La 4G l'interopérabilité entre les réseaux

Fondé sur un cœur de réseau IP, le futur système de télécommunication 4ème génération (4G) représente la convergence entre le réseau de 3ème génération (type UMTS) et les diverses technologies radio complémentaires. L'objectif est de fournir aux utilisateurs des services sans interruption dans un environnement hétérogène. Les utilisateurs mobiles peuvent accéder ou échanger des informations indépendamment de leur position, du temps et en utilisant des équipements différents en terme de capacité. Les réseaux 4G sont des réseaux hybrides qui intègrent différentes topologies et plate-formes réseaux.

Il existe deux types d'intégration : l'intégration des différents types de réseaux sans fil hétérogènes (wireless LAN, WAN, PAN ainsi que les réseaux ad hoc) ; et l'intégration des réseaux sans fil avec l'infrastructure fixe, l'Internet et le réseau téléphonique fixe. Cependant, beaucoup de travaux demeurent pour permettre une intégration sans couture.

La gestion de la mobilité est une opération fondamentale pour n'importe quel réseau mobile. Bien que les fonctionnalités et les implémentations varient entre les diverses technologies, certaines caractéristiques de base sont communes (voir section 3.1). Dans cette thèse, nous évoquons la nécessité d'une structure hiérarchique des réseaux sans fil hétérogènes, afin de mieux gérer la mobilité entre les réseaux hétérogènes. Cette structure est virtuelle et nécessite de nouvelles composantes et fonctionnalités pour gérer la communication entre les réseaux hétérogènes. Nous prenons comme exemple, l'interconnexion UMTS/WLAN afin de montrer les techniques réelles d'intégration. Ensuite, nous supposons que l'intégration est faite et nous nous focalisons sur les transferts cellulaires entre les différentes technologies des réseaux sans fil tout en tenant compte de leurs diversités (débit, latence, zone de couverture, etc).

3.1 Les principes de la gestion de mobilité dans les réseaux sans fi l et mobiles

L'objectif principal de la gestion de mobilité est de maintenir des informations sur la position des terminaux mobiles et de gérer leurs connexions lorsqu'ils se déplacent dans les zones de couvertures. La gestion de la mobilité inclut deux procédures [28] : la gestion de la localisation et la gestion des handovers. La gestion de la localisation permet de fournir au réseau des informations sur la position courante d'un terminal mobile. Cette fonctionnalité comprend : i) *le processus d'inscription de la localisation,* où le terminal mobile est authentifié et sa position est

mise à jour, ii) *le paging*, où la position du terminal mobile est recherchée pendant l'initialisation d'une nouvelle session.

Le handover est le processus par lequel une communication établie est maintenue alors que le terminal mobile se déplace à travers le réseau cellulaire ; elle implique que la communication puisse passer d'un canal physique à un autre avec une coupure sans conséquences (en moyenne < 100 ms pour une communication voix dans le GSM).

Dans un réseau cellulaire, la station de base (BS : *Base Station*) constitue le point d'accès de toutes les communications. Elle remplit le rôle de serveur pour les clients se trouvant dans sa cellule. Lorsqu'un utilisateur mobile quitte sa cellule pour entrer dans une autre cellule, il peut être amené à s'approcher d'une BS voisine. Pour poursuivre sa communication, cet utilisateur se voit contraint de changer de point d'accès. Il effectue alors ce que l'on appelle un *transfert intercellulaire*, ou *handover* ou encore *handoff*, qui consiste à demander à un gestionnaire du réseau, ou commutateur, de mettre en place la signalisation nécessaire au transfert. Ce type du handover se passe entre deux stations de bases qui utilisent la même technologie sans fil.

Le processus du handover peut être devisé en trois étapes : *initiation*, *décision* et *exécution* [29]. La phase d'initiation du handover est déterminée par des conditions spécifiques comme la qualité du signal, la disponibilité d'un point d'accès alternatif, etc.

La décision de s'attacher à un nouveau point d'accès peut être prise de trois manières différentes : un handover contrôlé par le réseau, un handover assisté par le terminal mobile et un handover contrôlé par le terminal mobile [30, 31]. Dans les handovers contrôlés par le réseau, c'est le réseau qui décide selon les mesures effectuées sur les signaux radio envoyés par les terminaux mobiles. Puisque cette solution est complètement centralisée, elle nécessite une puissance considérable de calcul. Elle est aussi pénalisée par l'absence d'une connaissance sur les conditions récentes de chaque terminal.

Afin d'épargner le réseau de la complexité de la tâche, le handover assisté par le terminal mobile permet au terminal d'effectuer les mesures et au réseau de prendre les décisions.

Dans un handover contrôlé par le terminal mobile, le terminal possède l'autorité et l'intelligence pour décider selon ses propres mesures. Cette solution a l'avantage d'être distribuée mais elle a des impacts sur la stabilité et la sécurité du réseau.

L'exécution d'un handover exige un échange de signalisation pour rétablir la communication et ré-acheminer les paquets de données via le nouveau point d'attachement.

3.2 Une structure hiérarchique des réseaux sans fil

Les technologies actuelles des réseaux sans fils varient fortement selon la bande passante offerte, le temps de latence, la fréquence utilisée, les méthodes d'accès, la puissance de consommation, la mobilité, etc. En tenant compte de l'hétérogénéité, l'ensemble de ces technologies peut être divisé en deux catégories : (i) ceux qui offrent des services avec un faible débit mais sur une large zone géographique ; (ii) ceux qui fournissent des services avec un débit important en couvrant des petits secteurs. Aucune technologie sans fil ne peut garantir toute seule un débit adaptable pour chaque application, nombreux services de donnée avec une connexion sans coupures, un nombre important d'utilisateurs sur une large zone de déplacement et une latence minimale lors de changement des cellules. Un terminal mobile équipé d'une interface WLAN se trouve dans un milieu restreint en taille (entreprise ou environnement personnel), il ne peut alors pas répondre aux besoins de bande passante de certaines applications dans le cas où il possède une interface WWAN.

Une solution possible consiste à faire coexister et combiner des technologies différentes de

réseaux sans fil pour obtenir une meilleure couverture et si possible un choix dynamique de réseau selon le type d'application. Dans ce cas, le terminal mobile est doté de plusieurs interfaces lui permettant de se déplacer librement et d'accéder aux infrastructures des réseaux filaires à travers des réseaux sans fil alternatifs. Prenons l'exemple d'un utilisateur possédant un ordinateur potable (*laptop*) ou un PDA (Assistant numérique personnel) connecté à Internet dans son bureau via une technologie infrarouge (permet de relier deux dispositifs dotés de ports à infrarouge). Lorsqu'il quitte son bureau avec sa machine, il se trouve dans le couloir de son entreprise avec une connexion Wi-Fi. Une fois que ce même utilisateur se déplace en dors de son entreprise, il continue à se connecter via une technologie WWAN (UMTS par exemple). Lorsqu'il rentre chez lui, il trouve une connexion Wi-Fi, etc. Cette combinaison des interfaces sans fil conduit à une hiérarchie appelée *structure hiérarchisée des réseaux sans fil* (OWN : Overlay Wireless Network structure). La figure 3.1 montre un exemple d'une telle structure. Les niveaux les plus bas permettent une couverture limitée en terme de portée mais des débits importants. Les niveaux les plus hauts couvrent un nombre important d'utilisateurs sur une large zone géographique mais limitée en bande passante.

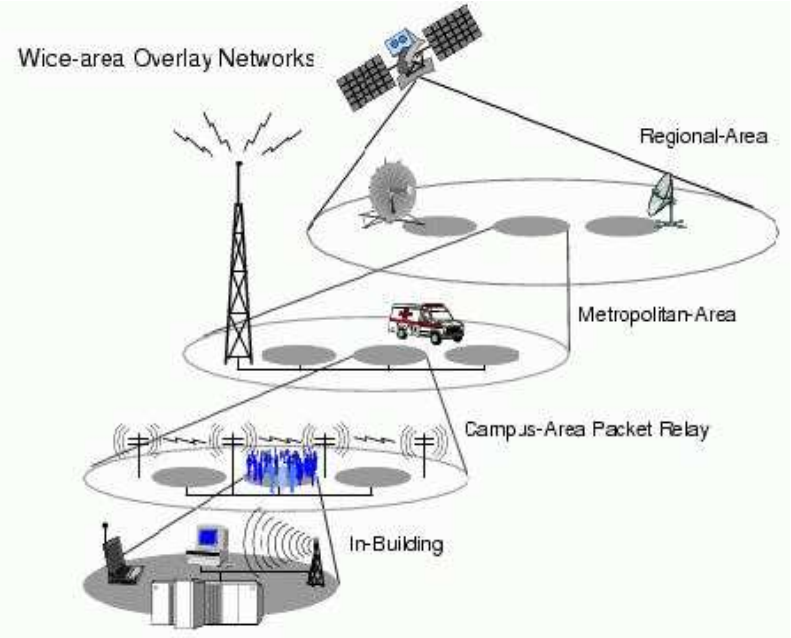


FIG. 3.1 – *Un exemple d'une structure hiérarchique des réseaux sans fil.*

Un bureau, un étage, une entreprise, un campus, etc, représentent en réalité des cellules de tailles différentes. Dans GSM, pour palier le manque de capacité réseau à l'explosion du nombre des abonnés, les opérateurs ont recours à différentes tailles de cellules : les macrocellules, les microcellules et les picocellules. Les macrocellules sont les plus courantes dans un réseau cellulaire GSM. Leur zone d'action s'étend jusqu'à 30 km selon les obstacles rencontrés. Les microcellules couvrent quelques rues d'un centre ville ou une station de métro (portée maximale de 500 m). Les picocellules couvrent un étage d'un grand bâtiment ou d'un centre commercial (portée maximale 100 m). Ces trois types de cellules permettent aux opérateurs de créer un réseau multicouches leur offrant ainsi plus de flexibilité pour augmenter la capacité de leur réseau. L'UMTS rajoute au sommet de cette structure multi-couches des satellites qui assurent une couverture sur l'ensemble

de la planète.

Dans une OWN, un ensemble de cellules de petite taille d'un niveau i sera regroupé dans une cellule de taille plus grande de niveau $i+1$. Un exemple de regroupement de cellules dans une OWN est montré dans la figure 3.1. Il n'existe pas vraiment de classification formelle selon la taille des cellules, mais en général on peut distinguer cinq types utilisés par la communauté de communications : picocellules, nanocellules, microcellules, macrocellules, et satellites. Les cellules utilisées dans une OWN offrent plusieurs technologies différentes des réseaux mobiles et sans fil (UMTS, GSM, GPRS, IEEE 802.11, Blutooth, Infrarouge, etc). Les caractéristiques de chaque type de cellule dans cette classification sont montrées dans la Table 3.1.

Niv.	Classification	Diamètre	débit	Latence	Mobilité	Application
0	Picocells	10m	1-50 Mbps	$\ll$ 10 ms	piéton	bureau
1	Nanocells	100m	1-2 Mbps	< 10 ms	piéton	bureau/étage
2	Microcells	500m	100 Kbps	100 ms	piéton/véhicule	bâtiment/campus
3	Macrocells	50km	19,2 Kbps	> 100 ms	véhicule	métropolitain
4	Satellite	> 50km	4,8 Kbps	$\gg$ 100 ms	véhicule	région

TAB. 3.1: Classification des réseaux mobiles et sans fil

3.3 Un exemple d'intégration : 3G/WLAN

La conception et l'intégration 3G/WALAN relèvent beaucoup de défis techniques. Comme l'Authentification, l'Autorisation et l'Accounting (AAA), la sécurité globale, la Qualité de Service (QoS) et les handovers. Malgré les progrès réalisés dans la standardisation de l'interconnexion WLAN/UMTS [32,33], la plupart des efforts ont été concentrés sur l'AAA plutôt que sur la gestion de mobilité et de QoS.

