. Ce fait lui donne l'avantage d'être utilisé par divers domaines d'application qui peuvent être catalogués suivant le but recherché selon trois grands types :

- ✓ Extraction des connaissances : reprend les applications qui utilisent le clustering pour extraire de la connaissance d'une base de données. Concrètement, ceci vise à déterminer des « *sous-concepts* » afin de donner du sens à l'information dont on dispose. On espère ainsi pouvoir s'attaquer à des tâches telles que la génération d'hypothèses (modélisation prédictive), le diagnostic médical en se basant sur des caractéristiques communes de patients,...
- ✓ Réduction des bases de données : L'objectif visé est de réduire l'espace des données sur lequel on travaille en espérant que le problème qu'on cherche à résoudre devient beaucoup plus simple. La compression d'information et la segmentation d'images sont des cas concrets de ce type d'approches.
- ✓ Etude de comportement (*profiling*) : Le but est de détecter des sous populations qui partagent des caractéristiques proches, ce qui leur donne un certain comportement commun. Ceci touche directement les applications où l'on cherche à regrouper des clients (Customer Relationship Management, en marketing), dans les transports, dans les banques, dans les commerces, dans les télécommunications, dans la gestion de ressources (énergie,...),... On trouve également de nombreuses utilités dans le domaine de la classification de documents.

➤ **Terminologie**

Avant de se lancer dans l'explication des techniques permettant d'aboutir à une classification non-supervisée il est important de définir quelques concepts.

- **Représentation d'une classe** : La représentation d'une classe est une description de l'ensemble des individus constituant la classe. Cette description permet de définir la forme et la taille de la classe dans un espace de données. Dans les algorithmes de classification, la description d'une classe est simplement caractérisée par une combinaison linéaire des individus (également appelé centroïde), ou par un axe médian (médoïde) ne donnant qu'une forme implicite, celle du centre de la classe. Une telle description permet d'assigner chaque individu à un centroïde ou à un axe médian suivant une règle, qui est usuellement l'affectation d'un individu au centre qui lui est le plus proche.

- **Le centroïde:** La représentation centroïde d'une classe  $C$  est définie comme étant la moyenne arithmétique des éléments présents dans  $C$ . Plus clairement un centroïde est un vecteur  $V(C)$  où chacune de ses composantes correspond à la moyenne arithmétique des attributs des objets  $d_i$  se trouvant dans la classe  $C$  (Fig II-4).

L'Equation 1 définit le vecteur du centroïde  $V(C)$  d'une classe  $C$  donnée :

Equation 1 :

$$V(C) = \frac{1}{|C|} \sum_{i=1}^{|C|} d_i$$

- **Le médoïde :** On prend  $k$  individus parmi tous les individus d'une classe. Ces  $k$  éléments sont centraux vis à vis de la classe, c'est-à-dire qu'ils sont proches du centre géométrique de la classe. Cette représentation permet de représenter la classe non pas par un point unique mais par un *noyau* censé définir au mieux la classe (Fig II-5). Le choix de la valeur de  $k$  est empirique. Le choix des éléments du noyau est fondé sur un calcul de distances entre tous les individus de la classe. Pour une classe donnée  $C_j$ , le noyau  $N(C_j)$  correspond aux individus qui ont la plus petite somme de distances par rapport aux autres individus de la classe.

Pour une classe  $C_j$ ,  $X$  est un élément du noyau s'il minimise l'inertie, définie ci-dessous.

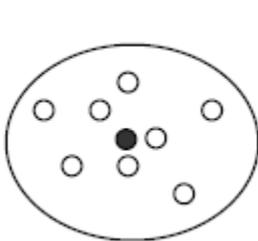


Fig III-4 :- Exemple de centroïde

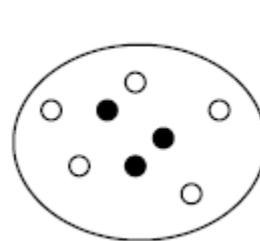
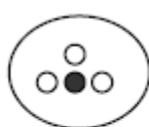


Fig III-5 :- Exemple de médoïde

- **La notion de proximité (distance) :** Une mesure de distance (appelée aussi mesure de similarité ou de proximité) exprime une similarité ou une dis-similarité : plus deux objets se rassemblent, plus leur similarité est grande et plus deux objets s'éloignent,

plus leur similarité est petite. La mesure de distance peut s'appliquer entre deux objets ou entre deux classes.

- **L'inertie** : L'inertie d'une population d'objets est la moyenne des carrés des distances des individus à leur centre. Cette entité mesure l'homogénéité des objets de la population.

L'inertie intra-classe est la somme des inerties des classes. Plus cette entité est petite, plus les objets de la même classe sont homogènes.

L'inertie interclasse est la moyenne des carrés des distances des centres des classes au centre global. Plus cette entité est importante, plus les classes sont éloignées.

Un bon clustering possède deux propriétés :

- Inertie intra-classe (variance des objets dans la même classe) minimale ;
- Inertie interclasse (variance des centres des classes) maximale.

Dans ce qui suit, nous allons présenter, globalement, les principaux algorithmes de clustering parmi les plus représentatifs et spécialement l'algorithme des k-means, celui utilisé dans ce mémoire.

- **Les algorithmes du clustering** : La majorité des algorithmes de clustering peuvent être répartis en quatre grandes familles :

#### Les algorithmes de partitionnement

- Construire k partitions et les corriger jusqu'à obtenir une similarité satisfaisante.
- k-Means, k-Medoids, PAM, CLARA et CLARANS ...

#### Les algorithmes hiérarchiques

- Ils sont descendants si, à partir d'une seule classe, ils cherchent à établir une partition par division ; ou ascendants s'ils cherchent à former des classes plus grandes par fusion de classes jusqu'à la satisfaction d'un critère d'arrêt.

- AGNES, DIANA, BIRCH, CURE, ROCK, ...

#### Les algorithmes à base de densité

- Grouper les objets tant que la densité de voisinage excède une certaine limite.
- DBSCAN, OPTICS, DENCLUE ...

#### Les algorithmes à base de grille

- Diviser l'espace en cellules formant une grille multi-niveaux et grouper les cellules voisines en termes de distance.
- STING, WaveCluster, CLIQUE

a) **Les algorithmes de partitionnement** : Leur principe général est de démarrer à partir d'un certain nombre de classes qui sont partitionnés d'une manière aléatoire en effectuant une "redistribution" des objets jusqu'à la rencontre d'un critère d'arrêt.

- ✓ **L'algorithme des k-Means** : L'algorithme des k-Means (ou l'algorithme des centres mobiles) a été introduit par J. MacQueen. Il est le plus utilisé dans les applications scientifiques et industrielles car il est le plus simple. C'est pour cause notre travail repose sur cet algorithme.

Dans cet algorithme, chaque classe est représentée par la moyenne (centroïde). k-Means est un algorithme itératif. Il commence avec un ensemble de  $k$  individus de référence choisis de façon aléatoire. Les individus de données sont ainsi partitionnés dans  $k$  classes ; un individu appartient à une classe si le centre de cette classe est le plus proche de lui (en terme de distance). La mise à jour des centroïde et l'affectation des individus de données aux classes sont réalisées pendant les itérations successives.

L'algorithme prend donc en paramètre  $k$ , le nombre de classes voulu, et se résume comme suit :

Entrée : Population de  $m$  individus,  $k$

Choisir aléatoirement  $k$  centres initiaux  $c_1, \dots, c_k$

Répéter

- Répartir chacun des  $m$  individus dans une classe  $i$  dont le centre  $c_i$  est le plus proche en termes de distance.
- Calculer les nouveaux centres : pour tout  $i$ ,  $c_i$  est la moyenne des individus de la classe  $i$  (Centroïde).

Jusqu'à ce qu'il n'y ait plus de changement (classes stables).

Sortie : Partitions.

La Figure ci bas montre un exemple de déroulement de l'algorithme k-Means sur un nuage d'objets bidimensionnels, avec  $k = 3$ .

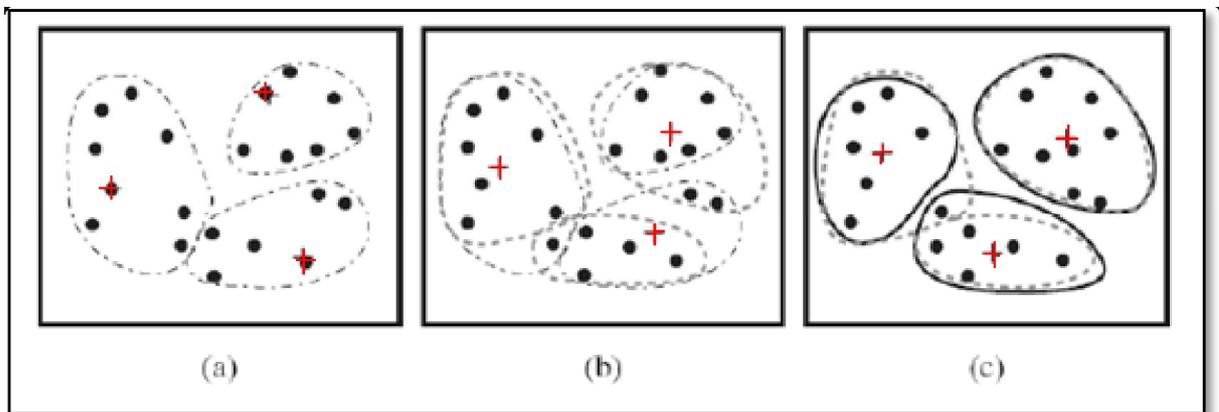


Fig III-6 : Partitionnement basé sur k-Means

Figure II-7. (a) : Choix aléatoire de trois centre initiaux (objets) marqués par (+) et affectation de chaque objet restant au centre le plus proche ;

Figure II-7. (b) : Calcul du nouveau centre pour chaque classe (moyenne des objets de la classe) et redistribution des objets selon les nouveaux centres ;

Figure II-7. (c): Le processus est répété jusqu'à stabilité des classes.

- ✓ **L'algorithme k-Medoids** : Il est introduit par Kaufman et Rousseeuw. Le principe de l'algorithme k-Medoids ressemble à celui de k-Means sauf que, contrairement à l'algorithme k-Means où la classe est représentée par une valeur moyenne, le centroïde, dans l'algorithme k-Medoids, une classe est représentée le médoïde.

Pour améliorer la qualité de clustering de L'algorithme k-Medoids, de nombreuses variantes s'y sont succédées, entre autres les algorithmes **PAM**, **CLARA** et **CLARANS**. Ces algorithmes adoptent le même principe de représentation par médoïde.

## 6) Conclusion

Le Data Mining sert à l'extraction d'information cachée dans de grandes bases de données. Il s'agit d'une technologie puissante très utilisé dans l'aide à la décision, car elle permet de prédire des tendances et des comportements. La résolution d'une problématique avec un processus de data mining peut se faire grâce à l'utilisation d'un grand nombre de méthodes et algorithmes différents. Cette richesse est due au fait que le data mining touche à plusieurs disciplines : les statistiques, l'analyse de données, l'intelligence artificielle et les bases de données.

Cela dit, le choix d'un algorithme approprié dépend fortement du contexte de son application, la nature des données et les ressources disponibles. Il est souvent nécessaire de combiner différentes techniques, afin de combler leurs lacunes.

Par conséquent, un mauvais choix de l'algorithme d'application engendre de lourdes conséquences sur le plan de la performance et des résultats. Dans certains cas, il peut faire échouer tout le projet.

Pour le développement de l'algorithme de gestion de localisation proposé dans ce travail, l'approche du clustering et de la méthode des k-means ont été choisies comme moyen de traitement servant à la construction automatique des zones de localisation propre à chaque abonné.

# Chapitre IV

## *L'algorithme De*

## *Gestion De Localisation*

## *Proposé, Basé Sur le Clustering*

**L**es algorithmes de gestion de localisation, comme ceux implémentés et présentés dans la majorité des propositions, supposent l'existence des zones de localisation prédéfinis et statiques. L'affectation des cellules aux zones de localisation est effectuée fréquemment, en utilisant des heuristiques et des lois de statistiques. Les zones de localisation résultantes s'appliquent à tous les abonnés. Cela dit, Il est évident que les abonnés inscrits au réseau n'ont pas tous le même comportement de mobilité, spatialement ou temporellement, et par conséquent, des zones de localisation statiques et globales sont loin d'être optimales.

L'algorithme proposé dans ce travail tente d'utiliser le profile de la mobilité des abonnés, pour créer automatiquement des zones de localisation individuelles pour chaque abonné, en utilisant un algorithme de clustering, à savoir l'algorithme des K-Means, et déterminer par conséquent les zones de paging les plus probables. La motivation derrière cette proposition, les données d'entrée de l'algorithme, les détails de l'algorithme sont présentés dans ce chapitre.

#### **4) Motivation**

Plusieurs propositions de la littérature suggèrent l'utilisation des profils utilisateurs pour individualiser les zones de localisation. Cependant, le contenu de ces profils, ainsi que leur création et leur maintenance, n'étaient pas bien définis. En plus, les algorithmes proposés déterminant dynamiquement des zones de localisation individuelles, souffraient généralement de complexité, n'étant pas supportables sur les équipements mobiles. Autres propositions de gestion de localisation sont basées sur le regroupement d'abonnés ayant un comportement similaire. Bien que plausible, un tel regroupement est assez arbitraire puisque les caractéristiques de mobilité d'un abonné peut changer. L'approche clé est que l'algorithme doit rester souple, pas trop compliqué, et efficace, même face à des changements de comportement de mobilité.

Dans ce travail, les idées ont été empruntées des études effectuées sur les déplacements d'individus, en particulier, ceux de la théorie du transport et des théories des modèles basés sur l'activité, ainsi qu'aux travaux menés dans [réf]. Dans la grande majorité des cas, un voyage est entrepris dans un but particulier survenant à une destination particulière. Le but derrière un voyage est généralement une activité précise, comme le travail ou le shopping. Ces activités peuvent mener à un nombre limité d'emplacements. Par exemple, la plupart des gens travaillent à des endroits fixes, les écoles et les zones de shopping sont situées sur des sites spécifiques. En général, le domicile est le point central à partir duquel tous les autres déplacements sont pris, et est généralement un emplacement où les abonnés passent une quantité de temps considérable durant la journée.