L'intégration devrait être faite par étapes. L'ensemble des étapes à suivre pour d'arriver à un schéma d'interconnexion *idéal* a été proposé par 3GPP (voir [33]). Cette proposition est basée sur les services demandés par les utilisateurs. À cet égard, six scénarios ont été spécifiés, chacun représente un type différent d'interconnexion et un niveau de satisfaction de l'utilisateur. Les scénarios sont : i) scénario 1, permet d'avoir seulement des clients et une facturation commune, ii) scénario 2, prévoit le contrôle d'accès basé sur le 3GPP, iii) scénario 3, permet l'accès aux services de 3GPP PSS (*Packet-Switched Streaming*) à partir de WLANs, iv) scénario 4, assure la continuité de service après l'exécution d'un handover inter-système, v) scénario 5, promet la fonctionnalité sans couture dans le scénario précédent, et enfin, vi) scénraio 6, permet d'accéder aux services du 3GPP CS (Circuit Switching) à partir de WLANs.

En se basant sur les scénarios ci-dessus, un certain nombre de propositions ont visé un certain degré d'intégration sans couture. Certaines d'entre elles se focalisent sur l'architecture générale et utilisent l'IP Mobile comme outil de base pour l'intégration des systèmes. Ceci permet une simple mise en œuvre avec des modifications légères aux composantes existantes de systèmes, mais en contre partie, un temps de latence de handover considérable. Ces propositions sont considérées comme des solutions *faiblement couplés* [32]. Bien qu'elles ne permettent pas d'accéder aux services 3GPP PS (Packet Switching), par exemple : les services WAP et MMS, à partir de WLANs, elles assurent la continuité des services (scénario 4), mais sans la fonctionnalité sans couture à tout moment (scénario 5). Les efforts de recherche dans cette catégorie de solutions se concentrent sur la conception des algorithmes d'initiation/décision pour les handovers afin d'offrir la meilleure

qualité aux utilisateurs avec une utilisation efficace des ressources.

Un autre groupe de propositions essaye d'intégrer UMTS et WLAN d'une manière plus serrée. Dans cette approche, le point d'interconnexion des deux technologies se situe soit dans le réseau cœur de l'UMTS (CN), soit dans le réseau d'accès (AN). Par conséquent, l'effort dans cette catégorie se focalise sur la résolution des problèmes techniques engendrés par les extensions ajoutées aux standards pour assurer l'interopérabilité entre l'UMTS et WLAN. La mobilité rapide et les fonctionnalités sans couture (scénario 5) sont les principales cibles de ces propositions sans oublier la complexité due aux améliorations introduites sur les composantes existantes. Ce groupe inclut les architectures *fortement couplées* [34–36] et *très fortement couplées* [37, 38].

3.4 Les handovers dans une OWN

Avec une OWN et d'après la figure 3.2, on distingue deux types de handovers : un handover horizontal (*horizontal handover*) et un autre vertical (*vertical handover*). Un handover horizontal est un handover classique entre deux cellules homogènes de même niveau hiérarchique dans la OWN. Par exemple, GSM vers GSM, WLAN vers WLAN, UMTS vers UMTS, UMTS vers GSM, etc. Les handovers horizontaux peuvent utiliser les algorithmes d'initiation des handovers montrés dans la section 3.1 et qui ont pour cause la mobilité des terminaux mobiles. Un handover vertical se déclenche lors du passage d'une cellule de niveau i à une autre de niveau $i+j$ ou $i-j$ avec $j>0$, c'est-à-dire entre deux cellules hétérogènes. Par exemple, UMTS vers WLAN, GSM vers Bluetooth, etc.

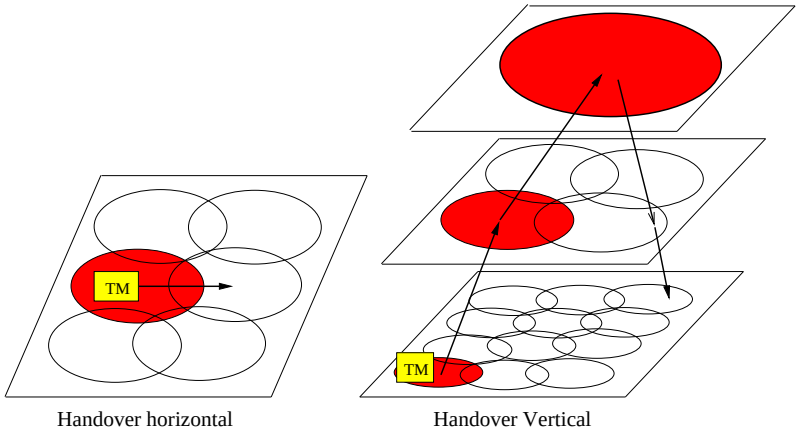


FIG. 3.2 – *Handover horizontal vs. Handover vertical.*

Il existe quelques différences importantes entre les handovers horizontaux et verticaux qui affectent les stratégies de mise en œuvre :

– Un terminal mobile procède à un handover horizontal d'une cellule A à une cellule B lorsqu'il se déplace hors de la zone de couverture de la cellule A vers celle de la cellule B. Ceci n'est pas forcément le cas pour les handovers verticaux. Le terminal mobile peut se déplacer hors de la zone de couverture d'une cellule A de niveau hiérarchique i à une zone de couverture d'une cellule B de plus grande taille de niveau $i+1$, comme il peut effectuer un handover vertical de la cellule B vers la cellule A tout en restant toujours sous la couverture de la cellule B (la cellule B enveloppe la cellule A).

– Pour pouvoir exécuter un handover vertical, le terminal mobile doit être doté de deux ou plusieurs interfaces de technologie sans fil. Chaque technologie sans fil possède ses propres caractéristiques : bande de fréquences, méthode d'accès et codages, infrastructure, etc. Par exemple, des fréquences infrarouges sont utilisées dans un bureau, alors qu'un ensemble de fréquences radios dans un bâtiment et un autre ensemble de fréquences dans un campus. Les équipements peuvent aussi utiliser différents types de couches physiques : FHSS (Frequency Hopping Spread Spectrum), DSSS (Direct Sequence Spread Spectrum), IR (Infrarouge), FDD (Frequency Division Duplex), TDD (Time Division Duplex), etc.

– Dans un réseau où les stations de base sont homogènes, le choix de la meilleure station de base est évident : le terminal mobile procède pour cela aux mesures d'un signal pilote, que chaque station de base transmet en permanence. Le signal reçu avec la meilleure qualité (selon les méthodes d'initiation des handovers) indique le meilleur point d'accès d'attachement. Dans une OWN, la puissance du signal n'est pas toujours le facteur déterminant de la meilleure station de base du réseau. Ceci est dû à la diversité des caractéristiques de chaque niveau de la OWN. Une faible puissance d'un signal provenant d'une station de base utilisant un réseau WLAN peut accomplir une meilleure performance qu'un signal puissant envoyé par une station de base qui utilise une technologie d'un réseau sans fil plus large (WWAN).

Le handover vertical est divisé en deux catégories : handover vertical montant (*upward*) et descendant (*downward*). Un handover vertical montant est un handover vertical d'une cellule de niveau i à une cellule de niveau $i+j,\ j>0$. Le passage à une cellule de niveau supérieur peut impliquer une dégradation de service due à la limitation de la bande passante. Un handover vertical descendant permet au terminal mobile d'accéder à une cellule de niveau $i-j,\ j>0$ avec une bande passante supérieure à celle offerte par la cellule de niveau i. Il est logique qu'un terminal mobile *cherche sans cesse à exécuter des handovers descendants* afin d'améliorer la qualité des services de ses applications. Les handovers verticaux descendants présentent un temps moins critique que les handovers montants vu que le terminal mobile reste dans la zone de couverture de l'ancienne cellule en effectuant un handover vertical descendant alors qu'il quitte définitivement l'ancienne cellule avant de précéder à un handover vertical montant.

3.5 Position des objectifs

La liberté de la mobilité entre les différents réseaux sans fil dans une OWN permet aux utilisateurs de se connecter au meilleur réseau à n'importe quel instant. Cette liberté crée de nouveaux problèmes, ouvre d'autres aspects de recherche et pose de nouveaux défis. Parmi les défis on trouve : les techniques de AAA, la sécurité globale, la QoS et réservation des ressources entre les différents systèmes, la mobilité horizontale (intra-technologie) et verticale (inter-technologie).

L'objectif principal de cette thèse est de minimiser la latence des handovers verticaux tout en maintenant une consommation de la bande passante la plus basse possible. Une faible latence offre la possibilité d'un transfert souple entre les réseaux avec des coupures de communications supportables et des pertes de données négligeables. Ceci permet une meilleure interactivité entre l'utilisateur et les services multimédias offerts par les réseaux. Comme un utilisateur peut se déplacer vers une zone de mauvaise connectivité, le seul changement possible et visible pour cet utilisateur doit être lié à la limitation de la bande passante de l'interface sans fil. Par exemple, Les niveaux les plus hauts d'une OWN supportent une interactivité vidéo et audio de très bonne

qualité, alors que les niveaux les plus élevés supportent uniquement des services audio.

La puissance de consommation de la batterie est un facteur important à optimiser surtout en présence d'interfaces multiples qui peuvent fonctionner simultanément. L'approche la plus simple pour gérer les interfaces sans fil est de les laisser fonctionner en même temps. Selon des mesures présentées dans [39] faites sur des interfaces disponibles pour le grand public, les interfaces IBM Infrared WaveLAN RF [40] consomment en tout 1,5 watts lorsqu'elles sont au repos (pas de transmission ni de réception). Ce qui présente approximativement 20% de la puissance consommée par un ordinateur portable. À ce niveau de consommation, la gestion de la puissance de consommation de la batterie devient cruciale. Une solution simple consiste à faire passer périodiquement l'interface de niveau hiérarchique directement inférieur du niveau d'attachement dans la OWN en mode veille (*power saving low duty cycle sleep state*) pendant un temps D (en secondes). Les interfaces supérieures restent toujours en mode veille. Ceci est dû aux mécanismes de handover vertical où le terminal mobile cherche sans cesse à écouter les interfaces des niveaux hiérarchiques inférieurs dans l'OWN. Cette solution est très efficace et des tests expérimentaux sont montrés dans [39]. Nous utilisons cette méthode pour calculer la latence dans notre système.

L'implémentation d'un handover vertical dans une OWN utilise une partie de la bande passante qui peut être importante si la signalisation, qui est en forme des messages de handovers envoyés par le mobile ou des signaux pilotes (trames balises) envoyés par les stations de base, est importante. Les performances d'une OWN augmentent avec la réduction de la signalisation supplémentaire.

Idéalement, un utilisateur essaie de se connecter au niveau le plus bas d'une OWN, où la bande passante est très large, le plus long temps possible avant qu'il sera nécessaire de se déplacer vers un niveau plus haut. Notre système avec des cellules hétérogènes doit avoir le même comportement qu'un système de cellules homogènes en terme de la faible latence des handovers horizontaux. La raison pour exécuter un handover vertical de niveau i vers un niveau $i+1$ réside dans le fait que le niveau i devient indisponible en se déplaçant hors de sa zone de couverture. La découverte de l'indisponibilité de réseau d'attachement prend un certains temps afin d'assurer que le terminal mobile s'est vraiment déplacé hors de la zone de couverture. Généralement en fonction de la non réception d'un certain nombre (seuil) des trames balises envoyées par la station de base d'attachement. Ce temps augmente considérablement la latence du handover vertical et un mécanisme de prédiction serait très utile. Les mécanismes de prédiction développés dans le cadre des systèmes homogènes ne sont pas applicables dans une OWN pour les handovers verticaux. La prédiction des handovers verticaux est l'un de nos objectifs qui a un effet direct sur la minimisation de la latence.

Il est pratiquement impossible d'atteindre ces objectifs à la fois. Cependant, il faut aussi éviter les situations où la réalisation d'un objectif dégrade la réalisation des autres objectifs. Par exemple, la réduction de la puissance de consommation par la désactivation des interfaces sans fil lorsqu'ils ne sont pas utilisées augmente le temps latence. Aussi, l'émission et la réception des données sur toutes les interfaces des réseaux sans fil simultanément permettent l'obtention d'un handover avec une latence nulle mais par contre il faut s'attendre à des

consommations fortes de puissance et un gaspillage de la bande passante. Le cas idéal consiste à faire un compromis entre ces objectifs.