Les techniques de gestion de localisation peuvent faire usage du fait qu'un abonné moyen a un nombre limité de zones fréquemment visitées, qui regroupent d'une certaine manière le modèle de mobilité quotidien. L'algorithme ainsi proposé, exploite les caractéristiques spatiales et temporelles des déplacements de l'abonné. Par exemple, lors d'un voyage typique, un certain nombre de cellules peuvent être traversées entre l'origine et la destination ; L'abonné devra passer un certain temps à la destination (en fonction de l'activité), qui sera probablement plus long que le temps passé dans les cellules intermédiaires. Les cellules intermédiaires sont incluses dans la zone de localisation, mais seulement la cellule de destination (ou une autre cellule dans laquelle une période de temps

relativement longue a été épuisée) doit être pager (message de paging) en premier. L'abonné est moins susceptible d'être trouvé dans les cellules intermédiaires.

Une logique similaire est utilisée pour déterminer les zones de localisation. En général, une zone de localisation large est préférable, vu que la fréquence des mises à jour sera réduite. L'inconvénient est le nombre proportionnellement élevé des messages paging. Ceci peut être atténué en utilisant la stratégie de paging décrite ci-dessus. Toutefois, même dans ce cas, les zones de localisation doivent être de taille limitée, pour borner le nombre de messages paging, à la fois dans un premier tour, et dans les pires cas de situation (le mobile n'est pas trouvé). Si une ville entière, par exemple, a été déterminée comme une seule zone de localisation, un abonné pourrait avoir des dizaines de cellules dans lesquelles il doit être pager dans le premier tour.

Si des zones de localisation fixes sont utilisées, les déplacements quotidiens des abonnés peuvent encore traverser plusieurs zones de localisation sur un seul voyage entrepris par jour, générant par conséquent des mises à jour inutiles. Une zone de localisation personnalisée et individuelle, ancrée autour de la cellule dans laquelle la mise à jour de localisation s'effectue, serait composée de cellules les plus susceptibles d'être traversées par l'abonné. Les informations nécessaires à la création de cette zone sont dérivées du profil utilisateur, contenant les enregistrements sur les précédentes transitions de cellule en cellule, ainsi qu'au temps passé dans chaque cellule.

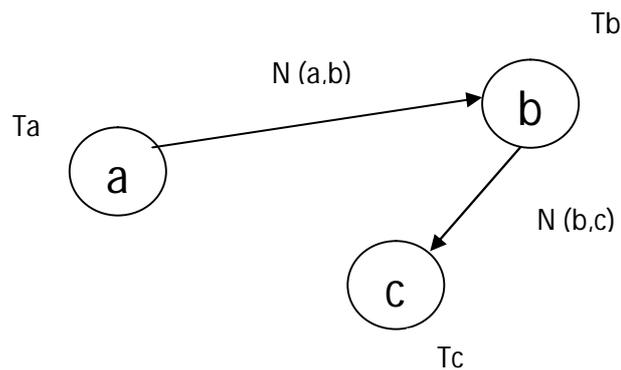
## **5) Historique de la mobilité et le profil utilisateur**

Le profil utilisateur se compose d'une base de données résumant l'historique de la mobilité des abonnés. Cette dernière consiste en le nombre de transitions qu'a effectué l'abonné pour aller d'une cellule à une autre, ainsi qu'à la durée moyenne qu'il passe dans chaque cellule visitée. Plus précisément, un compteur **N (a,b)** est calculé, représentant le nombre de fois que l'abonné passe de la cellule **a** à la cellule voisine **b**. On associe également à chaque cellule visitée un second compteur (chronomètre) **Ta**, représentant la durée moyenne passée dans la cellule **a**.

Quand la station mobile se déplace d'une cellule a vers une cellule b, le compteur  $N(a,b)$  est incrémenté ( $N(a,b) + 1$ ), et le timer  $T_b$  est lancé ; Il sera arrêté lorsque la station mobile quitte la cellule b vers une autre cellule voisine, ou alors lorsque on l'éteint. Quand  $T_b$  est

arrêté, la valeur moyenne de ce dernier est mise à jour, puisque le nombre total de visites de la cellule b peut être dérivée de la table stockant les compteurs N.

Les informations sauvegardées dans N (a,b) et Ta pour toutes les cellules visitées a et b peuvent être représentées par un graphe, où les nœuds représentent les cellules visitées, et les liens représentent les transitions entre ces cellules. Le poids d'un lien (a, b) est la valeur de N (a,b) et le poids du nœud a est la valeur de Ta.



Quand un abonné entre dans une nouvelle cellule n'appartenant pas à la zone de localisation actuelle, une mise à jour de localisation est alors effectuée et la nouvelle zone de localisation est déterminée. La zone de localisation est créée en utilisant le clustering, en se basant sur l'algorithme des K-Means, qui lui se base à son tour sur la fréquence de la mobilité de l'abonné à des moments précis de la journée. Les détails du traitement seront décrits ci-dessous.

## 6) Description de l'algorithme proposé

Compte tenu des exigences et des contraintes présentées plus haut, un algorithme de gestion de localisation a été conçu en utilisant comme données d'entrée l'historique de mobilité de l'abonné. Le profil utilisateur peut être présenté sous forme de graphe, comme décrit précédemment. Une zone de localisation contient les cellules les plus visitées par l'abonné à un moment donné de la journée. Une mise à jour de localisation est effectuée à chaque changement de cellule n'existant pas dans la zone actuelle où se trouvait l'abonné. À l'arrivée d'un appel entrant, un premier paging est effectué sur un sous-ensemble de cellules de la zone de localisation où l'abonné est plus susceptible d'y être, en se basant

cette fois sur la durée moyenne passée dans chaque cellule de la zone. Ce sous ensemble est appelé : Paging Area (Zone de paging).

### c) **Construction des zones de localisation et des zones de paging**

Les zones de localisation (LA) et les zone de paging (PA) sont construites automatiquement en utilisant comme données d'entrée les informations dérivées du profil utilisateur, représenté conceptuellement par le graphe de mobilité. Ces données regroupent l'ensemble des activités au quotidien de l'abonné, les cellules les plus visitées à chaque période de la journée, et la durée moyenne passée dans chaque cellule visitée. En termes de simulation, ceci consiste dans les deux types d'événement de mobilité, à savoir, l'événement génération de voyage et l'événement de changement de cellule. (Voir chapitre 5)

Les événements de changement de cellule sont classés par leur temps de survenance et sont utilisés dans la construction des zones de localisation et de paging. En effet, ces événements représentent dans la réalité l'ensemble des cellules visitées par l'abonné. Le quotidien de l'abonné est divisé en intervalles de temps, en fonction de la fréquence de mobilité et des activités entreprises à chaque moment de la journée. A titre d'exemple, les déplacements la nuit sont moins fréquents que le jour. Par conséquent, l'abonné a l'habitude de se localiser dans une seule zone de localisation, et a plus de chance d'être trouvé dans la cellule qui couvre son domicile. Donc, cette cellule devra être pagée en premier. De plus, les activités de la matinée sont différentes de celle de l'après midi ou de la soirée, et les activités entreprises les jours de semaine sont différents de celles du week-end ; Se rendre au boulot le matin, rentrer chez soi le soir, et shopping le week-end. Ceci dit, l'abonné pourrait reprendre le même chemin pour se rendre à son travail que celui du retour au domicile, donc il traversera les mêmes cellules.

Par contre, la fréquence de mobilité dans un intervalle de temps agit sur la taille de la zone de localisation. Dans le cas d'une forte mobilité de l'abonné, la zone de localisation construite sera large, comportant toutes les cellules que l'abonné à l'habitude de traverser dans cet intervalle. Ceci diminue considérablement le nombre de messages de mise à jour générés inutilement. D'autre part, si la mobilité est faible, la zone de localisation sera de petite taille. Cependant, dans une large zone de localisation, le nombre de paging a

tendance d'augmenter proportionnellement. C'est pour cela que le concept de zone de paging a été défini.

Chaque zone de localisation est divisée en zones de paging, selon des intervalles de temps, et les cellules pager dans une zone de paging seront celles où l'abonné passe le plus de son temps, comme les cellules domicile, travail et shopping. Les cellules intermédiaires seront pager en second dans le cas échéant.

Toutes ces contraintes et ces stratagèmes sont implémentés dans l'algorithme proposé ; Un algorithme basé sur le clustering, qui est une technique de classification dans le data mining, détaillé dans le chapitre 3.

Pour rappel, le clustering a pour principal objectif la représentation simplifiée des objets d'un environnement en les classifiant en  $k$  clusters homogènes. En d'autres termes, le clustering consiste à trouver une partition d'objets de telle sorte que les objets d'une même classe soient plus similaires entre eux qu'avec les objets des autres classes.

Dans cet algorithme, l'environnement représente la zone de couverture du réseau cellulaire. Les objets de l'environnement sont les cellules. Ces dernières seront classifiées dans un premier temps dans des clusters selon le nombre de visites. Ces clusters représentent les zones de localisation. Ensuite, chaque cluster est lui-même classifié en sous clusters, formant ainsi les zones de paging.

L'algorithme de clustering utilisé est celui des K-means. C'est l'algorithme le plus utilisé dans les applications scientifiques grâce à sa simplicité et son efficacité. Dans cet algorithme, chaque cluster est représenté par un centre appelé centroïde. La distance suivie dans les calculs est la distance Euclidienne et l'attribut d'un objet (cellule) se constitue dans le temps de survenance d'un changement de cellule. Dans chaque intervalle de temps, on a deux clusters. Le premier contient les cellules visitées par l'abonné, il représente donc la zone de localisation sélectionnée. Le deuxième comporte les cellules non visitées par l'abonné dans ce même intervalle de temps.

d) **La stratégie de mise à jour**

Une mise à jour de localisation est effectuée à chaque fois qu'un abonné entre dans une nouvelle cellule qui ne faisait pas partie de la zone de localisation ou il était, à chaque fois qu'une station mobile rejoint le réseau, ou périodiquement comme défini par le réseau. A l'issue de la mise à jour, une nouvelle zone de localisation est assignée à l'abonné, composée des cellules les plus susceptibles d'être traversées plus tard par l'abonné.

Les décisions concernant l'inclusion ou non d'une cellule à une zone de localisation, et les décisions concernant la procédure de mise à jour sont expliquées dans la séquence d'étapes décrites ci-dessous.

- a) A chaque fois que l'abonné entre dans une nouvelle cellule (par exemple de la cellule a à la cellule b), le nombre  $N(a,b)$  est incrémenté et la durée moyenne  $T_a$  est mise à jour. Dans la simulation, ceci est représenté par un événement de changement de cellule.
- b) Si une station mobile vient de se ré attacher au réseau après avoir été détachée (Eteinte de l'équipement mobile), un simple message est envoyé à la HLR, pour notifier qu'il est à nouveau accessible.
- c) Si la nouvelle cellule entrée appartient à la zone de localisation courante, aucune mise à jour n'est effectuée.
- d) Si la nouvelle cellule entrée n'appartient pas à la zone de localisation ou été l'abonné, un message de mise à jour est généré et fonction de mise à jour est alors appelée.
- e) La liste des cellules voisines à la nouvelle cellule entrée précédemment visitées est dégagée, ainsi que le nombre de fois que l'abonné s'est déplacé vers ces cellules, partant de la nouvelle cellule ( $N(a,b)$ ). La procédure de mise à jour attribue par conséquent à l'abonné une nouvelle zone de localisation individualisée construite par le clustering.
- f) Ces étapes ci-dessus sont répétées à chaque fois que l'abonné entre dans une cellule n'appartenant pas à sa zone de localisation.

e) **La stratégie de paging**

La stratégie de paging est très proche de celle de la mise à jour. En effet, dans cette proposition, un des paramètres maintenu par la station mobile est la durée moyenne passée dans chaque cellule visitée  $a$ , représentée par  $T_a$ . Les cellules avec une petite valeur  $T_a$  sont celles que l'abonné a tendance de traversée assez rapidement durant un déplacement, c'est les cellules intermédiaires. Une cellule intermédiaire peut être souvent traversée, mais l'abonné y passe un temps minimal, donc il est rarement trouvé dans cette cellule. D'autre part, l'abonné a de forte chance d'être trouver dans les cellules ayant une valeur  $T_a$  maximale.

La station mobile maintient une table de  $T_a$  pour chaque cellule. Durant une mise à jour de localisation, l'ensemble des valeurs  $T_a$  des cellules appartenant à la nouvelle zone de localisation sont transmis au réseau. Mettre à jour les valeurs  $T_a$  permettra au réseau d'effectuer un paging plus efficace.

La stratégie de paging nécessite l'utilisation des zones de paging et du paramètre  $T_a$  histoire de diminuer les coûts de paging. Les cellules de la zone de paging de la zone de localisation courante avec un haut  $T_a$ , et donc plus susceptible de contenir l'abonné, forment les cellules pager en premier par le réseau. Les cellules avec des valeurs inférieures de  $T_a$  sont ensuite pager s'il n'y a pas de réponse de l'abonné.

Pour régler le problème de détermination des cellules qui doivent être pager en premier ou pas, la valeur moyenne de  $T_a$  entre les cellules de la zone de paging est calculée. Les cellules ou  $T_a$  est supérieur ou égal à la moyenne forment les cellules pager en premier. Si cette tentative de paging échoue, toutes les cellules de la zone de paging sont alors pager (dans le cas où l'abonné se déplace vers les premières cellules pager juste après le premier tour de paging).

# Chapitre V

## *Description Du Modèle De Mobilité Et Evaluation Des Résultats*

**U**n modèle de mobilité, dans le contexte de la gestion de localisation, est un modèle résumant les déplacements quotidiens d'un abonné, ou plus exactement, les déplacements quotidiens d'un abonné inscrit au réseau. Un tel modèle est d'une importance primordiale dans les études de gestion de mobilité. Le nombre de messages de mise à jour de localisation et de paging, requis pour comparer l'efficacité des schémas de gestion de localisation, dépend fondamentalement des modèles de mobilité utilisateur.