3.6 OWN : La gestion de la mobilité

Le protocole IP Mobile [8] standardisé par l'IETF [4] présente une solution incontournable pour la gestion de la mobilité sur Internet. Cette solution est bien adaptée à la macro-mobilité,

mais elle est beaucoup moins adaptée à la micro-mobilité (c'est-à-dire, mobilité inter-domaine). Nous allons proposer deux architectures (Archi-1 et Archi-2) pour deux solutions différentes mais les deux sont basées sur le protocole IP Mobile pour gérer la macro-mobilité. Dans les deux architectures, les handovers sont contrôlés par les terminaux mobiles.

3.6.1 Description d'Archi-1

La figure 3.3 illustre notre première architecture (Archi-1) proposée pour la mobilité IP dans une OWN. Cette architecture est basée sur différents types de technologie de réseau sans fil regroupés dans un réseau d'accès de la OWN. Ce réseau d'accès constitue un sous-domaine IP et il est attaché à Internet via une passerelle. Le déplacement d'un terminal mobile à l'intérieur d'un réseau d'accès (mobilité intra-domaine) est géré par le système sans fil correspondant (WLAN, UMTS, etc) avec un handover vertical entre les systèmes hétérogènes. Le changement de réseau d'accès (mobilité inter-domaine) est géré par le protocole IP Mobile.

L'architecture est hiérarchique dans le sens où il existe un niveau intermédiaire entre la passerelle et les différentes stations de base. Ce niveau intermédiaire est constitué d'une partie fixe sous forme d'arborescence filaire reliant la passerelle et les stations de base ou des points d'accès sans fil. La OWN est une structure virtuelle qui permet de définir le handover vertical.

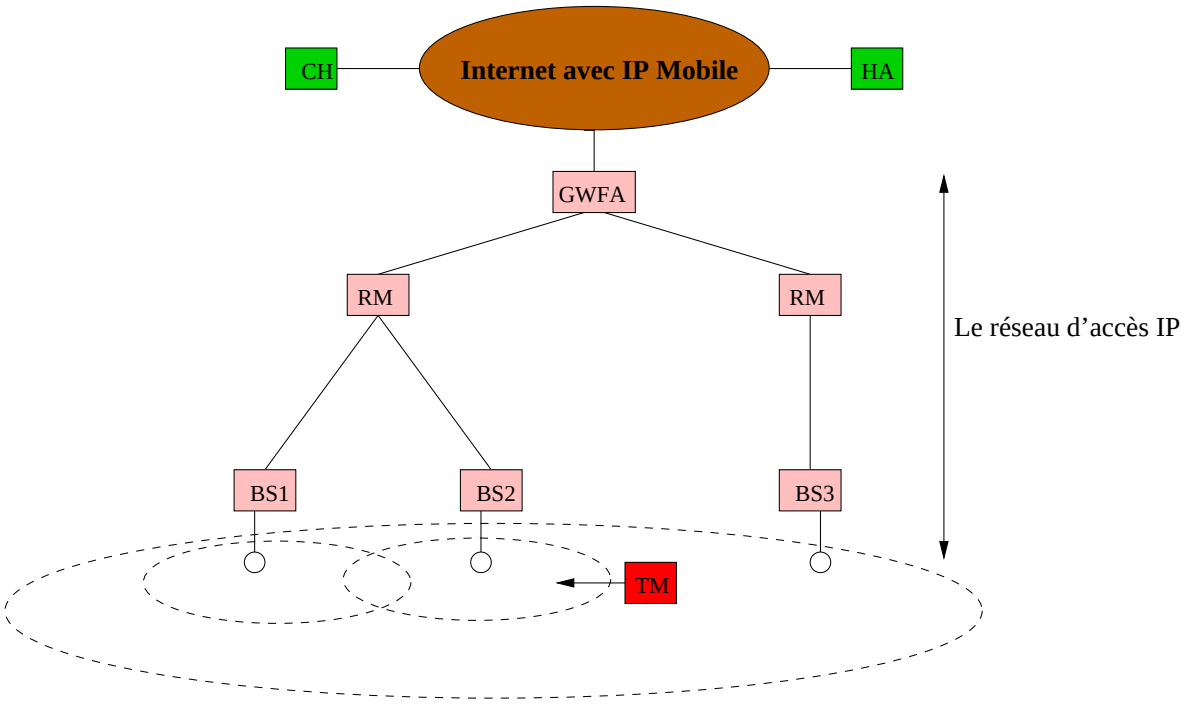


FIG. 3.3 – *Archi-1 : la 1ère architecture pour la gestion de la mobillité.*

L'architecture est composée de six entités fonctionnelles :

– Home Agent (HA) : routeur situé dans le réseau d'abonnement du terminal mobile. Il enregistre la localisation du terminal mobile et renvoie les paquets destinés à ce dernier vers son réseau visité.

– Correpondent Host (CH) : un nœud dans le réseau Internet qui communique avec le terminal mobile dans son réseau Home.

– Passerelle Foreign Agent (GWFA) : routeur tenant le rôle de Foreign Agent (FA) pour les nœuds visiteurs. Il tient également le rôle de HA dans le cas où le réseau d'accès est le réseau d'abonnement de certains nœuds mobiles. En outre, il implémente la fonctionnalité multicast pour router les paquets vers le terminal mobile à l'intérieur du réseau d'accès. Il diffuse périodiquement des messages d'annonce indiquant son adresse IP et il associe à chaque terminal mobile une adresse multicast unique dans son réseau d'accès.

– Routeur Mulicast (RM) : routeur intermédiaire entre la GWFA et les stations de base dans le réseau d'accès. Il implémente la fonctionnalité multicast.

– Station de base (BS) : nœud reliant les parties filaires et sans fil de l'architecture. Il possède deux interfaces, une sans fil pour communiquer avec les terminaux mobiles et l'autre filaire pour communiquer avec la passerelle via l'arborescence filaire.

– Terminal Mobile (TM) : un nœud dans le réseau d'accès possédant plusieurs interfaces sans fil. Il implémente la procédure d'enregistrement IP Mobile. Cette procédure est déclenchée lors d'un changement de réseau d'accès.

3.6.2 Fonctionnement d'Archi-1

Lorsqu'un TM est dans son réseau d'abonnement, un CH peut communiquer avec lui sur Internet via le routage IP normal. Une fois ce TM se déplace vers un réseau visiteur, le protocole IP Mobile intervient pour permettre au TM d'obtenir une COA par la GWFA du réseau d'accès auquel il est attaché et de s'enregistrer auprès de son HA. En plus, le TM doit acquérir une adresse IP muticast unique dans le réseau d'accès proposée par la GWFA. Un petit groupe de BSs est sélectionné par le TM pour écouter cette adresse multicast. Ce groupe forme le groupe multicast associe au TM. La raison de la sélection d'un groupe multicast de stations de base réside dans le fait que ces stations de base sont voisines et lorsque le TM se déplace, il va s'y trouver couvert par une des stations de son groupe. Une BS dans ce groupe est sélectionnée par le TM comme une *forwarding* BS, elle relaie immédiatement les paquets multicast envoyés par la GWFA. Les autres BSs du groupe multicast jouent le rôle de *buffering* BSs, elle stockent les paquets multicast envoyés par la GWFA dans un *buffer* circulaire afin de maintenir les paquets les plus récents et d'éviter le surchargement des *buffers*. La bufferisation des paquets par les BSs voisines de la *forwarding* BS permet au TM de récupérer les paquets non reçus lors de la procédure de handover.

La GWFA décapsule les paquets envoyés par le HA vers le TM. Elle utilise l'adresse multicast associée au TM correspondant pour envoyer les paquets en mode multicast.

Les BSs envoient périodiquement des trames balises pour annoncer leurs présences et relayer d'autres informations. Le TM écoute en permanence ces trames et détermine la *forwarding* BS, les *buffering* BSs et les BSs qui ne font pas partie de son groupe multicast. Lorsque le TM initialise un handover, il demande à l'ancienne BS de passer de l'état *forwarding* BS à l'état *buffering* BS. Les paquets stockés par la nouvelle *forwarding* BS et qui ne sont pas encore reçus pas le TM sont renvoyés à ce dernier.

3.6.3 Les handovers dans Archi-1

Dans les cellules homogènes (les cellules de même niveau hiérarchique), les puissances des signaux des trames balises reçues par le TM sont comparées et la BS avec le signal le plus puissant (selon les mécanismes d'initiation des handovers montrés dans la section 3.4) est sélectionnée comme la *forwarding* BS. La figure 3.4 montre en détail le mécanisme du handover horizontal. En plus des envois périodiques de trames balises, la *forwarding* BS relaie les paquets de données vers le TM. Si le TM reçoit une trame balise d'une BS de meilleure qualité que celle de *forwarding* BS, il décide d'initier un handover horizontal en demandant à la nouvelle BS d'arrêter la bufferisation et de se comporter comme une *forwarding* BS. À l'ancienne BS, il demande de passer de l'état de *forwarding* BS à l'état *buffering* BS. La latence d'un handover horizontal commence dès l'instant où le TM décide que le signal reçu de la nouvelle BS est meilleur que le signal de l'ancienne BS jusqu'à la réception du premier paquet de données provenant de la nouvelle BS.

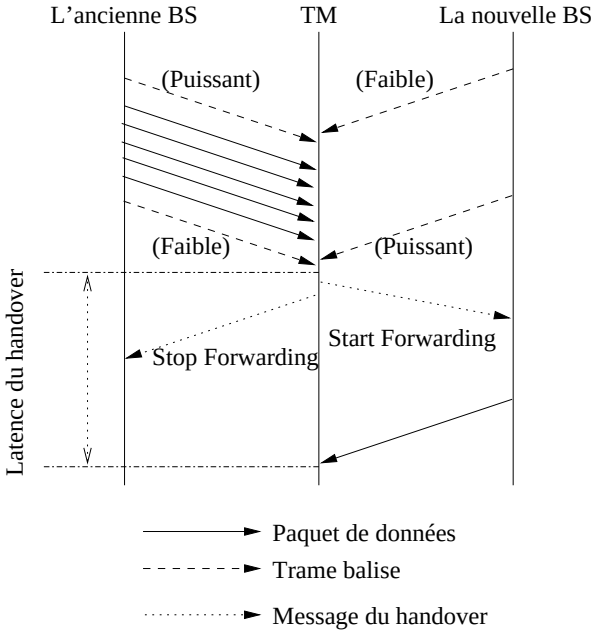


FIG. 3.4 – *Le mécanisme du handover horizontal dans l'Archi-1.*

Entre les cellules hétérogènes (les cellules de différents niveaux hiérarchiques), le TM écoute les différentes trames balises et choisit le meilleur signal non pas seulement en fonction de sa puissance mais aussi en fonction d'autres métriques (la qualité du signal, le débit offert, etc). La BS avec le meilleur signal est sélectionnée comme une *forwarding* BS et un petit ensemble des BSs est considéré comme *buffering* BSs. Un handover vertical montant est initialisé lorsqu'un certains nombre de trames balises (selon un seuil T_B) n'a pas été reçu du niveau hiérarchique d'attachement de la OWN (le niveau auquel appartient la *forwarding* BS). Le TM décide que le niveau hiérarchique d'attachement est inaccessible et il effectue un handover vertical montant vers un niveau accessible directement supérieur dans la OWN. Dans ce cas, le TM demande à la *buffering* BS de meilleur signal d'un niveau hiérarchique supérieur de passer à l'état *forwarding* BS. Comme l'ancienne *forwarding* BS n'est plus accessible directement, le TM envoie une requête à cette BS pour qu'elle cesse la transmission via la nouvelle *forwarding* BS. Ce procédé est montré dans la figure 3.5. Les croix sur cette figure signifient les points logiques de l'atténuation des

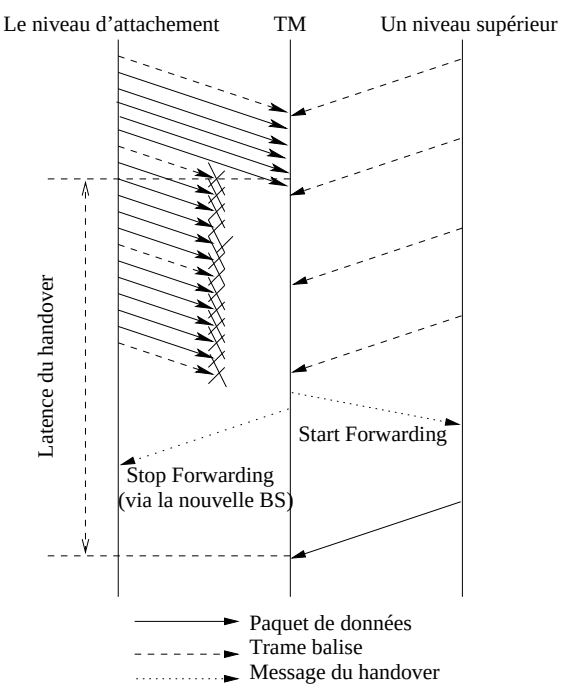


FIG. 3.5 – *Le mécanisme du handover vertical montant dans l'Archi-1.*

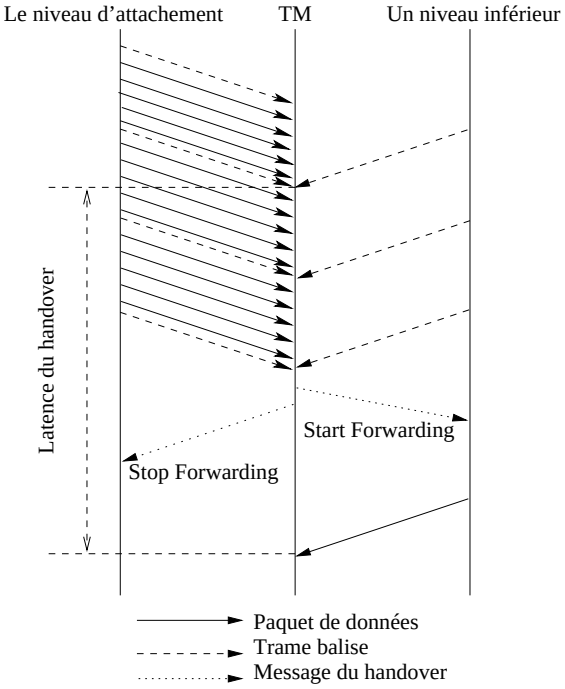


FIG. 3.6 – *Le mécanisme du handover vertical descendant dans Archi-1.*

messages et non pas les chemins empruntés par ces messages. La latence d'un handover vertical montant est donnée par la durée entre le moment où le niveau hiérarchique

d'attachement de la OWN commence à devenir inaccessible jusqu'à la réception du premier paquet de données envoyé par la nouvelle BS.

Lorsqu'un certain nombre de trames balises (selon un seuil T_B) est reçu successivement d'un niveau hiérarchique inférieur dans la OWN, le TM effectue un handover vertical descendant (figure 3.6). La latence d'un handover vertical descendant est donnée par la durée entre le moment où un niveau hiérarchique inférieur dans la OWN commence à devenir accessible jusqu'à la réception du premier paquet de données envoyé par la nouvelle BS.

3.6.4 Latence et overhead dans Archi-1

La latence d'un handover vertical L est définie par le temps nécessaire pour qu'un TM puisse se déconnecter de l'ancienne BS et recevoir le premier paquet de la nouvelle BS. L est donnée pour la formule suivante :

$$L = L_D + L_P + L_N + L_F$$

– L_D est le temps nécessaire pour qu'un TM découvre qu'il doit se connecter à un nouveau niveau hiérarchique de la OWN. En absence de T_B trames balises du niveau d'attachement, le TM décide que ce niveau est inaccessible et il doit initier un handover vertical montant vers un niveau plus haut accessible dans la OWN. La réception de T_B trames balises d'un niveau inférieur dans la OWN donne au TM la possibilité d'effectuer un handover vertical descendant. L_D dépend fortement de la fréquence d'émissions des trames balises par les

BSs. Si l'espace de temps entre deux trames balises est large, la valeur de L_D sera grande par le fait que le TM prendra un temps important pour découvrir l'accessibilité d'un niveau inférieur ou l'inaccessibilité du niveau courant d'attachement de la OWN. Soit N_B l'espace de temps entre deux trames balises (en secondes). La borne supérieure de L_D est donnée par $N_B \times T_B$. La borne inférieure est égale à $N_B \times (T_B - 1)$. Par conséquent, la moyenne est donnée par $N_B \times (T_B - 1) + N_B/2$. Cette composante de la latence est invisible comme une déconnexion pour les utilisateurs dans le cas d'un handover horizontal ou vertical descendant puisque le TM reste toujours connecter à l'ancienne BS pendant qu'il découvre qu'il est possible de se connecter à une nouvelle BS. Si on considère D (en secondes) le temps périodique de mode en veille de l'interface de niveau hiérarchique directement inférieur au niveau d'attachement, alors il faut rajouter à la latence moyenne L_D des handovers verticaux descendants un temps moyen de $D/2$.

– L_P est la latence nécessaire pour qu'un TM effectue une transition vers l'interface qui correspond à la technologie utilisée par la nouvelle BS. Cette composante peut prendre la valeur 0 si l'état de l'interface cible est déjà en marche lors de l'apparition du handover.

– Le TM demande à la nouvelle BS de se comporter comme une *forwarding* BS en un temps L_N. Ceci dépend généralement de la latence entre le TM et la BS. L_N prend un temps $L_U + S_M/Bw_U$ secondes pour les handovers verticaux montants et $L_L + S_M/Bw_L$ pour les handovers verticaux descendants. (L_U, Bw_U) et (L_L, Bw_L) représentent respectivement (la latence (en secondes), La bande passante (en bits/s)) de la technologie d'interface cible du niveau supérieur et inférieur dans la OWN. S_M (en bits) représente La taille de message de changement d'état envoyé par le TM à la nouvelle BS.

– L_F est la temps nécessaire pour que la nouvelle BS puisse envoyer le premier paquet de données vers le TM. $L_F = 0$ s'il n'y a aucun paquet à envoyer. L_F prend un temps $L_U + S_{Data}/Bw_U$ secondes pour les handovers verticaux montants et $L_L + S_{Data}/Bw_L$ pour les handovers verticaux descendants. S_{Data} est la taille de ce premier paquet destiné au TM.

L_D et L_P se chevauchent dans le cas où le TM peut prédire dans le temps l'inaccessibilité du niveau courant d'attachement dans la OWN et il met en marche à l'avance l'interface d'un niveau supérieur accessible. L_D, L_N et L_F peuvent aussi se chevaucher si la nouvelle BS est déjà dans l'état *forwarding* BS vers le TM. Ce cas est possible dans l'optimisation montrée dans la section 3.7.3 où le TM ordonne plusieurs BSs de se comporter comme *forwarding* BSs. Par contre L_P et L_N ne peuvent jamais se chevaucher.

L'overhead B est le nombre de bits de la signalisation envoyée par seconde. Les trames balises et d'autres messages que le TM utilise pour initier un handover sont considérés comme un overhead. La valeur moyenne de l'overhead pour un handover vertical montant ou descendant est $(1/N_B)S_B$, où S_B est la taille (en bit) d'une trame balise.

3.6.5 Description d'Archi-2

Dans Archi-1, le TM ne s'attache pas forcément à toutes ses *buffering* BSs ce qui engendre des bufferisations inutiles. Afin d'éviter ce problème, l'architecture 2 (Archi-2) remplace les routeurs multicast (RM) dans le réseau d'accès d'Archi-1 par des routeurs IP normaux. Le but de cette architecture est d'éviter l'utilisation de la technique de groupe multicast qui bufferise les paquets de données de chaque TM. La figure 3.7 illustre notre deuxième architecture (Archi-2).

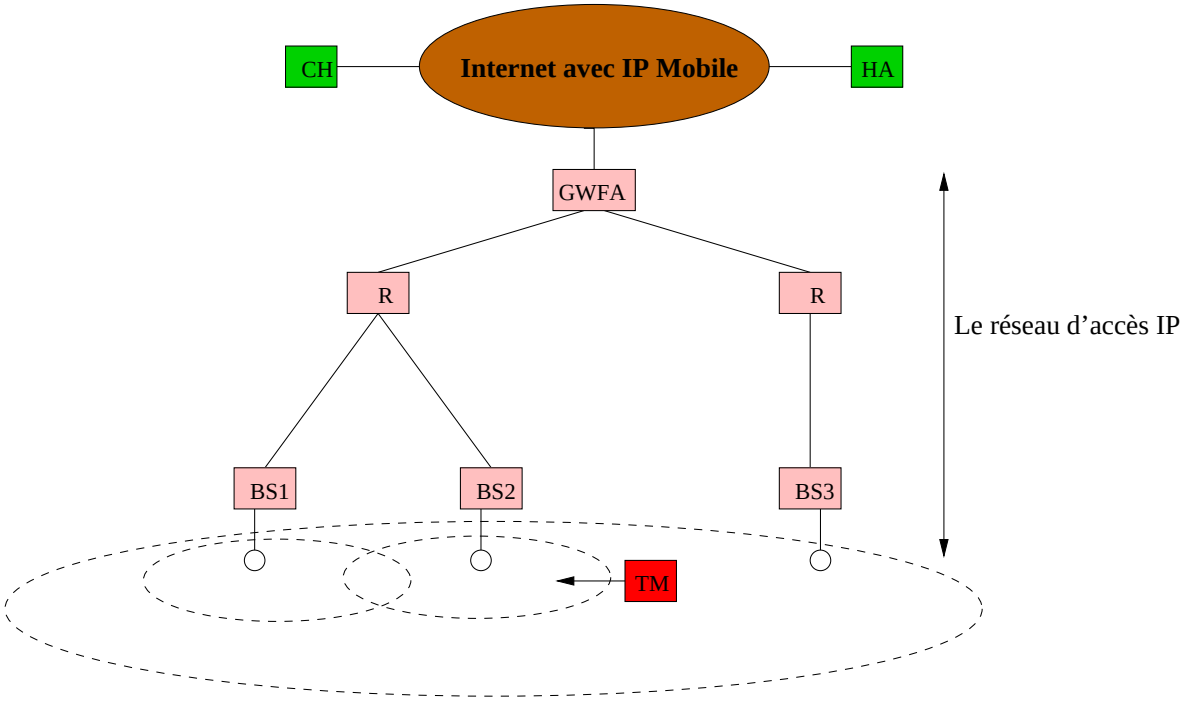


FIG. 3.7 – *Archi-2 : la 2ème architecture pour la gestion de la mobillité.*

3.6.6 Fonctionnement d'Archi-2

Dans cette architecture et contrairement à Archi-1, le TM visiteur doit acquérir seulement une COA par la GWFA du réseau d'accès auquel il est attaché et s'enregistrer auprès de son HA. Le TM recherche constamment la BS qui offre le meilleur signal. Une fois attaché à une cellule, la BS de cette cellule relaie les paquets de données vers lui. Il n'a y a pas de bufferisation par les cellules voisines dans cette architecture. Pour assurer une perte de paquets nulle lors des handovers, la BS où le TM est attaché stocke les derniers paquets de données envoyés. Ces paquets font l'objet d'une retransmission vers la nouvelle BS sélectionnée par le TM. Cette retransmission est effectuée par l'ancienne BS via le réseau filaire vers la nouvelle BS. Le fonctionnement d'IP Mobile est préservé dans Archi-2.