Dans le passé récent, il y a eu plusieurs propositions d'algorithme de gestion de localisation, afin d'améliorer les méthodes actuelles utilisées pour définir statiquement les zones de localisation. En tentant de quantifier les performances de leurs propositions, la plupart des auteurs ont utilisé des modèles de mobilité qui sont trop simplifiés et irréalistes, complètement aléatoires, ou dépendent de nombreux supposés typiques des valeurs des variables clés.

Différentes propositions de gestion de localisation réagissent différemment vis-à-vis du caractère aléatoire du comportement des utilisateurs. Un modèle de mobilité est donc nécessaire, un modèle qui soit plus réaliste qu'un modèle de mobilité aléatoire. Les modèles de mobilité aléatoire ont été fréquemment utilisés dans le passé en raison de leur simplicité, et le fait que les premiers utilisateurs des communications cellulaires ont été généralement des abonnés très mobiles dont le comportement pourrait être relativement bien modélisés comme aléatoire.

Les modèles de mobilité aléatoires ne reflètent plus la réalité du comportement de la majorité des abonnés, dont ce dernier a tendance à suivre une certaine routine. Un modèle de mobilité aléatoire peut entraîner des résultats très partiels pour les travaux ayant trait aux aspects comportementaux de la mobilité utilisateur - par exemple, le fait que le déplacement dépend du travail ou de l'activité de l'utilisateur.

Le modèle de mobilité basé sur l'activité de l'abonné utilisé pour évaluer notre algorithme utilise certaines idées issues de la théorie du transport, et est décrit dans ce chapitre. L'objectif principal de la plupart des modèles issus de la théorie du transport est d'analyser le comportement observé des déplacements urbains et de sa relation avec certaines variables socio-économiques, dans le but de prévoir le comportement futur du voyage face à la croissance de la population.

Un élément de base de la modélisation des transports est le voyage (Trip). Un voyage est habituellement défini comme un déplacement à sens unique d'une personne, d'une origine à une destination, en utilisant un mode de transport. Les voyages sont souvent classés par le but du voyage, comme le travail, le shopping, ou le retour à son domicile.

D'autres concepts importants dans la modélisation du transport comprennent la notion de zones de trafic. Les zones de trafic représentent la structure spatiale d'une zone géographique, comme une municipalité par exemple, et qui peuvent varier en taille de quelques centaines de mètres carrés, à plusieurs kilomètres carrés.

## **1) Les éléments de la demande de voyage (Travel Demand)**

Il y a plusieurs approches de modélisation de la demande de voyage. Tous les modèles intégrés tentent cependant de décrire et de prévoir les quatre composantes de base de demande de voyages [11]. Certaines méthodes visualise et modélise chaque composant séparément, tandis que d'autres combinent deux ou plusieurs étapes.

Génération de voyage : La première étape consiste à estimer le nombre total de voyages  $T_i$  en provenance de la zone  $i$  et/ou le nombre total de voyages  $T_j$  se terminant à la zone  $j$ . En pratique, les voyages sont classés selon leur objectif, comme le travail, le shopping, ou les loisirs. Si une des extrémités du voyage est le domicile de l'abonné alors le voyage est appelé home-based trip et Non-home-based trip dans le cas contraire.

Distribution de voyage : Compte tenu du nombre total de voyages en provenance d'une zone donnée zone  $Z$ , cette étape calcule la valeur  $T_{ij}$ , qui est le nombre de voyages allant de chaque zone  $i$  à toutes les autres zones  $j$ . Le nombre de voyages entrant dans une zone déterminée peuvent également être distribué à toutes les autres zones d'origine.

Le Modal Split (Modal Share) : le mode fait référence aux différents types de moyen de transport empruntés. Le modal split détermine le rapport relatif aux différents modes utilisés dans  $T_{ij}$ . Malgré son importance dans la modélisation du transport urbain, le mode n'est pas pris en compte dans le modèle de mobilité décrit ici. En effet, La mode a des implications majeures en termes de handover, dû aux différentes vitesses des différents modes, mais il n'affecte pas la gestion de localisation.

Affectation de voyage : Vu le nombre de voyages  $T_{ijm}$  d'une zone  $i$  à une zone  $j$  en utilisant un mode  $m$ , cette étape cartographie le parcours du voyage sur des infrastructures de transport existantes ou proposées. L'itinéraire choisi est typiquement celui avec le plus petit coût généralisé, généralement une combinaison de temps, distance, vitesse, et de commodité.

## **2) Le modèle de mobilité proposé basé sur l'activité**

Le modèle de mobilité utilisé, s'appuyant sur celui utilisé dans [11], est présenté dans cette section, en commençant par les besoins et les exigences généraux du modèle. L'enquête qui a fourni les données utilisées dans la simulation est également décrite, comme tout modèle est, dans une certaine mesure, limité par les données disponibles.

#### a) **Exigences du modèle et disponibilité des données**

L'objectif du modèle de mobilité est de fournir, au niveau de chaque abonné individuel, un ensemble réaliste de chemins parcourus sur une base quotidienne. Comme le modèle sera utilisé pour évaluer un algorithme de gestion de localisation qui repose sur un profil utilisateur à long terme, une période de plusieurs jours sera simulée afin de permettre la création d'un profil utilisateur. Le modèle sera appliqué à un système de télécommunication cellulaire, donc les chemins parcourus peuvent être superposés sur un cadre géographique de cellules radio. Par conséquent, la sortie finale du modèle sera une liste de cellules traversées par un abonné individuel sur une période de plusieurs jours, avec la durée moyenne passée dans chaque cellule.

Plusieurs informations concernant la mobilité sont recueillies par des enquêtes périodiques menées par les bureaux de planification de la circulation. Ces sondages ont de nombreux usages, tels que la planification du trafic et l'élaboration des politiques de transport public. Ces données comprennent normalement les informations concernant le voyage pour un échantillon de ménages pour un jour de semaine, bien que les données exactes recueillies varient d'une région à une autre et d'une enquête à une autre.

Les données d'entrée pour le modèle de mobilité utilisé dans ce mémoire a été tirée de l'enquête de voyage menée par la municipalité régionale de Waterloo, au Canada. Cette enquête a capturé des informations sur les voyages effectués par les résidents de cette ville pendant une semaine particulière. Une autre section de l'enquête contenait des informations sur les individus, dont certains ont été inaccessibles pour des raisons de confidentialité. Les données de l'enquête ont été agrégées dans un seul fichier contenant des éléments pertinents pour cette simulation.

Les données suivantes ont été incluses pour chaque voyage enregistré :

- L'identificateur d'une personne ;

- L'heure de départ du voyage ;
- L'heure d'arrivée du voyage ;
- Le but du voyage (ou activité) à l'origine et à la destination ;
- La zone du domicile de la personne ou le générateur de voyage ;
- La situation de l'emploi (travail à plein temps ou à temps partiel, à son domicile ou en dehors) ;
- Lieu de travail fixe (vrai ou faux).

L'enquête de voyage ne pouvait pas être utilisée dans son format original, principalement parce qu'il recueille des informations pour une semaine seulement. Plusieurs schémas de gestion de localisation, y compris celui proposé ici, impliquent un certain type de profil de déplacement utilisateur construit au fil du temps. Le modèle utilise les données d'entrée agrégées extraites de l'enquête pour simuler la mobilité des abonnés sur une période de plusieurs jours de semaine. Les données d'entrée pour la simulation et la procédure de simulation sont décrites ci-dessous.

#### **b) Aperçu de modèle de mobilité**

L'objectif final du modèle de mobilité est de simuler le mouvement quotidien des abonnés sur une période de plusieurs jours, en termes de cellules traversées et du temps passé dans chaque cellule. En utilisant ces informations, les algorithmes de gestion de localisation peuvent être comparés en fonction du nombre relatif des messages de mises à jour de localisation et des messages de paging qu'ils génèrent. Le modèle de mobilité suivi se base donc sur les approches basées sur les activités des abonnés.

Le paramètre de conduite dans la simulation est la notion de but du voyage, ce qui équivaut à une activité dans cette proposition. Ceci est un paramètre important dans l'enquête de voyage original, qui classe les raisons du voyage en neuf catégories, à savoir (1) le travail, (2) les travaux connexes, (3) l'école, (4) passager, (5) Shopping (6) travaux sociaux / loisirs, (7) affaires personnelles, (8) retour au domicile, et (9) autres.

Dans la simulation, un jour de 24 heures est divisé en un nombre de segments égaux, appelées périodes. Les périodes sont utilisées pour agréger les données de l'enquête en heures de la journée. Un jour simulé pour un abonné particulier consiste en une séquence d'activités qui ont une durée et un emplacement. Les chemins empruntés pour voyager

d'une activité à une autre, avec le temps passé à chaque activité, forment les données de sortie spatiale et temporelle nécessaire à la simulation.

La notion d'activité, ou but du voyage, est central à la simulation. Chaque activité se caractérise par son temps de la journée, sa durée, et son emplacement (la cellule). Une activité est sélectionnée en se basant sur l'activité précédente et la période actuelle. La probabilité de transition d'une activité à une autre utilise la matrice de transition d'activités décrite ci-dessous. Une fois l'activité suivante est sélectionnée, sa durée est déterminée en utilisant la matrice de la durée d'activité. Enfin, la localisation de l'activité est sélectionnée, en fonction du type d'activité, et de quelques heuristiques. Comme l'emplacement actuel de l'abonné est déjà connu, une fois l'emplacement de l'activité suivante est sélectionné, le parcours intermédiaire (en termes de cellules traversées) et la distance totale sont déterminés à partir d'un tableau de recherche. Le temps total et le temps passé dans chaque cellule intermédiaire sont calculés. L'abonné reste dans la cellule de destination pendant toute la durée de l'activité, et la séquence est répétée.

Chaque abonné simulé a certaines caractéristiques, l'une des plus importantes est le type de personne. Les distributions dans la matrice de transition d'activité et dans la matrice de durée d'activité sont indexées par type de personne. Quatre catégories de type de personne ont été définies, de façon similaire à [30]. Cette catégorisation a tenté de créer des groupes d'abonnés assez similaires, avec des schémas de mobilité similaire, en utilisant les informations collectées lors de l'enquête. Les quatre catégories sont :

- Travailleur à plein temps à l'extérieur du domicile ;
- Travailleur à temps partiel à l'extérieur du domicile, mais non étudiant ;
- Etudiant, pouvant travailler à temps partiel à l'extérieur de son domicile ;
- Ne travaillant pas à l'extérieur du domicile, et non étudiant.

Une autre caractéristique importante est de savoir si le lieu de travail est fixe, puisque le travail est une activité majeure. Il est important notamment de connaître les cellules du domicile de l'abonné, de son travail, et/ou de son école/université. Ces informations sont

utilisées pour sélectionner la cellule de destination lorsque l'activité correspondante (par exemple le travail) est sélectionnée.

### c) Les données d'entrée

Pour déterminer la prochaine activité, sa durée, son l'emplacement et le chemin géographique à prendre, les entrées sont recherchées dans plusieurs tableaux, parfois en utilisant un indice aléatoire. La dérivation et l'utilisation de ces tables sont spécifiées ci-dessous.

Matrice de transition d'activité: Après validation, les données de l'enquête ont été manipulées par un programme externe pour obtenir des données directement applicables à cette simulation. La matrice de transition d'activité est l'une des tables de données extraites à partir des données de l'enquête.

La matrice de transition d'activité est une répartition empirique des transitions d'une activité (ou but du voyage dans l'enquête) à une autre, enregistrée comme un tableau à quatre dimensions. La fonction est indexée par le type de personne de l'abonné, la période actuelle, et l'activité précédente. Une probabilité cumulée de transition à la prochaine activité est associée à chaque entrée [11], comme indiqué dans le tableau IV.1.

Type Personne	Période	Activité précédente	Activité Suivante	Probabilité Cumulée
1	4	8	1	0.351724
1	4	8	2	0.393103
1	4	8	3	0.420690
1	4	8	4	0.475862
1	4	8	5	0.696552
1	4	8	8	1.000000
1	4	9	1	0.000000

Tab IV-1 :- Exemple de données de la matrice de transition d'activité

La matrice de durée d'activité: La matrice de durée d'activité (tableau IV.2), qui a été également créé par un programme externe, est un tableau indexé par le type de personne, par la période, et par l'activité actuelle, donnant une probabilité cumulée des différentes

durées d'activité qui ont été observés [11]. La durée d'une activité particulière peut être dérivée des enregistrements de l'enquête classés par ordre chronologique en calculant la différence entre l'heure d'arrivée d'un voyage, et l'heure de départ du voyage suivant. La fréquence relative des durées (regroupées en intervalles de 5 minutes) a été enregistrée, avec le type de personne correspondant, la période et but du voyage.

En général, le premier voyage enregistré de la journée commence toujours du domicile de l'abonné, et il en est de même pour la terminaison du dernier voyage.

Type Personne	Période	Activité	Durée	Probabilité Cumulée
0	7	6	330	0.948718
0	7	6	360	0.961538
0	7	6	400	0.974359
0	7	7	0	0.125654
0	7	7	5	0.235602
0	7	7	10	0.413613
0	7	7	15	0.539267

Tab IV-2 :- Exemple de données de la matrice de durée d'activité

Le chemin géographique : Dans un modèle de mobilité réaliste, les cellules ne sont pas toutes visitées pendant la même durée. En effet, certaines cellules représentent l'emplacement des activités étendues, tandis que d'autres, sont traversées brièvement pendant que l'abonné se déplace pour rejoindre la destination finale. Ce concept de cellules intermédiaires est critique vis-à-vis du modèle de mobilité, afin de refléter la répartition du temps passé dans diverses cellules. Compte tenu des cellules de l'origine et de destination, une table a été créée pour rechercher le chemin pris, au niveau cellulaire. Une autre table contient la distance totale de l'origine à la destination.