3.6.7 Les handovers dans Archi-2

À la réception d'une trame balise de meilleure qualité d'une cellule voisine homogène (même technologie sans fil), le TM initie un handover horizontal en suivant les étapes suivantes (figure 3.8) :

- Le TM envoie un message *Greet* (un message pour initialiser le handover) à la nouvelle BS. Ce message contient l'adresse de TM, l'adresse de l'ancienne BS et une petite liste des *IDentificateurs* (IDs) des derniers paquets de données reçues par le TM.
- La nouvelle BS crée une nouvelle entrée dans sa table de routage pour le TM. Cette entrée représente la capacité de la nouvelle BS de relayer les données vers ce TM. La nouvelle BS acquitte le message *Greet* en envoyant un *Greet Ack* au TM. À la réception de l'acquittement, le TM est attaché définitivement à la nouvelle BS. Pour éviter la perte des données

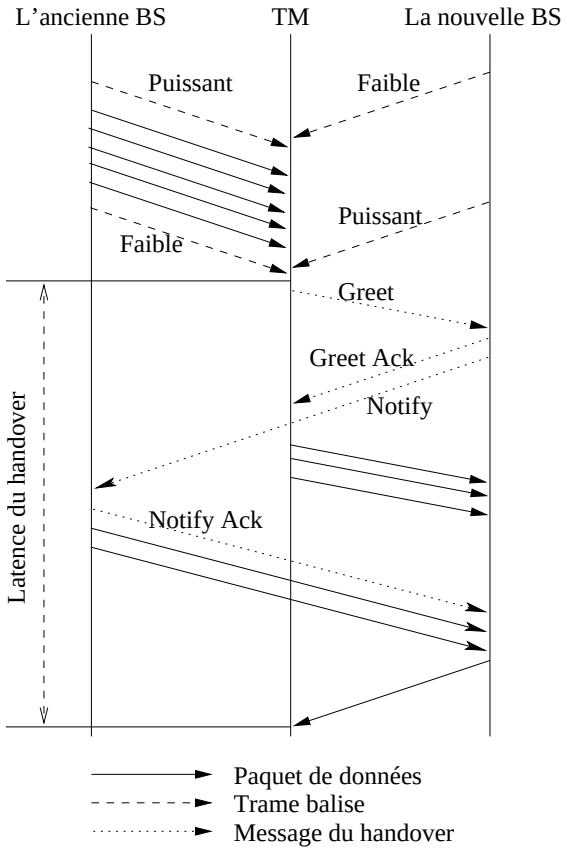


FIG. 3.8 – *Le mécanisme du handover horizontal dans Archi-2.*

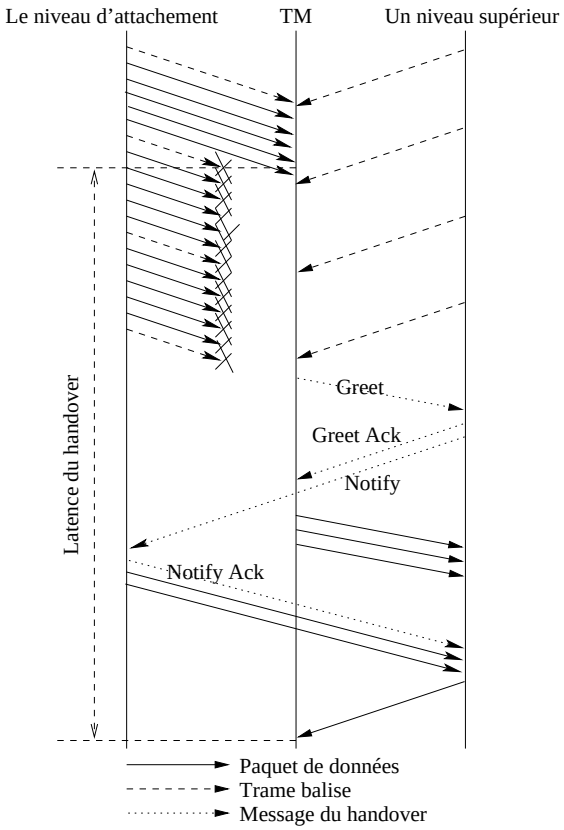


FIG. 3.9 – *Le mécanisme du handover vertical montant dans Archi-2.*

envoyées récemment à l'ancienne BS, et qui sont dans le buffer de TM, ces données sont renvoyées à la nouvelle BS.

– L'ancienne BS est informée que le TM a changé la cellule par un message de notification (*Notify*) envoyé via le réseau filaire par la nouvelle BS. Ce message contient aussi la liste des identificateurs des paquets reçus récemment par le TM. Cette liste est la même liste transmise précédemment par le TM dans le message *Greet* à la nouvelle BS. Cette technique permet à l'ancienne BS de retransmettre les paquets de données qui ne sont pas reçus par le TM lorsqu'il était attaché à sa cellule en passant par la nouvelle BS.

– À la réception de la notification, l'ancienne BS procède à une mise à jours de sa table de routage en retirant l'entrée correspondante à ce TM. Elle envoie les paquets de données disponibles dans son buffer dont les identificateurs ne figurent pas dans le message de notification vers la nouvelle BS. Elle répond aussi par un *Notify Ack*.

Ce processus du handover épargne les BSs des cellules voisines d'effectuer des bufferisations qui peuvent être inutiles. L'échange des messages de notification entre la nouvelle BS et l'ancienne BS permet d'assurer la retransmission des données qui peuvent être perdues durant le handover.

Le principe des handovers verticaux montant est illustré dans la figure 3.9. Les messages *Greet, Greet Ack, Notify* et *Notify Ack* sont décrit précédemment.

3.6.8 Latence et Overhead dans Archi-2

le temps latenc L des handovers dans cette Archi-2 est

$$L = L_D + L_P + L_{Greet} + L_{Notify} + L_T + L_N$$

- L_{Greet} est le délai entre la transmission et la réception d'un message *Greet* envoyé par le TM à sa nouvelle BS. L_{Greet} prend un temps $L_U + S_{Greet}/Bw_U$ secondes pour les handovers verticaux montants et $L_L + S_{Greet}/Bw_L$ pour les handovers verticaux descendants. S_{Greet} est la taille (en bits) de message *Greet*.
- L_{Notify} est la latence d'envoyer un message *Notify* par la nouvelle BS à l'ancienne BS. Cette latence dépend de la taille de message *Notify* (S_{Notify}), du nombre de liens ou sauts (H) séparant l'ancienne et la nouvelle BS dans le réseau filaire, de la latence (L_W) et la bande passante (Bw_W) de chaque routeur dans le réseau filaire qui connecte les deux BSs. La valeur de L_{Notify} est donnée par $[(S_{Notify}/Bw_W) + L_W] \times H$.
- L_T représente le temps entre la transmission du premier paquet de données dont l'identificateur ne fait pas partie de la liste des *IDs* par l'ancienne BS à la nouvelle BS passant par le réseau filaire. L_T prend $[(S_{Notify}/Bw_W) + L_W] \times H$.

La latence L de cet algorithme de handover est quasiment équivalente à celle présentée dans l'Archi-1 en considérant que la retransmission des données entre l'ancienne BS et la nouvelle BS sur le réseau filaire est négligeable vu que les deux BSs ne sont pas très éloignés l'une de l'autre et le réseau filaire présente un débit et fiabilité plus importants. L'overhead de la bande passante dépend seulement de la fréquence d'émission des trames balises. La valeur moyenne de l'overhead pour un handover vertical montant ou descendant est $(1/N_B)S_B$, où S_B est la taille (en bit) d'une trame balise.

3.7 Optimisation des handovers

Les applications temps réels tels que la voix ou la vidéo nécessitent un temps de latence très faible lors des handovers afin d'assurer la transparence auprès des utilisateurs. Dans la littérature, plusieurs méthodes d'optimisation des handovers classiques ont été développées. Ces méthodes peuvent être toujours utilisées lorsqu'un terminal mobile change sa cellule d'attachement en restant dans le même niveau hiérarchique dans la OWN. Les handovers verticaux descendants offrent aux TMs l'avantage de rester connectés à l'ancienne BS durant le processus du transfert vers la nouvelle BS. Par conséquent, l'optimisation de la latence des handovers verticaux descendants ne présente pas une priorité. Les handovers verticaux montants consomment beaucoup de temps avant qu'un TM réalise qu'il est hors de sa cellule d'attachement et qu'il est nécessaire de remonter dans les niveaux hiérarchiques de la OWN. Les composantes L_D et L_P de la latence L sont les seuls paramètres qui offrent la possibilité d'une optimisation. L'intervalle de temps avant d'initier un handover vertical montant et qui consiste en l'absence de T_B trames balises de l'ancienne BS peut être contourner par deux méthodes : la prédiction ou l'augmentation de la fréquence d'émission des trames balises par les BSs. La latence L_P dépend de la manière de faire fonctionner l'ensemble des interfaces d'un TM.

3.7.1 La cellule virtuelle

On a vu précédemment que les BSs envoient périodiquement des trames balises. Chaque TM conserve la trace de la puissance du signal reçu appelée RSS (Received Signal Strength) de sa BS d'attachement. Lorsque la RSS de cette BS commence à s'affaiblir au-delà d'un certain seuil, le TM commence à mesurer les RSSs des cellules voisines. L'initiation d'un handover horizontal est due à la réception d'une trame balise avec une RSS supérieure à la RSS de la BS d'attachement tandis qu'un handover vertical est initié dès l'absence ou la réception de T_B trames balises. On a vu aussi que la latence d'un handover vertical montant augmente avec l'augmentation de l'espace de temps séparant la transmission des trames balises. Ceci est dû au temps d'attente des T_B trames balises. Le concept des cellules virtuelles permet d'éliminer ce temps d'attente.

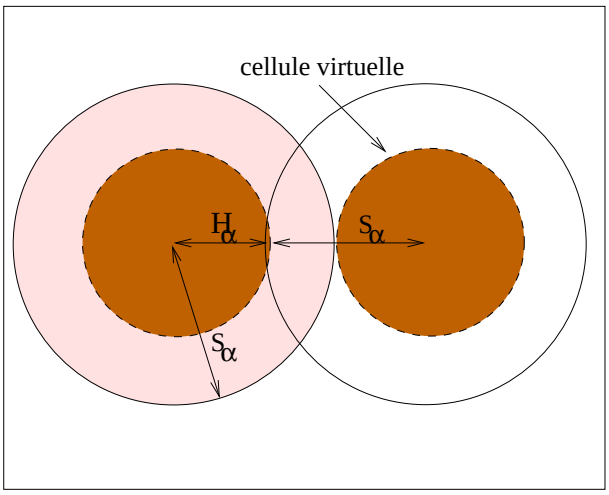


FIG. 3.10 – *La technique de la cellule virtuelle.*

On définit deux paramètres : S_α et H_α. Le seuil d'une puissance de signal acceptable (S_α) est la borne inférieure de la RSS d'une BS sous laquelle une communication devient impossible. Ce seuil forme les frontières d'une cellule. Le seuil H_α est la borne inférieure de la RSS de la BS d'attachement dans laquelle la RSS d'une BS voisine est supérieure ou égale à S_α. Un TM dans une cellule virtuelle reçoit des trames balises de la BS d'attachement avec une RSS supérieure à H_α. Chaque cellule possède sa propre cellule virtuelle. La figure 3.10 montre les deux paramètres qui permettent d'approuver la présence des cellules virtuelles. Un TM en dehors de la cellule virtuelle de la BS d'attachement doit recevoir des trames balises des cellules homogènes. Si ce n'est pas le cas, le TM procède à un handover vertical montant lorsque la RSS de la BS d'attachement devient inférieure à S_α sans compter l'absence de T_B trames balises.

La technique de la cellule virtuelle permet d'annuler la composante L_D de la latence pour les handovers verticaux montants. Les paramètres S_α et H_α sont faciles à fixer pour chaque type de cellule de la technologie sans fil utilisée. Ces paramètres sont calculés en fonction de la puissance de transmission de la BS d'attachement, des BSs voisines, de l'environnement et de la bande passante offerte par la technologie.

3.7.2 Émission rapide des trames balises

Le TM peut demander à la BS d'attachement (dans Archi-2) ou à un sous-ensemble de son groupe multicast (dans Archi-1) d'augmenter la fréquence d'émissions des trames balises et donc la réception des trames balises sera plus rapide. Cette technique permet de réduire la latence L_D. La figure 3.11 montre le gain en termes de latence L_D. Les autres composantes de temps latence L sont identiques à celles de la méthode classique.

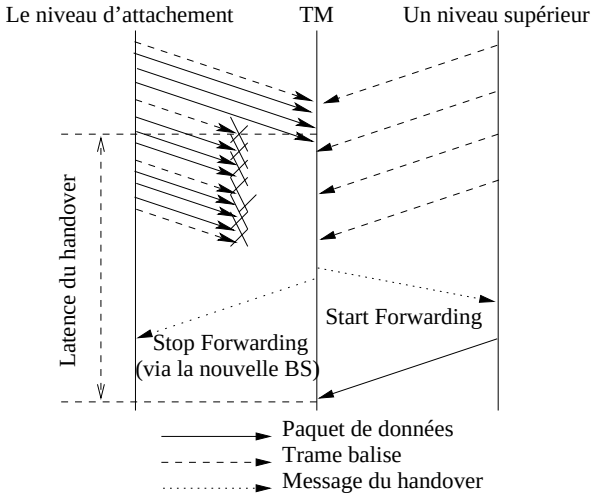


FIG. 3.11 – *Emission rapide des trames balises.*

Cette technique a l'inconvénient d'ajouter un coût supplémentaire en terme de consommation de puissance et d'overhead.