Les informations de routage ont été extraites en utilisant le logiciel de système d'information géographique appelé TransCAD \*. La grille de la ville a été entrée dans TransCAD, sur lequel se superpose une couche de cellules. Chaque cellule est une zone géographique

correspondant approximativement à une couverture radio cellulaire. Les cellules sont groupées dans des zones de trafic adjacentes. Ces dernières sont importantes car elles représentent l'unité géographique de base pour laquelle des données de l'enquête sont recueillies. L'affectation des zones de trafic aux cellules était en quelque peu arbitraire, mais en suivant certaines règles de base, à savoir, la convexité des cellules, et la taille des cellules qui devait être inversement proportionnelle à la route et la densité de population. Chaque cellule était représentées par un centroïde, un point au centre de la cellule, et est relié à la grille de la ville. Le chemin pris d'une cellule à une autre était en pratique un chemin d'un centroïde à un autre.

Un total de 45 cellules a été défini. En utilisant la fonction de chemin minimal de TransCAD, la distance minimale entre chaque paire de centroïdes a été calculée, et les cellules traversées par cette voie minimale ont été enregistrées manuellement. Dans le cas où les frontières de cellule ont coïncidé avec le chemin minimal, la règle suivie a été de minimiser le nombre de changements de cellule en restant dans la même cellule le plus longtemps possible. Dans quelques cas où l'angle d'une autre cellule a été touché avant de retourner à la même cellule, aucun changement de cellule n'a été enregistré, car en pratique les frontières des cellules sont flous, et un bon algorithme de sélection de cellules devrait éviter des tels changements. En utilisant la longueur du chemin minimal calculé, le temps du déplacement était calculé en divisant la longueur de chemin par la vitesse moyenne, supposée constante pour toute la région. Cette vitesse moyenne devrait être relativement faible pour tenir compte des temps d'attente aux panneaux d'arrêt, feux de circulation et des routes encombrées.

Les informations de l'abonné : Le modèle de mobilité décrit les mouvements des abonnés individuellement, et certaines informations doivent être maintenues pour chaque abonné. La sortie dérivée du modèle doit refléter l'individualité d'un abonné, sinon il serait impossible de le distinguer d'un modèle de mobilité aléatoire. En théorie de transport, les individus (ou groupes) peuvent être classés selon différents paramètres, comme le revenu,

possession d'une voiture, etc. Pour éviter une complexité inutile, mais encore dans une

\*TransCAD est un logiciel de modélisation spécialisé dans l'étude de problématiques liées à la planification des transports. Il dispose d'un Système d'Information Géographique intégré qui permet de meilleures représentations graphiques des résultats et intègre des procédures de calages particulièrement puissantes.

quatre valeurs, fondées principalement sur la situation d'emploi. Certains endroits (cellules)

qui ont été associés à un individu ont également été inclus, comme le domicile, le travail, et les lieux scolaires. Le lieu de travail peut être ou ne pas être fixé au cours du temps, et une variable conserve ces informations. Cependant, le domicile et les lieux scolaires sont supposés fixes. Cette information est utilisée pour déterminer la cellule de destination à chaque fois que l'activité correspondante est sélectionnée. Ainsi, le modèle de mobilité proposé intègre la structure des modes de déplacement quotidien, qui est complètement ignoré dans les modèles de mobilité aléatoire. L'activité, la cellule et la zone de localisation courantes sont également enregistrées pour chaque abonné dans la simulation.

#### **d) La procédure de simulation**

En utilisant les données des matrices décrites ci-haut, une simulation à événements discrets est développée. En bref, une simulation à événement discrets ordonnance les événements dans une file d'attente qui surviennent à un certain moment en se basant sur l'horloge globale, et ces événements sont dépilés dans un ordre chronologique et agissent selon leur type d'événement.

Dans la simulation de la mobilité, on distingue trois types d'événement : événement génération de voyage (trip generation event), événement changement de cellule (cell change event) et événement appel entrant (incomming call event). La survenance d'un type d'événement particulier fait appel à sa fonction correspondante. La simulation est débutée par un événement génération de voyage, dont les paramètres sont entrés par l'utilisateur. La fonction de cet événement crée et met dans une file un ensemble d'événements de changement de cellule. Ces derniers correspondent aux cellules traversées quand l'abonné sélectionne une nouvelle activité et se rend à son nouvel emplacement. La fonction de l'événement changement de cellule effectue la procédure de mise à jour, qui peut être déclenchée quand un changement de cellule se produit. La fonction de génération de voyage crée notamment à autre événement de génération de voyage, correspondant à la fin de l'activité de destination.

Les appels entrants ne sont pas modélisés par le processus de Poisson, vu que le modèle n'est pas très réaliste pour les appels reçus de chaque abonné. En effet, un abonné a plus de chance de recevoir un appel à 14h qu'à 2h, un fait qui n'est pas pris en compte par la loi de Poisson. Un paramètre clé utilisé dans la simulation est le nombre d'appels entrants, **Ca**, au cours des 24 heures, qui est défini à l'exécution. La fonction d'appel entrant crée et met dans

une file **Ca** appels entrants pour les prochaine 24 heures en se basant sur une heuristique de loi de probabilité pour l'heure d'appel. Elle enfile notamment un événement qui va déclencher la génération des prochains appels entrants.

Quatre variables indépendantes sont utilisées comme entrées de la simulation. Le type de personne et la variabilité de l'emplacement du travail paramètrent les aspects de la mobilité, alors que le nombre quotidien d'appels entrants et la taille maximale de la zone de localisation affectent l'algorithme de gestion de localisation. Pour chaque combinaison des variables indépendantes, plusieurs simulations sont exécutées. Les sorties détaillant les messages de mise à jour de localisation et de paging générés par les différents algorithmes de gestion de localisation sont renvoyées vers des fichiers externes.

Événement génération de voyage (de déplacement) : Les événements de génération de voyage causent tous les déplacements de l'abonné pendant la simulation. Les entrées de la fonction sont l'activité précédente, le type de l'abonné, et le temps d'horloge de la simulation (qui est converti en période correspondante). La sortie est l'ensemble d'événements de changement de cellule. Après la mise en file d'attente des événements de changement de cellule, un autre événement de génération de voyage est ajouté à la file, représentant ainsi la prochaine activité et destination, comme décrit ci-dessous. Les événements de changement de cellule correspondent à l'ensemble de cellules traversées par l'abonné, allant de la position actuelle à l'activité suivante.

La détermination de l'activité suivante de l'abonné conduit la génération du modèle de mobilité quotidienne, comme le montre la figure IV.1. La prochaine activité est sélectionnée en utilisant la matrice de transition d'activité. Compte tenu de l'activité actuelle, de la période actuelle, et du type de l'abonné, l'activité suivante est sélectionnée aléatoirement depuis les entrées correspondant de la matrice de transition d'activité. En fonction de la période et du type, l'activité suivante sélectionnée est utilisée pour déterminer la durée de l'activité, en utilisant cette fois-ci, la matrice de durée d'activité. La cellule de destination dépend également de la prochaine activité sélectionnée. Si l'activité a pour but d'aller au travail (pour un emplacement du lieu de travail fixe), à l'école, ou retourner à son domicile, les localisations constantes sont utilisées depuis les informations de l'abonné déjà

enregistrées. La localisation retournée aux autres activités (excepté le shopping) est un nombre aléatoire uniformément distribué représentant les cellules de 1 à 45.

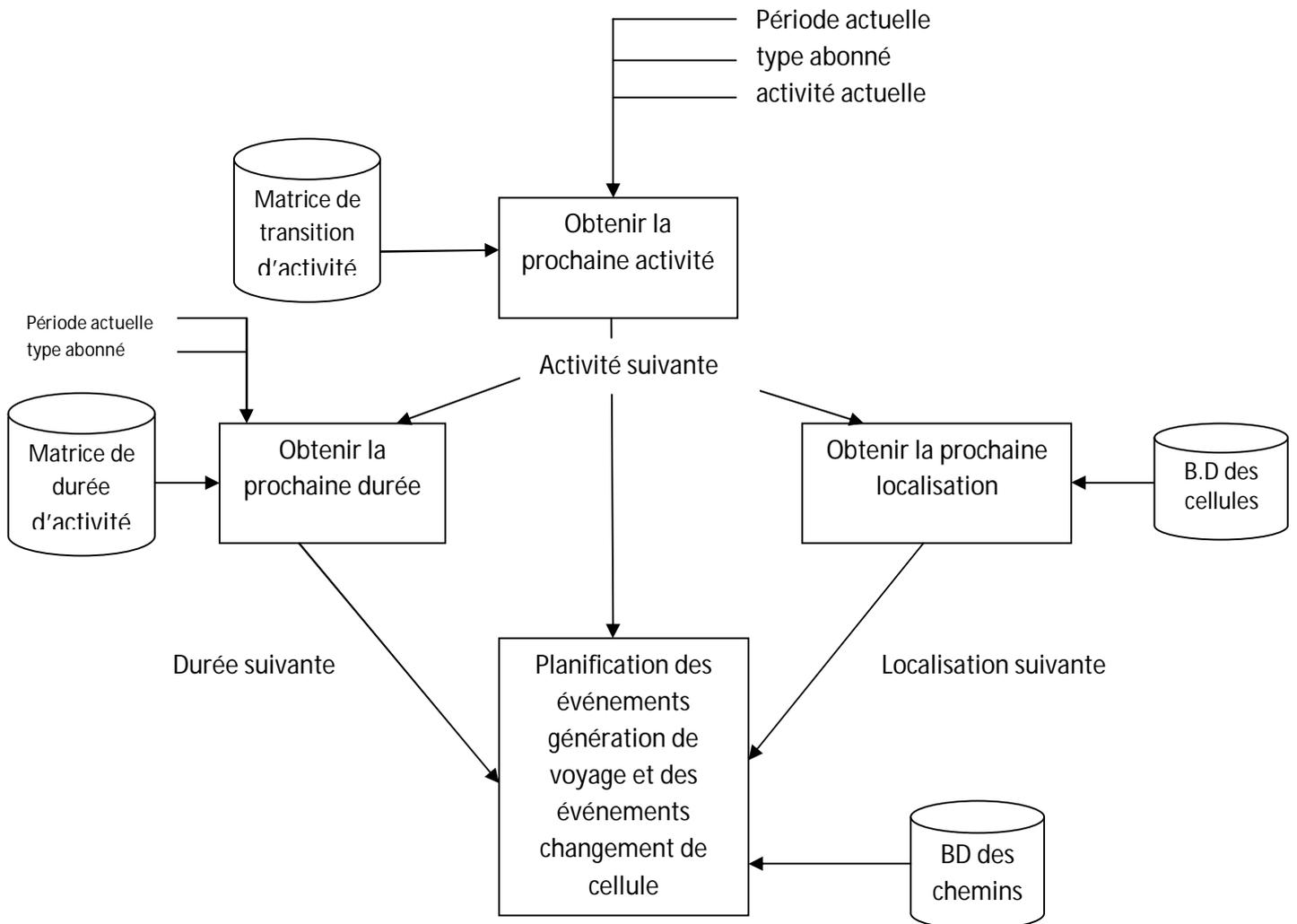


Fig IV-1 :- Schéma de flux de données pour la génération de voyage dans le modèle de mobilité proposé.

Pour le shopping, les destinations possibles sont limitées, en raison de contraintes de localisation du shopping, ainsi que l'habitude et la commodité des individus. Par conséquent, le shopping peut être modélisé de manière plus réaliste en termes de distance à l'emplacement du shopping, ou de zone de travail en détail (zonal retail employment) [11]. La méthode utilisée ici, est une combinaison des deux approches. La zone de travail en détail

est divisée par la distance depuis la localisation actuelle à la cellule cible. La cellule de destination est sélectionnée aléatoirement parmi les cinq premières cellules cibles. Intuitivement, cette méthode privilégie les cellules qui sont plus près de l'emplacement actuel de l'abonné, et est plus susceptible de sélectionner les quartiers de shopping.

En utilisant la cellule actuelle et la cellule de destination, le chemin et la distance totale sont recherchés dans les tables des chemins et des distances. La distance totale est divisée par un système défini par l'utilisateur de la vitesse moyenne pour calculer le temps total, qui est divisé par le nombre de cellules traversées pour dériver le temps correspondant. Une hypothèse a été faite disant que les temps de la traversée des cellules sont égaux pour toutes les cellules intermédiaires. Pour chaque cellule intermédiaire, un événement changement de cellule est créé et mis dans la file d'attente, avec un temps d'événement égal au temps actuel plus un multiple approprié du temps de la traversée de la cellule.

Enfin, un second événement génération de voyage est ajouté à la file d'attente, avec un temps d'événement égal au temps actuel plus le temps total du déplacement à la destination plus la durée de l'activité (une fois l'activité choisie est terminée, une nouvelle activité devrait être dérivée).

Événement de changement de cellule et d'appel entrant: Quand un événement de changement de cellule survient, la fonction de gestion de localisation est appelée. Les actions de la mise à jour dépendent de l'algorithme de gestion de localisation implémenté. Les détails de la mise à jour, telles que le temps et les cellules dans lesquelles elles se sont produites et les cellules qui composent la nouvelle zone de localisation, sont notés dans un fichier externe pour un traitement ultérieur.

Indépendants de tout mouvement, les appels peuvent arriver à l'abonné, représentés par un événement appel entrant. La simulation va créer aléatoirement **Ca** heures d'appels entrants en utilisant une loi de probabilité. Une journée est divisée en trois périodes, chacune avec un

Temps de la journée	Probabilité d'appels entrants
---------------------	-------------------------------

00:00 to 08:00	0.2
08:00 to 18:00	0.5
18:00 to 24:00	0.3

Tab IV-3 : - loi de probabilité des appels entrants au cours d'une journée

pourcentage correspondant des appels entrants au quotidien. Cela représente approximativement les périodes occupées observées durant la matinée et l'après-midi, et les périodes creuses de fin de nuit et de début de matinée. La distribution heuristique utilisée ici est montré dans le tableau IV.3.

Similairement à l'événement de changement de cellule, quand un événement d'appel entrant survient, la fonction de gestion de localisation est appelée. Cette fonction manipule les actions de paging, selon l'algorithme de gestion de localisation implémenté. Les détails concernant le nombre de message paging générés, sont notés dans un fichier externe pour un traitement ultérieur.

### **3) Evaluation des résultats**

Dans cette partie, l'algorithme de gestion de localisation proposé est comparé à des propositions existantes, principalement en termes d'efficacité radio de la bande passante. Tel que déjà mentionné dans les chapitres précédents, le but le plus important dans un algorithme de gestion de localisation est l'utilisation efficace de la bande de fréquence, qui est une ressource fixe. Cette exigence peut être satisfaite en minimisant le nombre de message de mise à jour et de paging qui doivent être transmis à travers l'interface radio.

#### **1. Description des données et des procédures**

L'analyse implique la comparaison de l'algorithme proposé basé sur le clustering avec des algorithmes de gestion de localisation standards **fixes** et **dynamiques**. Les 45 cellules radio dans l'environnement de simulation sont groupées dans 4 dispositions de zone de localisation fixes (**0, 1, 2, 3**) de manière aléatoire. Ces zones de localisation sont équivalentes aux zones de paging, et les termes sont utilisés indifféremment dans les paragraphes ci-dessous. Le groupe fixe 0 contient une cellule par zone de localisation. Le groupe fixe 1 possède treize zones de localisation de trois ou quatre cellules chacune. Le groupe fixe 2 est similaire, avec trois à cinq cellules, généralement différentes que celles du groupe fixe 1, dans chacune des onze zones de localisation. Enfin, le groupe fixe 3 est divisé en cinq zones de localisation suffisamment large, avec huit à dix cellules par zone de localisation.

Concernant l'algorithme dynamique, présenté dans le chapitre deux et proposé dans [24], les zones de localisation sont construites dynamiquement en se basant sur des statistiques des déplacements de l'abonné. Un algorithme dynamique a été implémenté avec le paramètre de la taille maximale d'une zone de localisation variable. Ces différentes combinaisons d'algorithmes ont été comparées avec l'algorithme proposé basé sur le Clustering.

Les résultats de la simulation fournissent les détails concernant les messages de mise à jour et de paging générés au fil du temps par un abonné simulé, pour chacun des algorithmes de gestion de localisation, qu'il soit fixe, dynamique, ou basé sur le clustering. Les déplacements et les messages générés pour chaque abonné ont été enregistrés pour 50 jours simulés, un intervalle assumé dès le départ pour être suffisamment long pour le modèle histoire d'atteindre un état stable. Pour chaque ensemble de variables de contrôle (type personne,

le lieu de travail fixe, taille de zone de localisation dynamique, et le nombre d'appels entrants quotidiens), plusieurs répétitions ont été effectuées, chacune avec un ensemble différent de paramètres internes (temps de démarrage initial, l'emplacement du domicile, du travail, et de l'école). Le nombre de messages de mise à jour et de paging générés ont été calculés et classifiés, par jour simulé, pour chaque algorithme, et pour chaque ensemble de variables de contrôle.

## **2. Comparaison des coûts liés à la signalisation**

Le nombre moyen des messages de mise à jour et de paging générés par les différents algorithmes fixes et dynamiques a été comparé avec celui de l'algorithme développé dans ce travail basé sur le clustering. Le coût total de la gestion de localisation a également été calculé et utilisé pour comparer la performance globale des différents algorithmes.

### **c) Performance globale**

**Mise à jour** : la représentation graphique dans la figure (V-1) montre le nombre moyen de messages de mise à jour pour les différents algorithmes, et sur toutes les variables de contrôle (type personne (0, 1, 2, 3), le lieu de travail fixe ou non (vrai ou faux), la taille d'une zone de localisation dynamique (10, 15). Elle montre la relation globale entre les différents algorithmes fixes, dynamiques, et clustering. Comme prévu, le groupe fixe 0, avec une cellule par zone de localisation, signe le plus mauvais résultat. Les groupes fixes 1 et 2, avec une moyenne de quatre cellules par zone de localisation, signent un résultat meilleur que le précédent, mais moins bon que le groupe fixe 3, qui a en moyenne, neuf cellules par zone de localisation.

Pour l'algorithme dynamique, à chaque fois que le temps passe, le profil utilisateur est en mesure de fournir une meilleure information pour la sélection des cellules. En effet, après la première journée, l'algorithme dynamique est plutôt comparable à celui des groupes fixes, mais ces résultats s'améliorent en fait et à mesure que le temps s'écoule.

La courbe d'apprentissage associée à l'algorithme du clustering proposé est clairement visible sur le graphe. En effet, Le nombre de messages de mise à jour est plus stable et les

coûts de signalisation est à son minimum. Ceci témoigne de la performance et de l'efficacité du clustering par rapport aux autres algorithmes.

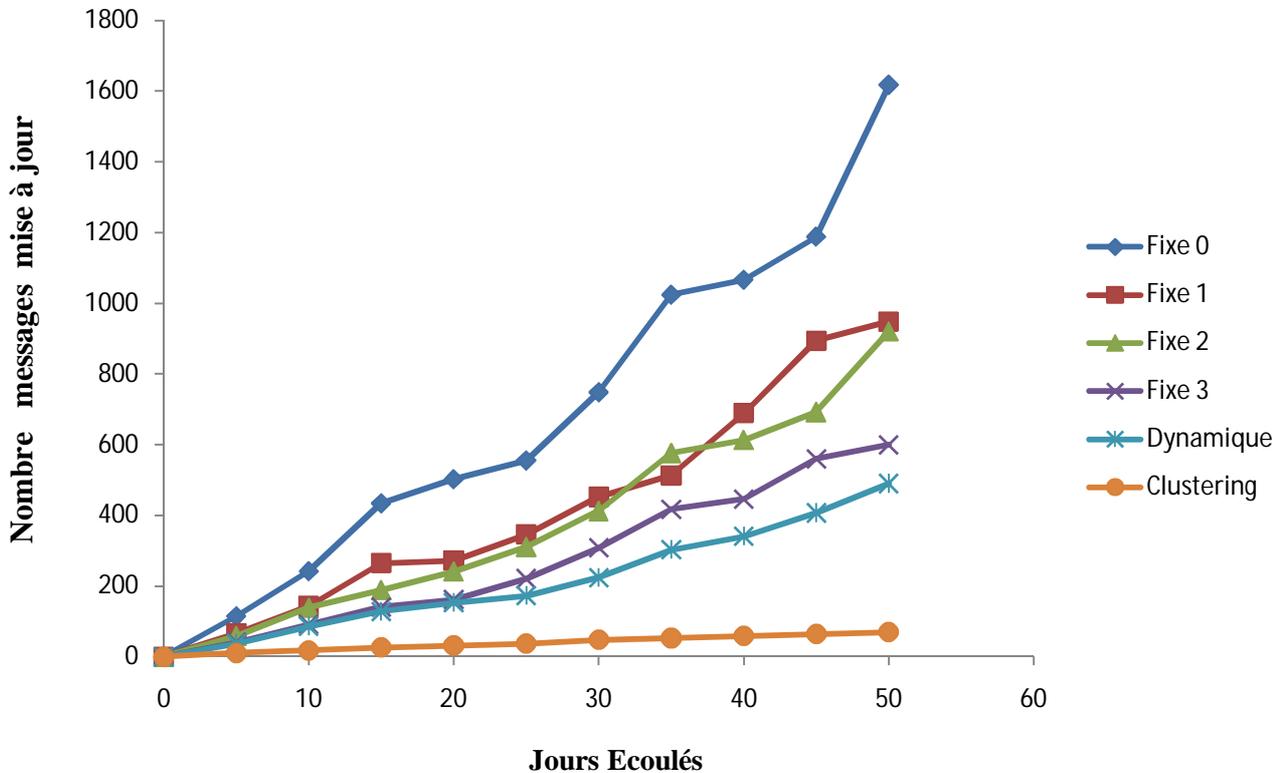


Fig V – 2 :- Nombre moyen des messages de mise à jour des différents algorithmes

**Paging**: Le nombre moyen global des messages de paging générés par les différents algorithmes implémentés est décrit par les courbes dans les figures (V-2), (V-3), (V-4), et ceux pour 3, 9, et 12 appels entrants, respectivement. Les résultats grimpent presque identiquement pour les différents niveaux d'appels entrants indiquant que l'algorithme du clustering proposé signe pratiquement les meilleurs résultats par rapports aux autres algorithmes.

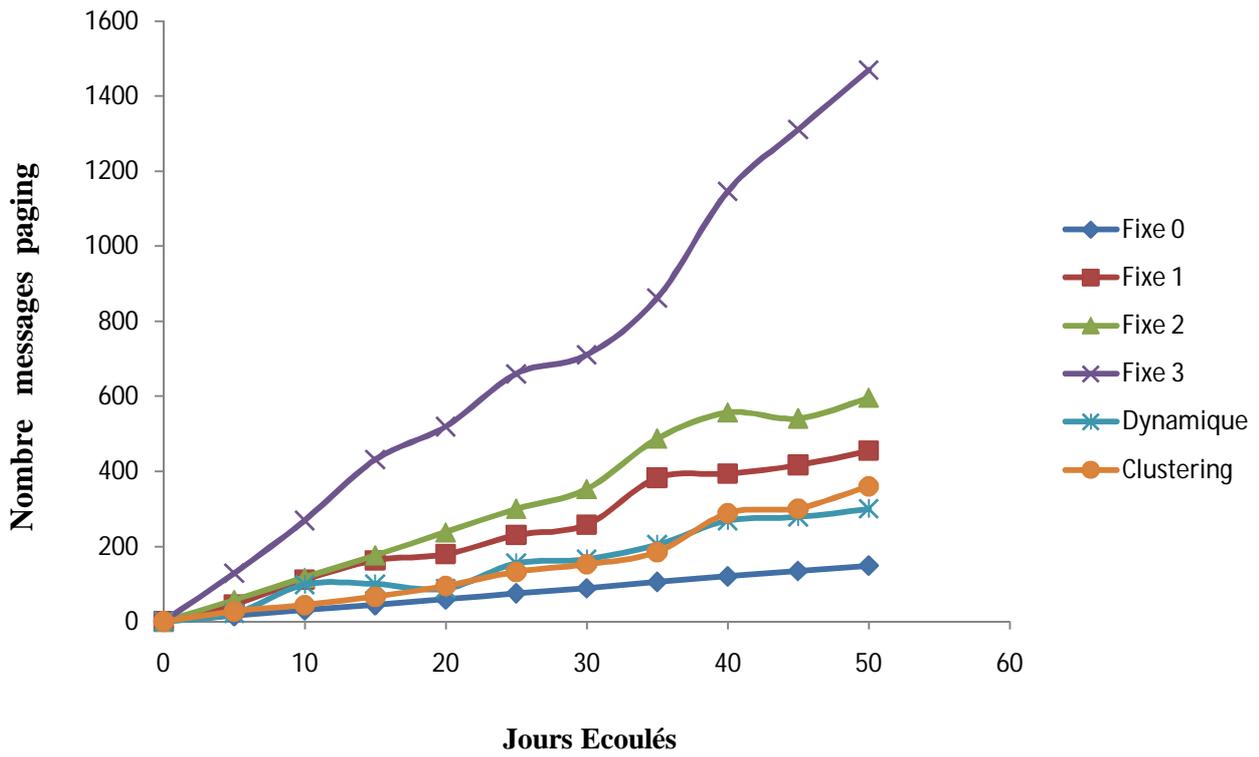


Fig V – 3 :- Nombre moyen des messages paging générés pour 3 appels entrants quotidiens

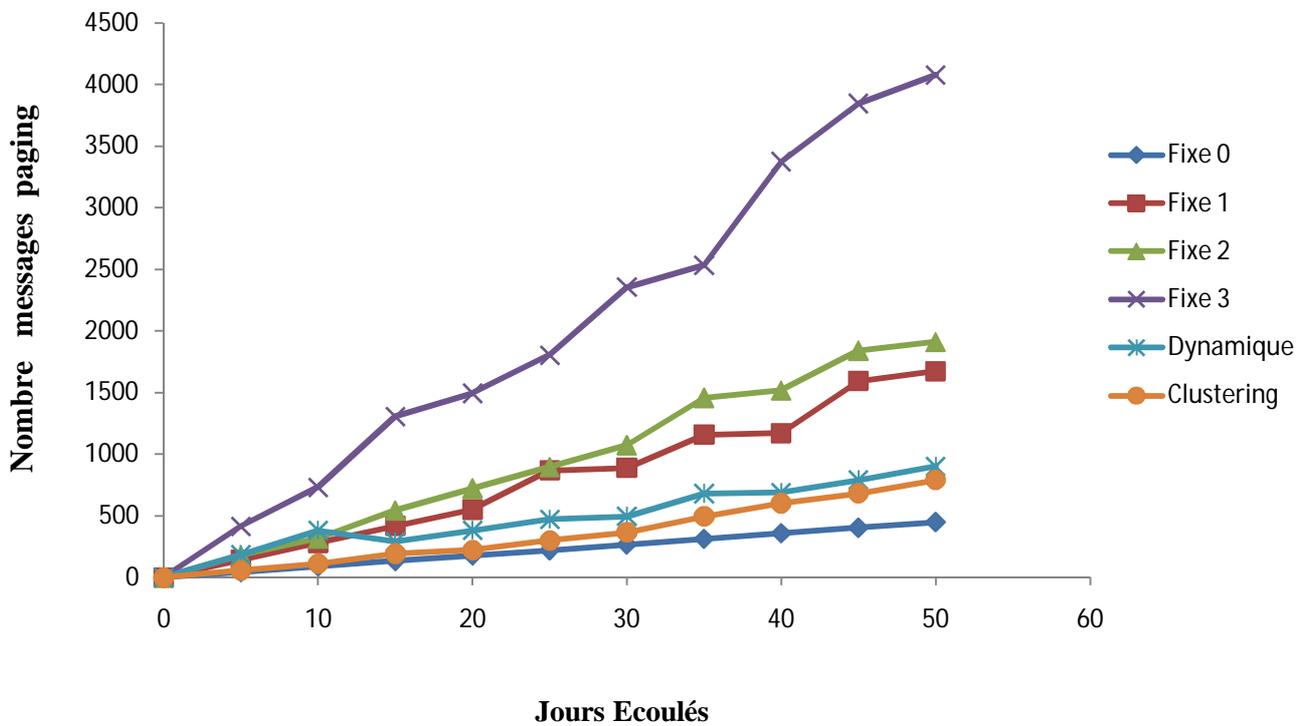


Fig V – 4 :- Nombre moyen des messages paging générés pour 9 appels entrants quotidiens