3.7.3 Diffusion multiple des paquets (DMP)

Cette technique concerne la procédure des handovers verticaux dans Archi-1. Le TM peut demander à plusieurs BSs de son groupe multicat de relayer ses paquets de données au lieu d'une seule *forwarding* BS. Ceci implique la réception de plusieurs copies de chaque paquet de données (figure 3.12). Afin de minimiser la réception multiple des paquets tout en optimisant la latence des handovers verticaux montants, le TM ordonne à deux BSs de se comporter comme *forwarding* BSs. La première *forwarding* BS est sélectionnée selon la procédure classique et la seconde est une BS d'un niveau hiérarchique supérieur de la première *forwarding* BS et qui possède le meilleur signal. La duplication des paquets de données reçues est filtrée au niveau couche IP des TMs en conservant dans un cache l'ID de chaque paquet récemment reçu. La couche IP des TMs conserve aussi la trace des paquets reçus par chaque interface. Lorsque T_D consécutifs paquets de données sont reçus par la nouvelle interface et non pas par l'ancienne interface, le TM décide que l'ancien niveau de la OWN est inaccessible et il initie un handover vertical montant vers la nouvelle interface.

Le TM doit recevoir T_D paquets de données (ou T_B trames balises si aucun paquet de données n'est envoyé) par une de ses interfaces supérieures avant l'initiation d'un handover vertical montant. Dans ce cas, L_D prend un temps $N_D(T_D-1)+N_D/2$ secondes où N_D est l'espace de temps entre la transmission de deux paquets de données. Il est clair que les autres composantes, L_N et L_F, n'existent pas vu que le nouvelle BS est déjà une *forwarding* BS. Cependant, cette approche

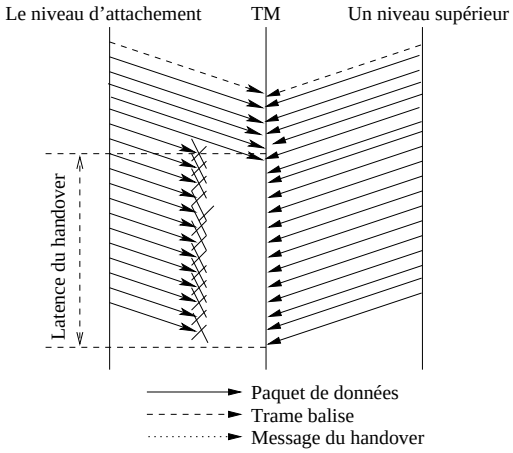


FIG. 3.12 – *Diffusion multiple des paquets.*

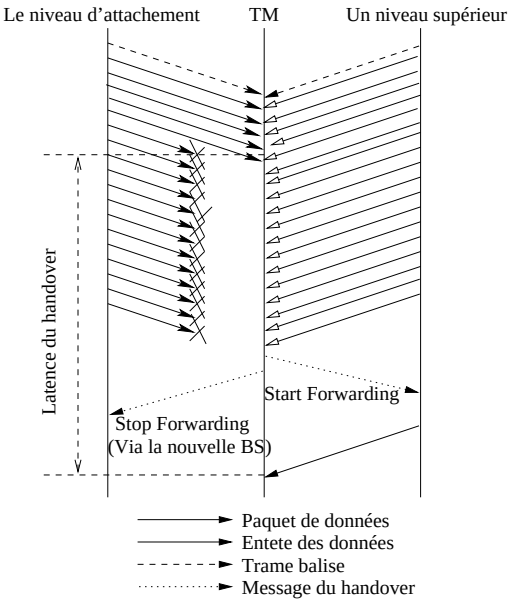


FIG. 3.13 – *Diffusion multiple des entêtes.*

présente une consommation supplémentaire de la puissance due au fonctionnement simultanée de deux interfaces. L'overhead compte, en plus des trames balises, les copies de paquets de données envoyées par la BS de niveau hiérarchique supérieur dans la OWN. Au moyenne on peut avoir $(1/N_B)S_B + (1/N_D)S_D$ b/s. Cette technique ne peut fonctionner que si les deux niveaux de la OWN sont capable de supporter le même chargement de bande passante.

3.7.4 Diffusion multiple des entêtes (DFE)

Dans la méthode précédente, une BS d'un niveau hiérarchique supérieur dans la OWN se comporte comme une *forwarding* BS en même temps que la BS d'attachement. Avec la technique de diffusion multiple des entêtes, la même BS de niveau hiérarchique supérieure dans la OWN bufférise les paquets de données comme dans la méthode classique mais par contre elle envoie uniquement les entêtes IP des paquets (figure 3.13). Le cache de la couche réseau de TM conserve la trace des paquets ainsi les entêtes IP reçus de chaque interface. Ce TM bascule sur la nouvelle interface lorsque T_D entêtes IP sont reçues par la nouvelle interface alors qu'aucun paquet n'est reçu par l'ancienne interface. Cette approche permet de réduire l'overhead généré par la diffusion multiple des paquets.

Etant donnée que le nouvelle BS était en état de *buffering* BS, alors contrairement à l'approche de diffusion multiple des paquets, les composantes de latence L_N et L_F ont les mêmes valeurs que la méthode de base. C'est-à-dire, $L_N = (L_U + S_M/B_U,\ L_L + S_M/B_L)$ et $L_F = (L_U + S_D/B_U,\ L_L + S_D/B_L)$ pour le handover vertical (montant, descendant) respectivement. La puissance de consommation de la batterie d'un TM est la somme des consommations des deux interfaces qui fonctionnent en même temps tandis que l'overhead dépend fortement de la taille et du nombre des trames balises sans oublier le nombre des entêtes IP envoyés. La valeur moyenne de l'overhead est donnée par $((1/N_B)S_B + (1/N_D)S_H$ b/s où S_H est la taille des entêtes IP en bits).

3.8 Les handovers inter-domaines

Afin d'éviter le coût d'exécution d'IP Mobile à chaque changement d'un réseau visiteur (le domaine da la GWFA), une chaîne de retransmission entre les GWFAs est utlisée. Cette chaîne est maintenue par une signalisation supplémentaire entre la nouvelle GWFA et l'ancienne. Lorsqu'un TM se déplace d'un réseau visiteur à un autre, la nouvelle GWFA informe l'ancienne GWFA de la localisation de ce TM dont le but de relayer les paquets de données destinées au TM vers la nouvelle GWFA (figure 3.14). Au fur et à mesure que le TM traverse différents domaines GWFA, une chaîne de retransmission est formée entre les GWFA. Lorsque cette chaîne atteint une certaine limite, IP Mobile est appliqué à nouveau entre le réseau visiteur actuel et le HA. Par conséquent, la chaîne est brisée et la nouvelle GWFA devient le début d'une possible nouvelle chaîne.

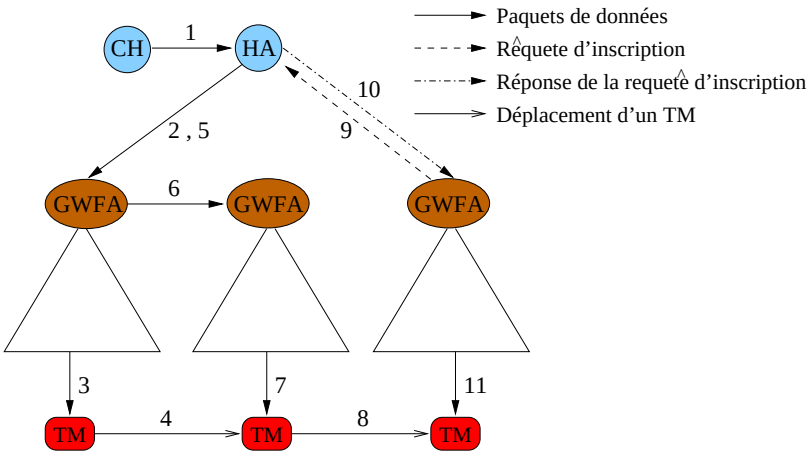


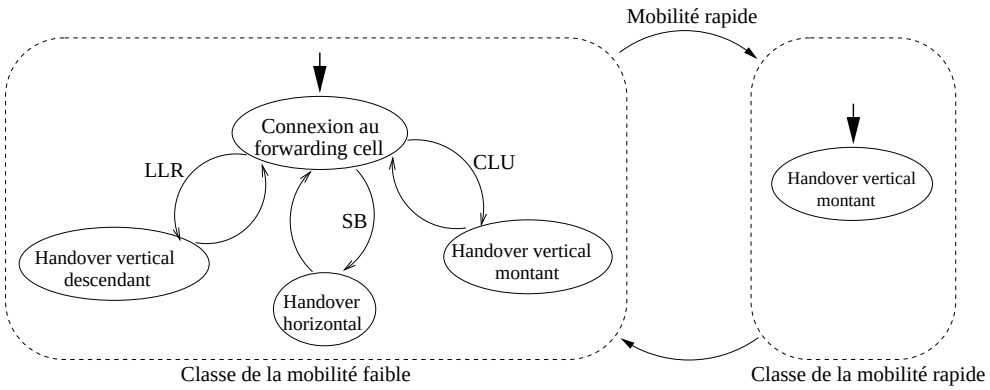
FIG. 3.14 – *Un exemple d'une chaîne de retransmission*.

3.9 La mobilité rapide et faible

Les cellules de chaque niveau hiérarchique de la OWN supportent une mobilité raisonnable imposée par l'environnement (bureau, étage, campus, région, etc). Les Picocells et Nanocells sont en charge de la couverture intérieure (*indoor*) à faible mobilité, tandis que les Macrocells et Satellite assurent une couverture des environnements extérieure (*outdoor*) desservant des mobiles qui se déplacent à grande vitesse. Les Microcells peuvent supporter une mobilité rapide dans un environnement extérieur ou une mobilité faible dans un environnement intérieur.

Nous proposons deux classes de mobilité. La mobilité rapide, représentée généralement par les TMs qui se déplacent dans un véhicule ; et la mobilité faible des piétons. Les TMs avec une mobilité rapide nécessitent plus de handovers et signalisations dans les niveaux hiérarchiques les plus bas de la OWN. Dans ce cas et afin de minimiser le nombre de handovers, les niveaux hiérarchiques supérieurs de la OWN où les cellules couvrent une zone géographique plus large sont adaptables à la mobilité rapide.

La figure 3.15 montre le procédé de la transition entre les deux classes. Une vitesse seuil (*speed-threshold*) est associée à chaque niveau hiérarchique de la OWN qui limite la vitesse maximale des utilisateurs ou le nombre des handovers dans un intervalle de temps. Il est clair qu'un niveau $i+1$ nécessite une vitesse seuil supérieur à celui d'un niveau i. Initialement, chaque TM



CLU : Current Level is Unreachable (le niveau d'attachement est inaccessible)
LLR : Lower Level is Reachable (le niveau inférieur est accessible)
SB : Strong Beacons of homogeneous BS (Un signal puissant d'une cell homogène)

FIG. 3.15 – *La gestion de la mobilité rapide*.

se considère dans la classe où la mobilité est faible. Dans cet état, le TM fonctionne normalement en exécutant des handovers selon les algorithmes présentés dans la section 3.6. Lorsque ce TM détecte qu'il se déplace à une vitesse plus grande que le seuil autorisé par le niveau hiérarchique d'attachement, il change son état d'une mobilité faible à une mobilité rapide et procède à un handover vertical montant. Ensuite il retourne à l'état de la faible mobilité.

3.10 Évaluation des performances

Dans cette section, les performances de notre système de handovers sont étudiées par des simulations sous OPNET [41]. Le simulateur OPNET offre plusieurs outils pour le développement des modèles, l'exécution de la simulation et l'analyse des résultats.