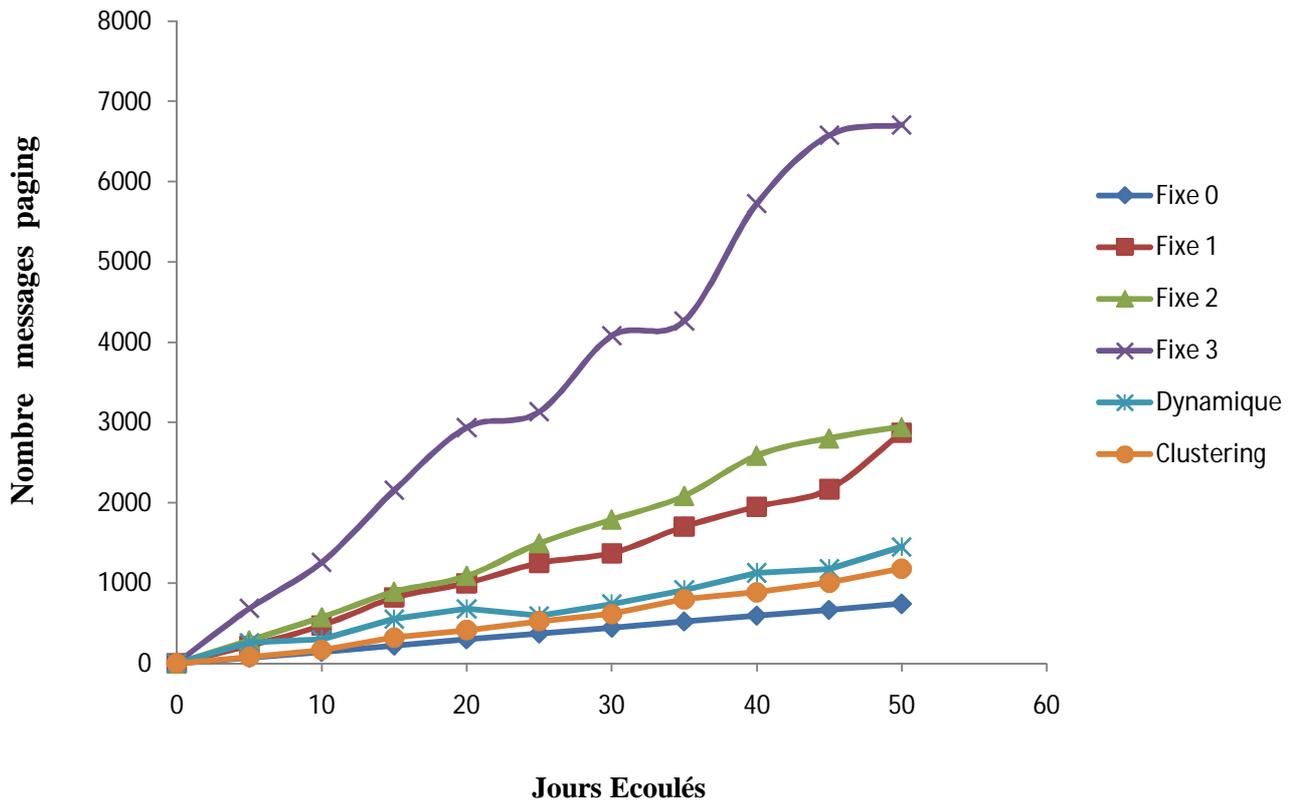


Fig V – 5 :- Nombre moyen des messages paging générés pour 15 appels entrants quotidiens

Le coût du paging est optimal pour le groupe fixe 0, vu que la cellule contenant l'abonné est toujours connue, et par conséquent, une seule cellule est pagée à chaque appel entrant. Néanmoins, ce coût optimal de paging est en général largement compensé par le grand coût de mise à jour. Inversement, le coût de paging du groupe fixe 3 est au minimal, dû aux zones de localisation relativement larges. Comme prévu, le coût du paging des groupes fixes 1 et 2 se situe entre les deux extrêmes, avec le groupe fixe 1 étant légèrement meilleur en raison de son petit nombre de cellules par zones de localisation.

Les coûts de paging du clustering sont quasiment identiques avec ceux du dynamique. Cela dit, les résultats de l'algorithme proposé s'améliorent à chaque niveau d'appels entrants en dépassant les résultats signés par le dynamique et en s'approchant petit à petit de la performance des résultats du groupe fixe 0.

#### **d) Coût total de la gestion de localisation**

L'algorithme du clustering proposé effectue les meilleurs résultats en termes de mise à jour, mais n'est pas à son optimal en termes de paging, en raison du compromis entre une mise à jour efficace et un paging efficace. Le coût total de gestion de localisation, défini comme une combinaison linéaire du nombre de messages de mise à jour et du nombre de messages de paging, est un des moyens qui permettent d'observer les effets globaux et comportement général des différents algorithmes.

Le coût est un concept relativement abstrait dans la gestion de localisation. Il peut être défini par plusieurs approches, comme par exemple le nombre total des signalisations générés par les messages, circulant sur interface spécifique, ou la taille totale en bits des messages échangés sur le réseau. Quelques auteurs représentent simplement le coût par une constante abstraite. Dans cette analyse, une approche intermédiaire et non reposante sur une seule technique a été suivie. Le coût total de gestion de localisation est donné par l'équation suivante [11]:

$$\text{Coût total} = c * N_{lu} + N_{pg}$$

Où :  $N_{lu}$  est le nombre de messages de mise à jour générés,  $N_{pg}$  est le nombre de messages paging générés pour un appel entrant, et  $c$  est une constante représentant le coût relatif d'une mise à jour vis-à-vis d'un paging. Vu que la mise à jour est la procédure la plus coûteuse en termes de signalisation par rapport au paging, les deux valeurs représentatives sont 5 et 10. Ces valeurs sont en quelque sorte arbitraires, mais représentent la taille et le nombre approximatif des messages requis pour une mise à jour, comparé à un message paging.

Des différences significatives dans le coût relatif de gestion de localisation des différents algorithmes surviennent avec l'augmentation du nombre des appels entrants. Pour  $c=5$ , et pour un niveau d'appels entrants assez bas égal à 3 (Fig V-5), le groupe fixe 0 signe le plus grand coût. Ceci est dû au coût excessif de la mise à jour requis par cet algorithme, ce qui n'est pas compensée par son coût de paging essentiellement faible, puisqu'il ya peu d'appels entrants. Le groupe fixe 1 et le groupe fixe 2 ont des performances similaires avec un coût qui augmente de manière proportionnelle au nombre d'appels entrants.

La performance du groupe fixe 3, qui est assez bonne au départ, décline rapidement avec l'augmentation du nombre d'appels entrants, de sorte que pour les 9 et 12 appels entrants par jour, sa performance est nettement mauvaise par rapport aux autres algorithmes. Ceci était prévisible, vu qu'il a signé le plus mauvais résultat en termes de paging, ceux qui forme une proportion importante du coût total de gestion de localisation pour  $c=5$ .

L'algorithme dynamique signe, au fil des temps, les meilleurs résultats par rapport aux autres algorithmes fixes. Malgré que ses coûts de paging soient similaires à ceux des groupes fixes 1 et 2, le faible coût de la mise à jour compense ces coûts de paging. Cependant, l'algorithme proposé basé sur le clustering surpasse en coût et performance tous les algorithmes précédents et ses coûts de signalisation restent assez stables, comme le montre la courbe du graphe.

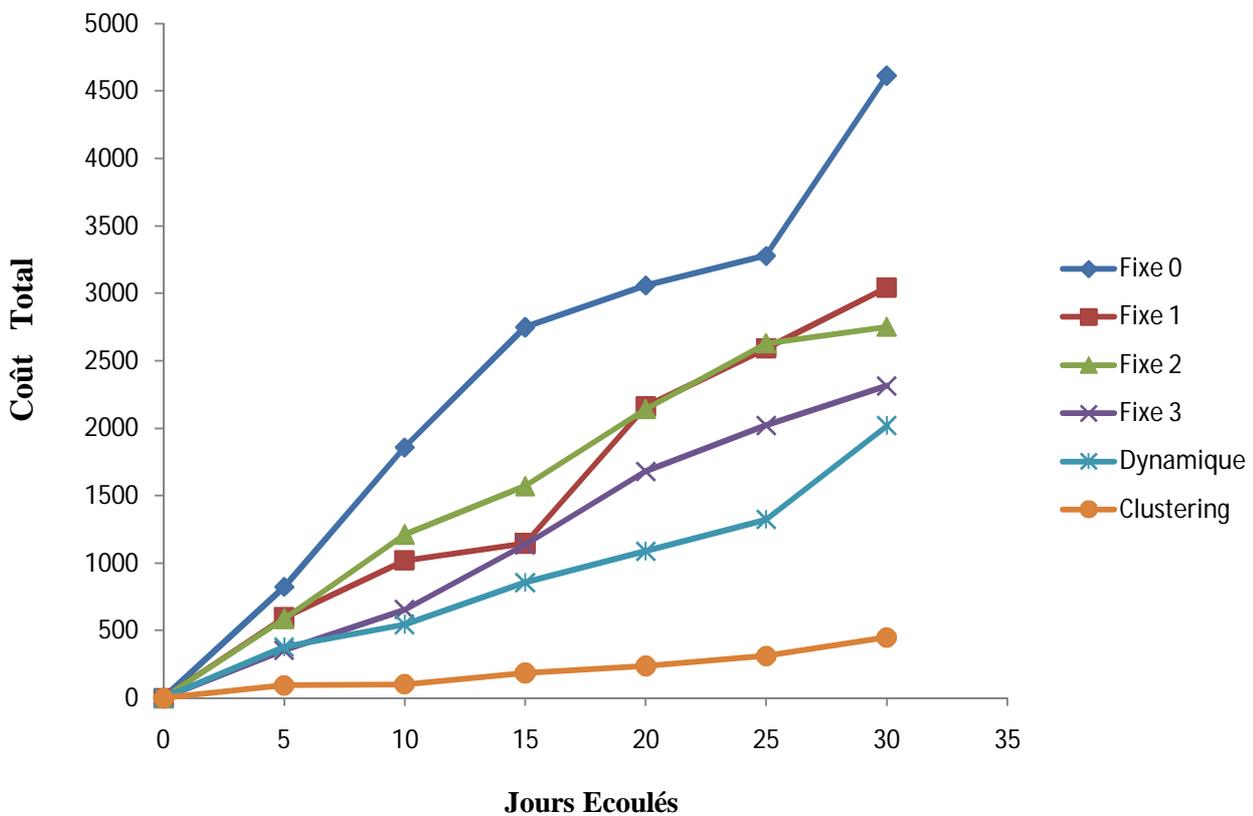


Fig V – 6 :- Coût total de la gestion de localisation ( $c=5$ ) pour 3 appels entrants

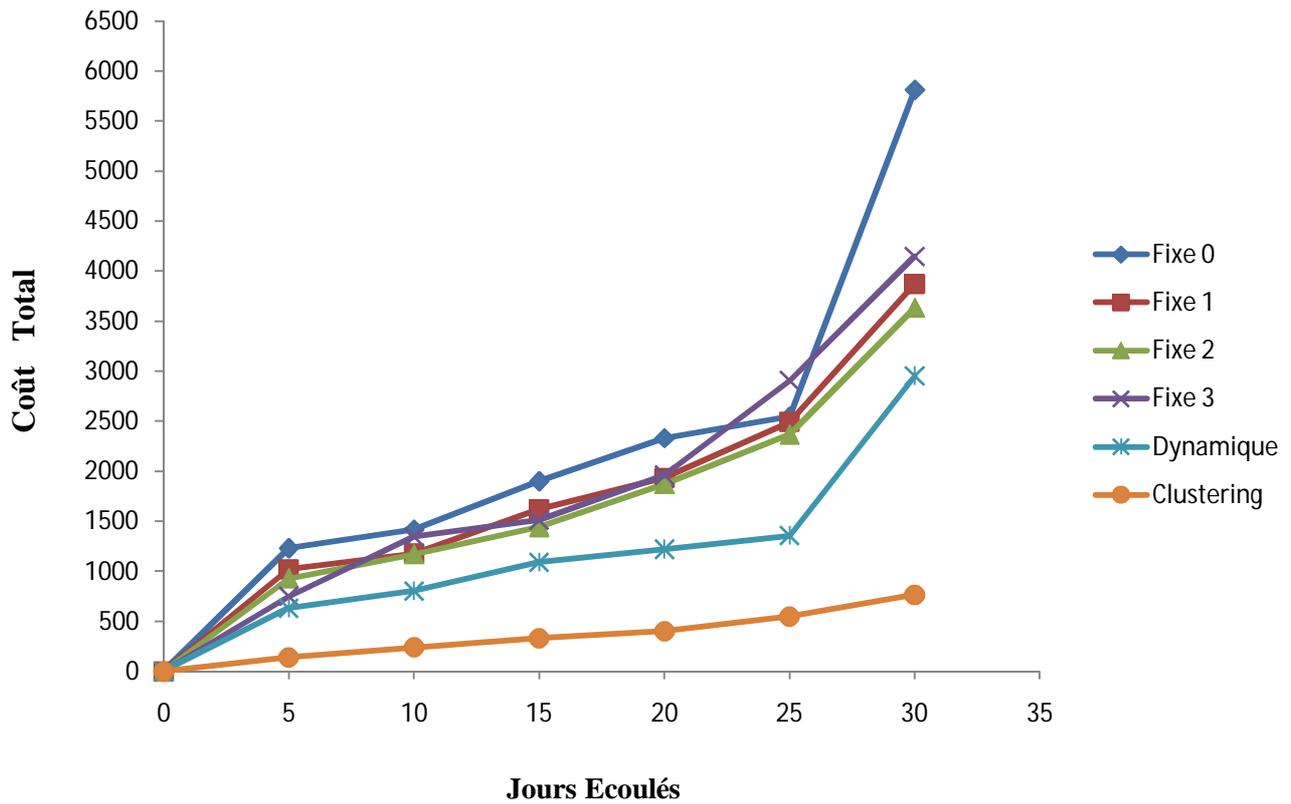


Fig V – 7 :- Coût total de la gestion de localisation (c=5) pour 9 appels entrants

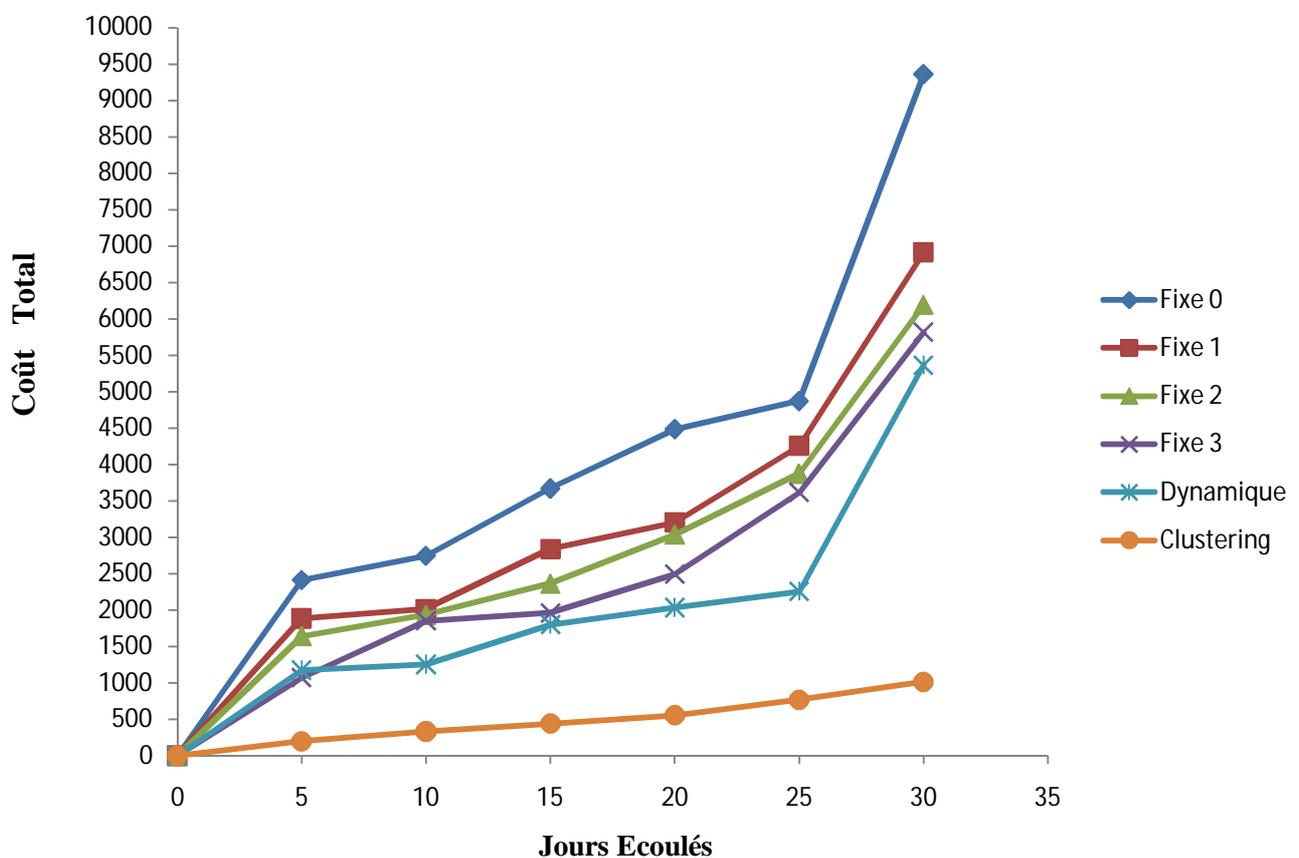


Fig V – 8 :- Coût total de la gestion de localisation (c=10) pour 9 appels entrants

Pour  $c = 10$ , le groupe fixe 0 signe le plus mauvais résultat pour tous les niveaux d'appels entrant, à savoir 3, 9 et 12. Ceci est dû à une mise à jour très coûteuse. Comme précédemment, les groupes fixes 1 et 2 se comportent identiquement. Le groupe fixe 3, avec ses larges zones de localisation et son faible coût de mise à jour, voit ses résultats se dégringoler au fur et à mesure que le nombre d'appels entrants augmente. Ceci est dû au coût de paging qui s'élève proportionnellement avec le nombre d'appels entrants.

Même remarque qu'avant concernant l'algorithme dynamique et le clustering. Notre proposition est la plus performante et la moins coûteuse de toutes. L'écart entre le coût total des algorithmes fixes et dynamique et de l'algorithme du clustering est encore plus grand pour  $c = 10$ , car l'efficacité de la mise à jour de l'algorithme proposé devient encore plus apparente.

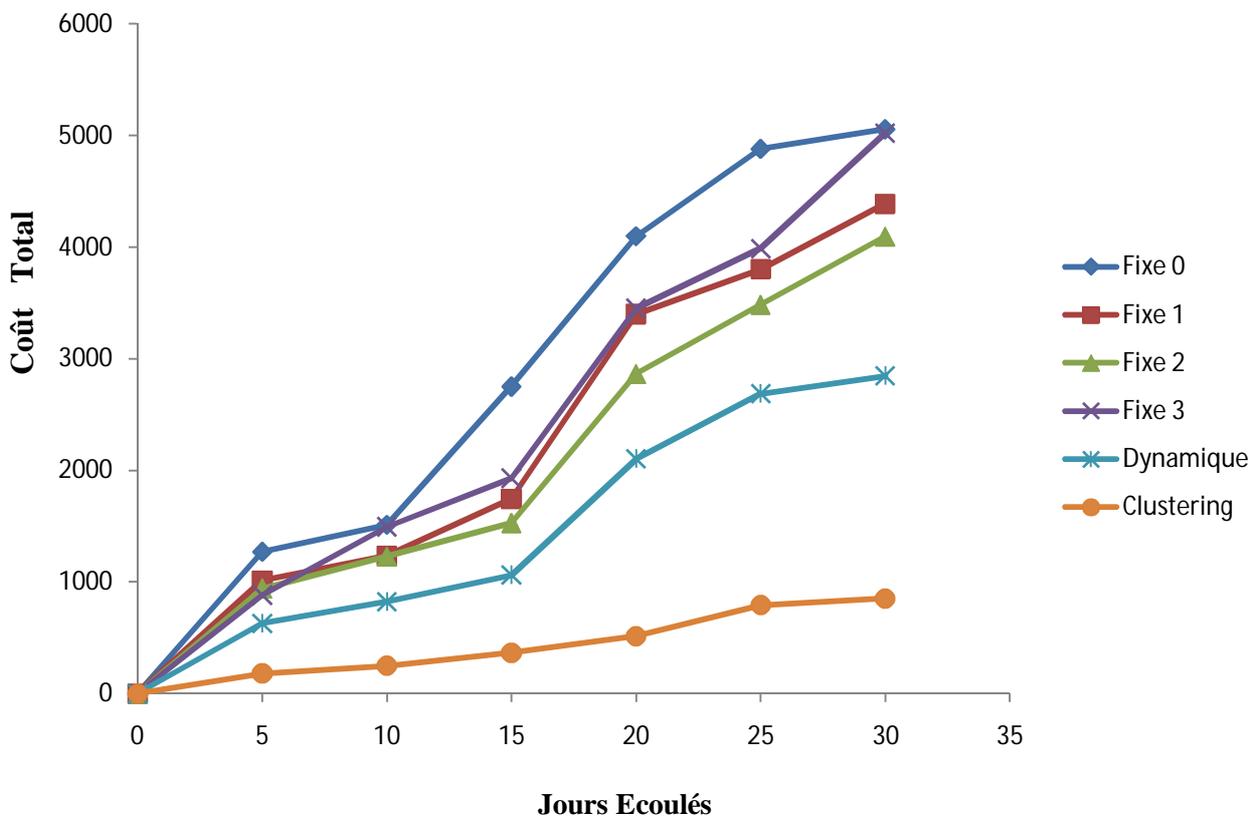


Fig V – 9 :- Coût total de la gestion de localisation ( $c=5$ ) pour 12 appels entrants

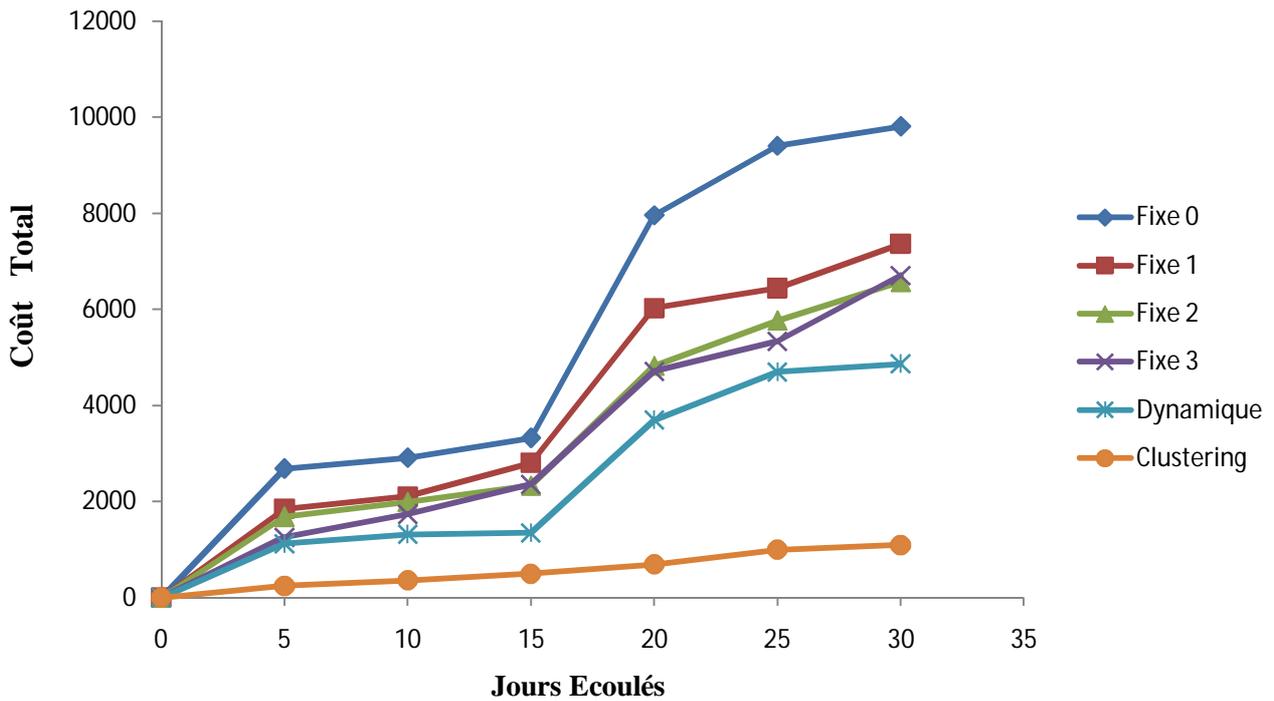


Fig V – 10 :- Coût total de la gestion de localisation ( $c=10$ ) pour 12 appels entrants

#### 4) Conclusion

A travers ces deux derniers chapitres, nous avons explicité les détails concernant la réalisation et l'implémentation de notre algorithme de gestion de localisation basé sur le clustering ainsi que du modèle de mobilité réaliste utilisé pour évaluer notre proposition. Nous avons en outre, grâce à des multiples exécutions, effectué une analyse comparant plusieurs algorithmes déjà proposés dans la littérature avec celui proposé dans ce mémoire. Ce dernier a signé les meilleurs résultats en termes de coût total de gestion de localisation.

En effet, l'algorithme proposé permet de déterminer la taille optimale d'une zone de localisation et de trouver le modèle optimal de regroupement de cellules en zones de paging. Ceci garantit l'atteinte du minimum global de la fonction de coût total en tenant compte des coûts de miss à jour  $N_{lu}$ , de paging  $N_{pg}$  et du rapport mise à jour/paging  $c$ .

# Conclusion et perspectives

---

**D**eux sujets distincts mais connexes ont été abordés dans les chapitres précédents. Le premier concerne l'algorithme de gestion de localisation proposé, qui utilise le profil utilisateur pour construire automatiquement, à base du clustering, les zones de localisation et les zones de paging. En utilisant le nombre de fois que l'abonné traverse une cellule à un instant donné, l'algorithme construit les zones de localisation individuellement et propres à chaque abonné. Un historique concernant la durée moyenne passée sur chaque cellule traversée est utilisé pour déterminer l'ensemble des cellules de la zone de localisation courante qui sont les plus susceptibles de contenir l'abonné. Ces zones sont appelés les zones de paging et sont pager en premier.

Le deuxième point concerne le modèle de mobilité réaliste utilisé pour tester et évaluer notre travail. Ce modèle est conçu pour simuler des comportements réalistes de mobilité, basé sur les activités dérivés statistiquement en utilisant les données d'une enquête de voyage réalisée précédemment. Un emplacement et une durée particulière sont associés pour chaque activité. Ainsi, un déplacement ou un voyage consiste du chemin entre les emplacements des activités adjacentes. Le modèle de mobilité est adapté aux abonnés d'un réseau cellulaire, en représentant ses chemins et ses emplacements par des cellules radio.