3.10.1 Le modèle de simulation

Une GWFA est placée au sommet de notre architecture de simulation pour envoyer des paquets de donnée en multicast ou unicast vers les TMs selon le type d'architecture utilisé (Archi-1 ou Archi-2). Nous avons considéré Deux niveaux hiérarchiques qui permettent aux TMs d'avoir deux interfaces de technologies sans fil.

Le modèle topologique

La OWN contient deux niveaux hiérarchiques (figure 3.16). Le niveau le plus bas (niveau 0) consiste en une collection de cellules identiques de format circulaire et de rayon R qui offrent une bande passante très élevée. Le niveau supérieur (niveau 1) consiste en un ensemble de cellules de format carrée de coté X qui couvrent une large zone géographique avec une faible bande passante. T% des cellules carrées contiennent des cellules circulaires identiques et uniformément distribuées. Le nombre des cellules circulaires est suffisant pour couvrir toute la surface d'une cellule carrée. Si $T = 50$, la moitié des cellules carrées comportent des petites cellules circulaires.

Les stations de base sont modélisées par des files d'attente et placées au centre des cellules correspondantes (cercle ou carré). Le nombre maximal des TMs (*Nombre_max_mobiles*) est l'un

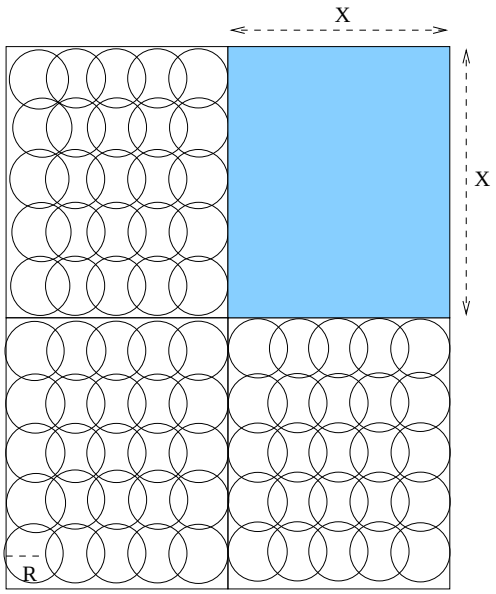


FIG. 3.16 – *Le modèle topologique.*

des paramètres de simulation. Chaque TM est placé aléatoirement dans la zone de déplacement par la sélection aléatoire de ses coordonnées *x* et *y*.

Le modèle de mobilité

Le modèle de mobilité [42] proposé est un processus stochastique continu dans le temps, qui caractérise les mouvements des nœuds dans un espace à deux dimensions. Les mouvements de chaque nœud consistent à une séquence d'intervalles de taille aléatoire, durant lesquels la direction et la vitesse sont constantes. La vitesse et la direction varient aléatoirement d'un intervalle à un autre. Par conséquent, durant un intervalle *i* de durée T_n^i, le nœud *n* parcourt une distance de $V_n^i T_n^i$ sur la même ligne avec un angle θ_n^i. La figure 3.17 illustre le mouvement d'un nœud *n* sur six intervalles de mobilité.

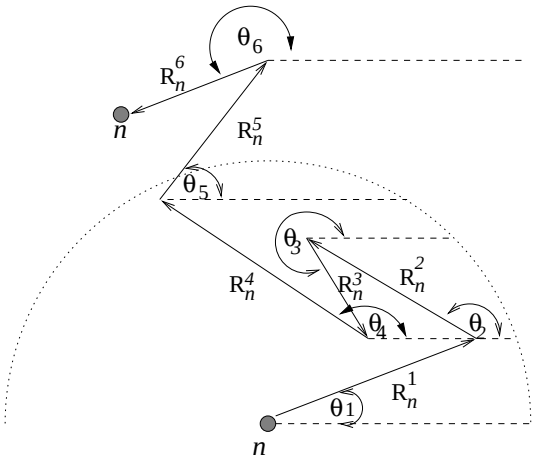


FIG. 3.17 – *Le modèle de mobilité.*

Pour calculer les coordonnées d'un nœud n à l'instant t durant un intervalle i d'une durée T_n^i, d'un angle θ_n^i et d'une vitesse V_n^i, on calcule d'abord la distance D parcourue par n selon la formule $D = V_n^i T_n^i$. Puis, on calcule les coordonnées locales (x,y) : $x = D\sin(\theta_n^i)$ et $y = D\cos(\theta_n^i)$. Enfin, on calcule les coordonnées globales par un changement de repère.

Pour obtenir un équilibre entre les arrivées et les départs dans notre région, tous les nœuds sortants de la zone de périphérie sont réinjectés dans la zone qui lui est symétriquement opposée, éliminant ainsi les effets de bord. La figure 3.18 montre la réinjection des nœuds $M1$, $M2$ et $M3$.

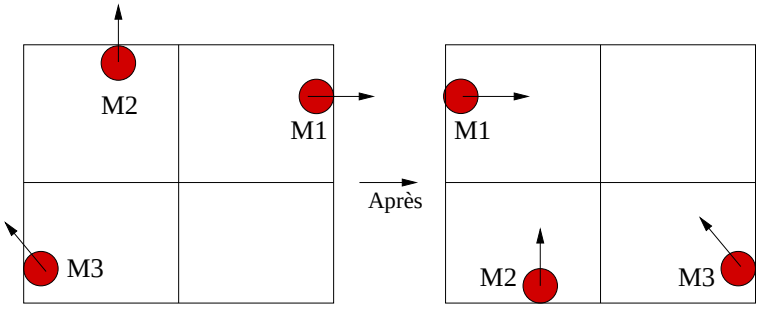


FIG. 3.18 – *La réinjection des nœuds.*

La mobilité d'un nœud n a besoin de trois paramètres : λ_n, $Speed\_Max$, $Speed\_Min$. Le calcul se fait comme suit : la taille des intervalles évolue selon une loi Exponentielle de paramètre $\frac{1}{\lambda_n}$; la direction du mobile durant chaque intervalle suit une loi Uniforme entre $(0, 2\Pi)$; la vitesse du mobile durant chaque intervalle suit une loi Uniforme entre $[Speed\_Min, Speed\_Max]$.

Le modèle de trafic

La GWFA génère des paquets de données UDP selon une loi Poisson de paramètre *inter-arraival*. Il représente l'intervalle de temps moyen séparant deux générations successives des paquets de données. La taille des paquets de données et la sélection d'un nœud récepteur se font selon une loi Uniforme entre $[Taille\_min, Taille\_max]$ et $[0, Nombre\_max\_mobiles - 1]$. Un paquet de donnée contient trois champs : l'adresse de nœud destinataire, la taille de paquet et le numéro de séquence. Le champ numéro de séquence permet de déterminer les paquets de données perdues.

La taille des buffers

Afin d'assurer une perte de données nulle, la taille des buffers des BSs doit être suffisamment grande pour stoker la quantité maximale des données qui peuvent être perdues durant le processus du handover. Ces paquets de données seront renvoyés par la nouvelle BS. La figure 3.19 présente le temps latence L_D nécessaire pour qu'un TM découvre qu'il doit se connecter à une nouvelle BS.

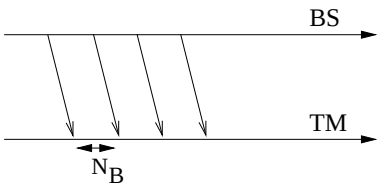


FIG. 3.19 – *Le temps de réception des T_B trames balises.*

La borne supérieure de L_D est donnée par $N_B \times T_B$ où N_B est l'espace de temps entre deux trames balises et T_B présente le nombre des trames balises successives qui provoquent une initiation d'un handover en cas où elles ne sont pas reçues. Par conséquent, le nombre maximal des paquets de données qui peuvent être perdues durant un handover sans un schéma de bufferisation est de

$$(\lceil L_D \rceil / \textit{inter-arraival}) + 1.$$

À partir de cette valeur, la taille maximale d'un buffer qui garantit un zéro-perte de données est extrapolée facilement. Il est clair que l'augmentation de la fréquence d'émission des trames balises réduit la taille des buffers.

3.10.2 Les résultats des simulations

L'objectif de ces simulations est de valider nos protocoles de gestion des handovers qui visent à minimiser la latence et la perte des paquets. Les cellules de niveau inférieur de la OWN (niveau 0) implémentent une technologie sans fil AT&T WaveLAN [40]. Cette technologie offre une bande passante de 1.6 Mb/s, 100 m de diamètre par cellule et 2 ms de temps latence. Les cellules du deuxième niveau (niveau 1) implémentent la technologie sans fil Metricom Ricochet Network [43]. Cette technologie permet à chaque cellule d'avoir une bande passante de 60 kb/s, 1km de diamètre et 100 ms de temps latence. La taille des paquets générés par la GWFA varie ente 1 kb et 65 kb. La meilleure valeur pour T_B est de 3 [44]. 80% (le seuil T) des cellules de niveau hiérarchique inférieur sont couvertes par une cellule de niveau supérieur. Nous avons fixé le nombre des nœuds à 50 ; La vitesse maximale des nœuds est de 5 km/h.

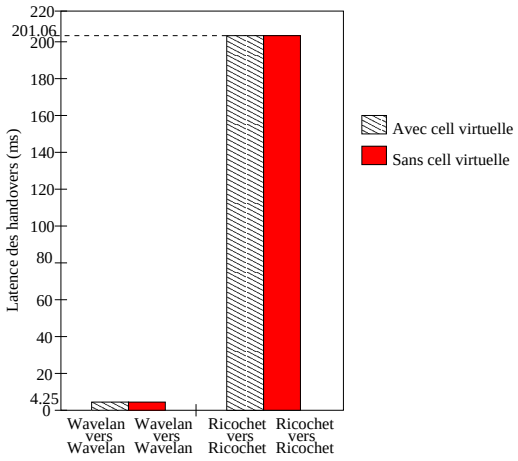


FIG. 3.20 – *La latence des handovers horizontaux.*

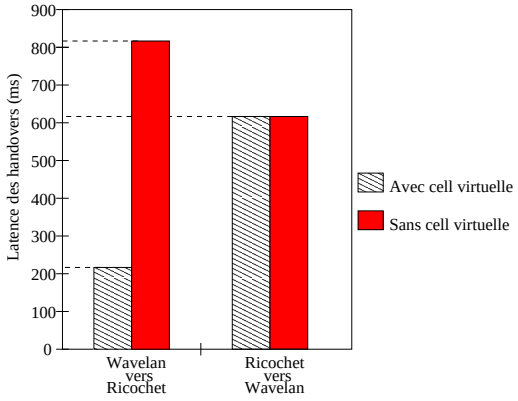


FIG. 3.21 – *La latence des handovers verticaux.*

La figure 3.20 montre une comparaison entre le système d'un handover horizontal sans et avec l'utilisation de la cellule virtuelle. Dans notre simulation, les handovers horizontaux possibles sont d'une cellule WaveLAN à une autre WaveLAN ou d'une cellule Ricochet à une autre Ricochet. La latence d'un handover horizontal à base de WaveLAN technologie est largement inférieure à celle d'un handover horizontal à base de Ricochet (Ricochet→Ricochet). 4,25 ms dans WaveLAN tendit que 201,6 ms dans Ricochet. Ceci est dû à la différence de la latence et la bande passante offerte par la technologie. La technologie WaveLAN a un débit plus important que celui de la technologie Ricochet et une latence très faible. Il est évident que la latence d'un système de base

d'un handover horizontal sans et avec l'utilisation de la cellule virtuelle est la même. La technique de la cellule virtuelle n'intervient que dans le cas d'un handover vertical montant.

L'efficacité des cellules virtuelles est illustrée dans la figure 3.21. Avec une période de 200 ms entre deux trames balises, les handovers verticaux montants d'une cellule WaveLAN à une cellule Ricochet réalisent une meilleure performance que les handovers verticaux descendants en termes de temps de latence. Un handover vertical montant (WaveLAN→Ricochet) consomme en moyenne 800,422 ms contre 600,285 ms pour un handover vertical descendant (Ricochet→WaveLAN), ce qui est conforme avec l'analyse présentée dans la section 3.6. L'utilisation des cellules virtuelles optimise les performances des handovers verticaux montant de 75%. La latence est passée de 800,422 ms à 200,321 ms en éliminant l'attente des 3 trames balises. Les cellules virtuelles ne peuvent pas prédire qu'un TM va recevoir T_B trames balises d'un niveau hiérarchique inférieur dans la OWN. Un TM connecté à une cellule Ricochet doit attendre impérativement la réception de 3 trames balises successives envoyée par la même cellule WaveLAN. Par conséquent, les cellules virtuelles n'influent pas sur le temps de latence des handovers verticaux descendants.