## **Evaluation et contribution**

***Minimiser les signalisations, sur l'interface radio et sur le réseau, afin de réduire les signaux de la bande passante et ainsi d'accroître sa qualité de service.***

Comme détaillé dans le chapitre précédent, l'algorithme proposé surpasse, en termes de coût total de signalisation, tous les autres algorithmes fixes et dynamiques, pour presque tous les types d'abonnés définis, les niveaux d'appels entrants, la variabilité de l'emplacement du travail, et la taille des zones de localisation dynamiques. L'algorithme proposé basé sur le clustering permettrait de réduire considérablement les coûts de gestion de localisation par rapport aux techniques actuellement utilisées. En raison de la nature des variables et le nombre de paramètres impliqués, une mesure exacte de l'amélioration serait plutôt insensée, mais des améliorations significatives peuvent être observées dans presque toutes les analyses effectuées dans le chapitre 6.

***Minimiser les temps d'attente lors d'un appel sortant, dû spécialement à la procédure de paging, dans le but de maintenir une bonne qualité de service.***

Dans la majorité des approches de gestion de localisation (vue au chapitre 2), plusieurs zones de paging sont pagée successivement à travers des zones de localisation larges histoire de localiser l'abonné. En raison des retards impliqués dans le paging (attente d'une réponse, paging répété ... etc.), en répétant le processus entier depuis plus d'une ou deux zones de paging entraînerait des temps d'attente inacceptables, en termes de qualité de service. Par contre, dans l'algorithme proposé, une seule zone de paging est pagée. Ceci diminue considérablement les signalisations dû au paging.

***Maximiser l'applicabilité des algorithmes de gestion de localisation.***

Certains algorithmes de gestion de localisation suggèrent que les abonnés soient classés dans des classes utilisateur, comme une classe contenant les abonnés très mobile, ou une classe pour piétons. Sauf si une méthode peut être utilisée pour mettre à jour ce statut, il s'agit d'un mécanisme inefficace car les classes utilisateur changent pour chaque abonné

tout au long de la journée (par exemple, très mobile lors qu'il est dans un véhicule, immobile au travail ou à domicile, ou un piéton en faisant du shopping).

L'algorithme proposé ne pas faire d'hypothèses sur le statut de mobilité d'un abonné. Le profil utilisateur associé à chaque abonné fournit une méthode relativement efficace pour stocker les informations de gestion de localisation.

***Minimiser la complexité des calculs de l'algorithme et la quantité de données stockés pour chaque abonné, pour réduire les coûts.***

En donnant les économies de notre proposition en termes de coûts de gestion de localisation, l'algorithme du clustering exige une certaine logique additionnelle et de la mémoire. En termes de calcul, l'algorithme proposé requiert d'enregistrer le temps moyen passé dans chaque cellule, et le nombre de transitions d'une cellule à l'autre. Ceci exige seulement un timer (et un ensemble d'opérations élémentaires pour mettre à jour la durée moyenne précédente) et un ensemble de compteurs.

En outre, le processus de comparaison en entrant dans une nouvelle cellule est plus complexe, nécessitant la comparaison d'un identificateur de cellule contre une liste d'identificateurs de cellules (la zone de localisation personnelle), par opposition à la comparaison d'un identificateur de zone de localisation (LAI) contre le LAI courant. Un autre coût supplémentaire de l'algorithme proposé est l'enregistrement du profil utilisateur sur le node B.

## **Perspectives**

La portée du modèle de mobilité et de l'algorithme proposé est assez grande, et plusieurs possibilités et les alternatives ont été couvertes partiellement ou pas du tout.

La notion de vieillissement des données du profil utilisateur a probablement besoin d'être pris en compte dans une mise en œuvre effective, afin de réduire le temps requis par l'algorithme pour s'adapter aux changements dans le comportement des abonnés. Par exemple, considérons un abonné qui déménage dans une nouvelle maison (nouvelle cellule). La cellule de la maison ancienne aurait une grande durée moyenne ( $T_a$ ), et sera donc très probablement pager en premier si elle est incluse dans la zone de localisation. Il en est de même pour  $N(a,b)$ . Dans les deux cas, le comportement de l'algorithme peut être différent de celui souhaité pour une gestion de localisation efficace. Le vieillissement des données du profil utilisateur réduira progressivement l'influence des cellules précédemment actives qui ne sont plus aussi importantes.

Coté algorithme de gestion de localisation, il serait intéressant d'utiliser une approche de prédiction de mouvement qui permettra d'anticiper et de prévoir les déplacements de l'abonné, afin de déterminer les cellules qu'il va traverser plus tard. Ces cellules constitueront la nouvelle zone de localisation.

Dans l'analyse des propositions sélectionnées, on a supposé que les messages de mise à jour de localisation sont le principal contributeur au coût total de gestion de localisation et le nombre d'appels entrants est considéré comme fixe et aléatoire pour chaque série de test. Comme le témoigne l'analyse du cinquième chapitre, les appels entrants influencent à leur tour sur la performance des systèmes de gestion de localisation, et en particulier lorsque l'abonné n'est pas très mobile ou reçoit beaucoup d'appels au cours d'itinérance. Il serait donc intéressant d'utiliser des données réalistes concernant le nombre d'appels entrants afin de bien évaluer l'algorithme sélectionné.

# Références

---

- [1] Javier Sanchez et Mamadou Thioune. ISTE Ltd. *UMTS*. 2007. Support en ligne : books.google.com.
- [2] B. Walke, P.Seidenberg et M.P Althoff. John Wiley & Sons Ltd. *UMTS The fundamentals*. 2003. Support en ligne: books.google.com.
- [3] Heikki Kaaranen, Ari Ahtiainen, Lauri Laitinen, Siamak Naghian et Valteri Niemi. *UMTS Networks : Architecture, Mobility and Services, second edition*. 2005. Support en ligne : mohandesoon.com library. Disponible sur : [http://www.mohandesoon.com/library/electric/Wiley\\_UMTS\\_Networks\\_Architectu.pdf](http://www.mohandesoon.com/library/electric/Wiley_UMTS_Networks_Architectu.pdf)
- [4] Pierre Lescuyer. Editions Dunod. *UMTS : Les origines, l'architecture, la norme*. 2001. Support en ligne : books.google.com.
- [5] Harri Holma et Antti Toskala. Editions Osman Eyrolles Multimedia. *UMTS : Les réseaux mobiles de 3<sup>ème</sup> génération*. 2000. Support en ligne : books.google.com.
- [6] Samuel Pierre. Presses Internationales, Ecole Polytechnique Montreal. *Réseaux et systèmes informatiques mobiles: Fondements, architectures et applications*. Avril 2003. Support en ligne : books.google.com.
- [7] Michael J. A. Berry et Gordon Linoff. *Data Mining, Techniques appliqué au marketing, à la vente et aux services clients*. 1997. Support en ligne : books.google.com.
- [8] René Lefébure et Gilles Venturi. Edition EYROLLES. *Le Data Mining*. Deuxième tirage 1998. Support en ligne : books.google.com

**[9]** Hakim Badis. *Etude Et Conception D'algorithmes Pour Les Réseaux mobiles Et AD HOC*. Thèse Doctorale. Université Paris-Sud 11. Support en ligne : [igm.univ-mlv.fr](http://igm.univ-mlv.fr). Disponible sur : <http://www-igm.univ-mlv.fr/~badis/these/Badis-these.pdf>

**[10]** Frédéric Evennou. 2007. *Techniques et technologies de localisation avancées pour terminaux mobiles dans les environnements indoor*. Thèse doctorale. Université Joseph Fourier, Grenoble I. Support en ligne : [hal.archives-ouvertes.fr](http://hal.archives-ouvertes.fr). Disponible sur : [http://hal.archives-ouvertes.fr/docs/00/13/60/64/ANNEX/Soutenance\\_These\\_v21.pdf](http://hal.archives-ouvertes.fr/docs/00/13/60/64/ANNEX/Soutenance_These_v21.pdf)

**[11]** John Scourias. 1997. *Dynamic Location Management And Activity-Based Mobility Modelling For Cellular Networks*. Mémoire de maitrise. Département informatique, université de Waterloo, Ontario, Canada. Support en ligne: [uwspace.uwaterloo.ca](http://uwspace.uwaterloo.ca) <http://uwspace.uwaterloo.ca/bitstream/10012/30/1/mq21539.pdf>

**[12]** Atif Akram Siddiqi. 1999. *Location Management Proposals: Survey And Evaluation*. Mémoire de maitrise. Département Informatique, université Carleton, Ottawa, Ontario, Canada. Support en ligne : [collectionscanada.gc.ca](http://collectionscanada.gc.ca) . Disponible sur : [http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape8/PODD\\_0006/MQ43363.pdf](http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape8/PODD_0006/MQ43363.pdf)

**[13]** Irina Kozatchok. 1999. *Dépistage d'unités mobiles dans les systèmes cellulaires*. Mémoire de maitrise en science appliquée. Département de génie électrique et de génie informatique, école polytechnique de Montréal. Support en ligne : [collectionscanada.gc.ca](http://collectionscanada.gc.ca) [http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape7/PODD\\_0019/MQ48859.pdf](http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape7/PODD_0019/MQ48859.pdf)

**[14]** Emmanuel TONYE et Landry EWOUSSOUA. *Planification et Ingénierie Des Réseaux De Telecoms, Gestion De La Mobilité*. Mémoire de maitrise. Ecole Nationale Supérieure Polytechnique, Université de Yaoundé I. Support en ligne : [foad.refer.org](http://foad.refer.org). Disponible sur : [http://foad.refer.org/IMG/pdf/4-Gestion\\_mobilite.pdf](http://foad.refer.org/IMG/pdf/4-Gestion_mobilite.pdf)

**[15]** Menouer Tarek et Dermouche Mohamed. 2009/2010. *Application de techniques de data mining pour la classification automatique des données et la recherche d'associations*. Mémoire de fin d'étude. Ecole nationale Supérieure d'Informatique, Alger. Support en ligne : [share.esi.dz](http://share.esi.dz)

**[16]** Nicolas BECK. 2005. *Application de méthodes de clustering traditionnelles et extension au cadre multicritère*. Mémoire d'ingénieur. Faculté des sciences appliquées, Université libre de Bruxelles.

**[17]** Pierre Rouillet. 20/11/2008. *La 3G : De l'UMTS à La 4G*. Cours. Support en ligne : pierre.rouillet.perso.sfr.fr. Disponible sur : <http://pierre.rouillet.perso.sfr.fr/3G4G.pdf>

**[18]** *Réseaux D'accès UMTS : Architecture et Interfaces*. Support en ligne : efort.com. Disponible sur : [http://www.efort.com/r\\_tutoriels/ACCES\\_UMTS\\_EFORT.pdf](http://www.efort.com/r_tutoriels/ACCES_UMTS_EFORT.pdf)

**[19]** Xavier Lagrange. Décembre 2004. *Réseaux mobiles 2G et 3G*. Cours. ENST Bretagne, Ecotel. Support en ligne : <ftp://138.96.0.43>. Disponible sur : <ftp://138.96.0.43/mascotte/David.Coudert/ecotel/2004/Cours/cb1-reseaux2G-lagrange-ecotel04-final.pdf>

**[20]** Samuel Pierre et Max Maurice. 2008. *Introduction Aux Réseaux mobiles*. Présentation pour Geninov Inc. Support En ligne : conatel.gouv.ht. Disponible sur : [http://www.conatel.gouv.ht/portabilite/Partie%20Introduction%20aux%20reseaux%20mobiles\\_new-Y.pdf](http://www.conatel.gouv.ht/portabilite/Partie%20Introduction%20aux%20reseaux%20mobiles_new-Y.pdf)

**[21]** Eric Meurisse. Février 2007. *UMTS et le haut-débit mobile*. Rapport. Support en ligne : igm.univ-mlv.fr. Disponible sur : [http://igm.univ-mlv.fr/~dr/XPOSE2006/eric\\_meurisse/pdfs/rapport\\_umts\\_eric\\_meurisse.pdf](http://igm.univ-mlv.fr/~dr/XPOSE2006/eric_meurisse/pdfs/rapport_umts_eric_meurisse.pdf)

**[22]** Xavier Lagrange. 2002. *Gestion des ressources radio dans les réseaux mobiles*. Cours. ENST Bretagne. Support en ligne : <ftp://138.96.0.43>

**[23]** Mohamed Moussaoui. *L'UMTS et les réseaux de 3<sup>ème</sup> Génération*. Cours. Ecole nationale des sciences appliquées, Université Abdelmalek Essaadi, Tanger.

**[24]** John Scourias et Thomas Kunz. *A Dynamic Individualized Location Management Algorithm*. Département informatique, université de Waterloo, Ontario, Canada. Support en ligne: big.uwaterloo.ca. Disponible sur : big.uwaterloo.ca/publications/pdfs/conf22.pdf

**[25]** Kardi Teknomo. Décembre 2004. *K-Means Clustering Tutorial*. Support en ligne: Kardi Teknomo's Page. Disponible sur : <http://people.revoledu.com/kardi/tutorial/kMean/>

**[26]** M. Gettler – Summa et C. Pardoux. *La classification automatique*. Support de cours. Ecole Doctorale de gestion, université Paris-Dauphine. Support en ligne : ceremade.dauphine.fr. Disponible sur :

<http://www.ceremade.dauphine.fr/~touati/EDOGEST-seminaires/Classification.pdf>

**[27]** Riadh Ben Messaoud. Avril 2007. *Data Mining (Fouille De Données)*. Support Licence C.E.STAT, Institut Universitaire de technologie lumière, Lyon. Support en ligne : eric.univ-lyon2.fr. Disponible sur :

[http://eric.univ-lyon2.fr/~rbenmessaoud/supports/dm\\_fsegn/1/dm\\_s1.pdf](http://eric.univ-lyon2.fr/~rbenmessaoud/supports/dm_fsegn/1/dm_s1.pdf)

**[28]** U. Madhow, M. L. Honig and K Steiglitz. June 1994. *Optimization of Wireless Resources for Personal Communications Mobility Tracking*. Support en ligne: cs.princeton.edu.

Disponible sur: <http://www.cs.princeton.edu/~ken/OptWireless.pdf>

**[29]** Naoki Matsumaru, Kshirasagar Naik and David S.L Wei. *Comparing three location management strategies for tracking mobile systems*. International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, NV, USA, 13 - 16 Jul 1998, pp.771-777.

**[30]** Edouard Kutter. 1973. *A model for individual travel behaviour*. Urban Studies.