L'utilisation des cellules virtuelles s'avère très utile pour réduire le temps de latence durant les handovers. Une application classique pour mieux voir l'efficacité de cette technique est la téléphonie mobile. L'interactivité des conversations exige un délai de tolérance inférieur à une valeur maximale de 200 ms. La latence des handovers ne doit pas dépasser cette valeur maximale. Nous conservons la technologie sans fil WaveLAN de niveau hiérarchique 0 dans la OWN. Nous supposons que les cellules de niveau hiérarchique 1 dans la OWN offrent une bande passante de 1 Mb/s, une latence de 10 ms et un diamètre de 400 m. Chaque BS envoie périodiquement, chaque 100 ms, des trames balises. Afin de simuler le relayage en temps réel des paquets de téléphonie aux TMs, la GWFA génère et envoie des paquets de données de 200 octets chaque 20 ms. Les figures 3.22 et 3.23 montrent les temps séparant la réception de deux paquets successifs. Logiquement, l'espace de temps entre la réception des paquets est environ de 20 ms. Le premier paquet reçu après un handover, paquet de numéro de séquence 13, est espacé par un écart supplémentaire de 4,2 ms qui correspond à un temps de latence d'un handover horizontal au niveau hiérarchique 0 de la OWN. Le paquet de données avec 14 comme un numéro de séquence est reçu en un délai inférieur à 20 ms ce qui est logique vu que le paquet numéro 13 est reçu en retard. La latence des handovers horizontaux est la même avec ou sans l'utilisation des cellules virtuelles. Après la réception du paquet numéro 18, le TM effectue un handover vertical montant et par conséquent le paquet numéro 19 est reçu après un temps de 322 ms dont 3×100 ms représentant le temps de perte de 3 trames balises. Les paquets de données avec les numéros de séquence de 20 à 35 sont déjà disponibles dans le buffer de la nouvelle BS à cause du retard engendré par la transmission de paquet numéro 19. Le mécanisme de base des handovers verticaux ne permet pas l'interactivité des conversations puisque la latence introduite est supérieure à la valeur maximale autorisée. L'utilisation des cellules virtuelles réduit l'espace de temps entre le 18ème est 19ème paquet à 22 ms en éliminant la composante de latence $T_B \times N_B$.

D'après les numéros de séquence, aucun paquet n'est perdu ou arrivé dans le désordre. Ceci est assuré par les retransmissions des paquets après les handovers, la taille des buffers et une mobilité très faibles (2 m/s).

Lorsque les TMs se déplacent avec une vitesse très faible, le système des handovers est très efficace et assure une perte des paquets de données nulle. Avec une mobilité rapide, le nombre des handovers augmente et par conséquent, des paquets de données peuvent être perdus. Dans cette simulation, les systèmes sans fil WaveLAN et Ricochet sont considérés. L'*inter-arraival* est fixé à 60 ms et le nombre des TMs à 50. La figure 3.24 illustre une comparaison entre le système de base sans et avec la gestion de la mobilité rapide. Les courbes représentent le pourcentage de la perte des

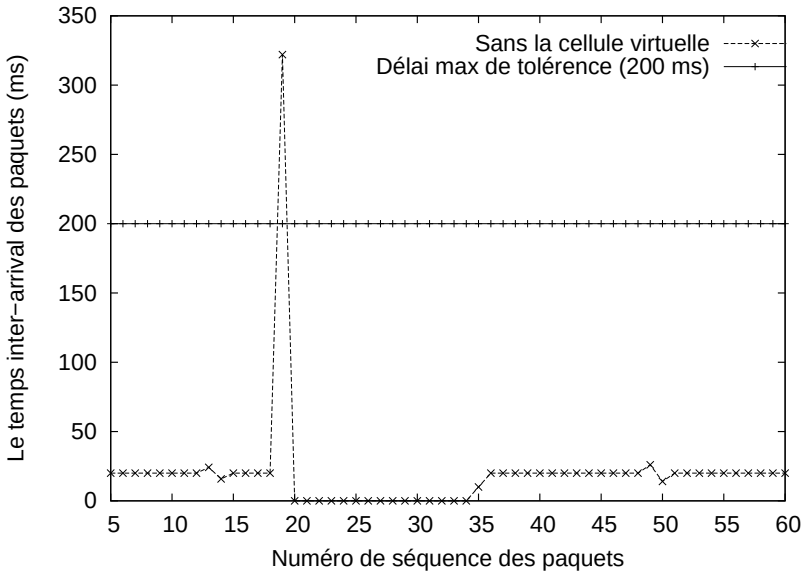


FIG. 3.22 – *Le temps d'arriver des paquets sans l'utilisation des cellules virtuelles.*

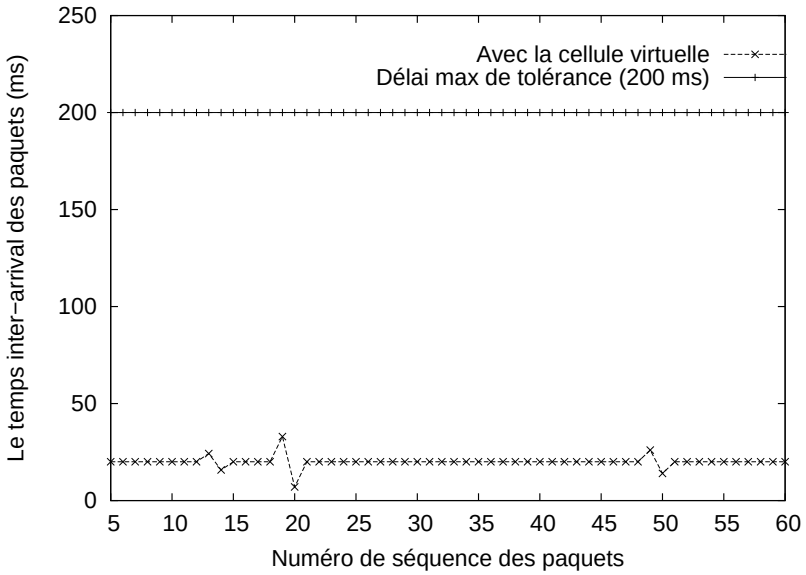


FIG. 3.23 – *Le temps d'arriver des paquets avec l'utilisation des cellules virtuelles.*

paquets de données en fonction de la vitesse de mobilité. Lorsque la mobilité est inférieure à 2 m/s, une perte nulle est réalisée. Avec l'augmentation de la vitesse, les pertes de données augmentent. Ceci est due essentiellement à la taille des buffers qui ne peuvent pas stocker pendant une large durée les données à retransmettre puisque les TMs qui se déplacent à grande vitesse changent les cellules rapidement et les anciens paquets de données sont écrasés par les nouveaux. La procédure de gestion de la mobilité rapide fixe des seuils de vitesse pour chaque niveau hiérarchique de la OWN. Ce seuil doit être bien choisi pour minimiser les pertes. Comme notre OWN contient deux niveaux, seulement le seuil de niveau 0 est significatif qui permet aux TMs dont la vitesse excède

ce seuil de passer au niveau hiérarchique 1. Nous avons considéré les seuils suivants : 2 m/s (1 handover chaque 200 s), 5 m/s (1 handover chaque 40 s) et 10 m/s (1 handover chaque 20 s). Les pertes engendrées par le système doté de la procédure de gestion de mobilité rapide avec un seuil de 2 m/s présente les meilleures performances. Les pertes augmentent avec l'augmentation du seuil. 2 m/s est le seuil le plus approprié pour le niveau 0.

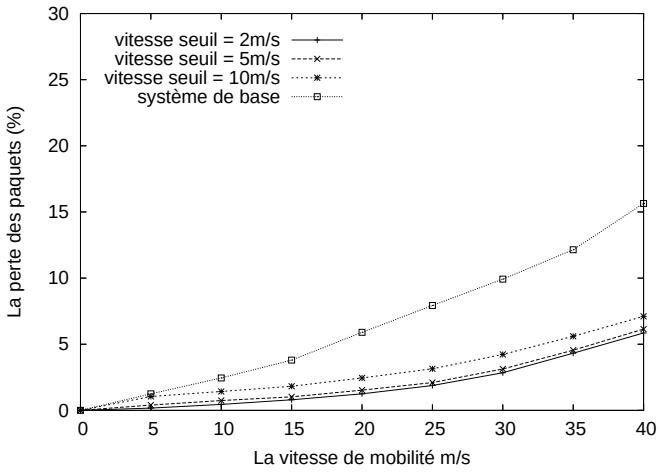


FIG. 3.24 – *La perte des paquets en fonction de la mobilté*.

Le tableau 3.2 montre les valeurs maximales de la latence et l'overhead engendrés par une émission rapide (100 ms) des trames balises.

Handover	L (ms)	B (bits/s)
WaveLAN→Ricochet	500,397	2480
Ricochet→WaveLAN	300,456	2480

TAB. 3.2: textitÉmission rapide des trames balises.

L'émission rapide des trames balises permet de réduire le temps de la latence contre une consommation large de la bande passante. À partir du tableau 3.3, la latence des handovers verticaux est inférieure à celle du système en prenant $T_D = 5$ pour le seuil des paquets (le choix de T_D est fait selon les heurustiques montrées dans [44]). Par contre, l'overhead généré est très important. L'émission rapide des trames balises présente des performances nettement supérieures à celles de diffusion multiple des paquets.

Handover	L (ms)	B (bits/s)
WaveLAN→Ricochet	500	546133
Ricochet→WaveLAN	500	546133

TAB. 3.3: *Diffusion multiple des paquets*.

La diffusion multiple des entêtes permet l'optimisation de l'overhead généré par les diffusions multiples des paquets (tableau 3.4).

Handover	L (ms)	B (bits/s)
WaveLAN→Ricochet	700,419	3733
Ricochet→WaveLAN	500,401	3733

TAB. 3.4: *Diffusion multiple des entêtes.*

3.11 Conclusion

L'utilisateur de la 4ème génération de mobiles a plusieurs technologies d'accès sans fil à sa disposition. Cet utilisateur veut pouvoir être connecté au mieux, n'importe où, n'importe quand et avec n'importe quel réseau d'accès. Pour cela, les différentes technologies sans fil doivent coexister de manière à ce que la meilleure technologie puisse être retenue en fonction du profil de l'utilisateur et de chaque type d'application et de service qu'il demande. Nous avons proposé une structure hiérarchique des réseaux sans fil et mobiles. Cette structure est composée de plusieurs niveaux selon les caractéristiques de chaque technologie sans fil. Chaque technologie est enveloppée par une autre de large zone de couverture mais avec un débit inférieur. L'ajout et/ou la modification de certaines composante et protocoles est nécessaire afin de réaliser l'interconnexion et d'assurer un transfert cellulaire transparent entre les différentes technologies par rapport aux utilisateurs.

Dans ce contexte, l'équipement terminal devra rechercher en permanence le meilleur réseau d'accès en fonction des besoins de l'utilisateur. C'est-à-dire, le terminal mobile cherche sans cesse à descendre dans la structure hiérarchique (OWN) en exécutant des handovers verticaux descendants. L'inaccessibilité du réseau d'attachement et les niveaux inférieurs obligent le terminal mobile d'effectuer des handovers verticaux montants après un temps considérable de déconnexion. Il est nécessaire de réduire la latence des handovers verticaux montants pour assurer la continuité de service sans dégradation. Nous avons proposé deux architectures qui permettent de récupérer les paquets de données perdus durant les handovers. L'utilisation de la technique des cellules virtuelles anticipe les handovers verticaux montants et réduit au maximum la composante L_D de la latence sans signalisation supplémentaire. D'autres méthodes plus agressives ont été proposées mais seulement applicables dans le contexte d'Arch-1. L'inconvénient principal de ces techniques est l'overhead qui peut dégrader les performances lorsque plusieurs terminaux mobiles exécutent ces techniques à la fois.