Network selection in heterogeneous wireless networks

Lusheng Wang

Télécom Paris (ENST)

THÈSE

Pour Obtenir le Grade de Docteur

De l'Ecole Nationale Supérieure

Des Télécommunications

Spécialité: **Réseaux et Informatique**

Lusheng WANG

Sélection de Réseau dans les Réseaux Sans Fil Hétérogènes

Date de Soutenance: 26 Janvier 2010

Composition du Jury:

Rapporteur	Christian Bonnet (Eurecom)
Rapporteur	Vicente Casares-Giner (UPV, Espagne)
Examinateur	Jean-Marie Bonnin (Télécom Bretagne)
Examinateur	Geng-Sheng Kuo (NCCU, Taiwan)
Examinateur	David Binet (France Télécom)
Directeur de Thèse	Daniel Kofman (Télécom ParisTech)

Acknowledgement

First and foremost, I would like to thank my supervisor Prof. Daniel Kofman for his support during the past three years and four months and for his helpful suggestions during the difficulties. His kind help is the main reason that I could complete this work, and the freedom he gave to me is the chance that I could learn how to do research independently.

I wish to express my gratitude to Prof. Michel Riguidel, the previous director of our department. His help, encouragement and trust greatly helped me during difficulties.

I am grateful to the members of the committee (Prof. Christian Bonnet, Prof. Vicente Casares-Giner, Dr. Jean-Marie Bonnin, Prof. Geng-Sheng Kuo and Mr. David Binet) for their time, support and comments.

I would like to thank the administrative personnel in Telecom ParisTech and especially INFRES department for their kind help during the past four years.

I would like to thank my dear friends in Telecom ParisTech, Myriam Morcel, Maryna Komarona, Quynh Huu Nguyen, Hai Lin, Bing Han, etc., for their warm-hearted help, especially when I just arrived in France and when I had difficulties in both work and life. Also, I want to thank my friends from the same school (Bin Liu, Liang Qian, Yifei Fu and Fei Cui) for their help during the reception of thesis defense. At the same time, I would like to express my appreciation to Mr. Foudil Bretel at CITI Lab (INSA & INRIA) for his kind help and time on the verification of my French resume. Taking this chance, I would like to thank all the friends who have ever helped me in both work and life during the past four years.

I would like to express my special appreciation to my parents for their support, encouragement and understanding of my persistence on study during the past twenty years.

To Wen who enriches my life with infinite support and love.

Abstract

In the context of present trend towards ubiquity of networks and global mobility of services, we notice that network access can be supported by a large diversity of access technologies with overlapped coverage. Networks based on these access technologies could be constructed at the same place, which compose an environment of heterogeneous wireless networks. Within this environment, mobile terminals (e.g. smart phones and laptops) and mobile networks (e.g. the local network formed by a mobile router in a bus or a train) usually have more than one interface to connect to the Internet. If they have protocols to support the connection to the Internet through multiple access interfaces with multiple IP addresses, they are called multi-homed mobile nodes (MNs). If they can only use one access interface at one time, they are called multi-mode MNs. No matter for any of the two types of terminals, the usage of a wrong access interface (and network) might result in obvious inconvenience, such as lack of bandwidth for video applications, poor mobility support for high-speed terminals, large signaling cost, traffic congestion, etc. Therefore, both multi-homed and multi-mode MNs require to be always connected to the best access network at any time anywhere, which is well known as ‘always best connected (ABC)’.

ABC brings plenty of advantages to customers and operators. To realize ABC, information of available networks and terminal-side entities should be gathered, and the best network should be selected. In the literature, various network selection schemes and mathematical models have been proposed for this task. In this dissertation, we firstly provide a survey of existing network selection schemes, which use various mathematical models. There are also integrated schemes which combine and take advantage of multiple mathematical theories. By analysis and comparison of these models, multi-attribute decision making (MADM) is considered as the most appropriate mathematical model for the network selection issue.

Then, a simulator for network selection is established using Matlab, which is capable of simulating a large number of scenarios and is going to be used frequently during this research. Simulation results show that MADM mathematical model works well in plenty of scenarios. Moreover, some important observations are summarized and several existing issues are identified. Main issues addressed in this dissertation include: usage of mobility-related factors, requirement of a subjective weighting method for evaluating various network properties’ importance, traffic load assignment during network selection, vertical handover (VHO) decision on whether a better network is worth handing-over to, etc.

First, the usage of mobility-related factors in the MADM-based network selection framework is studied. If these factors are not used or not correctly used, networks with poor mobility support capability might be selected for a high-speed MN, which disturbs its live applications and increases signaling cost. Our proposal for this issue can be divided into two parts: a first scenario with two networks; a second scenario, more generic, with *n* networks. Since VHO properties are related with the permutation of networks, the selection of the best network becomes the selection of the best permutation when VHO properties are taken into account. Network selection schemes for both scenarios are described and methods to get rapidly the best permutation for the generic scenario are discussed.

Second, the requirement of a new subjective weighting method is analyzed. Network selection is a fast and automatic procedure, but the traditional analytical hierarchy process is not suitable for calculating the subjective weights of various attributes in such a procedure. Hence, a TRigger-based aUtomatic Subjective weighTing (TRUST) method is proposed. Since a network selection procedure is triggered only when certain event happens, we suggest calculate the subjective weights based on the feature of the current trigger event. For example, video streaming applications require large bandwidth, so the weight of bandwidth should be large when the network selection procedure is triggered by an event 'video streaming starts'. Similarly, terminals with high speed require a network with good mobility support feature, so mobility-related attributes should have large weights when an event 'terminal speed becomes high' triggers the network selection procedure.

Third, mobility signaling cost in HMIPv6 networks is evaluated, which shows the way to evaluate total handover cost in the proposed best permutation scheme. Meanwhile, based on the evaluated mobility signaling cost, a mobility anchor point (MAP) selection scheme for HMIPv6 networks is proposed, which minimizes the total additional cost of HMIPv6, including mobility signaling cost and packet tunneling cost.

Finally, several other issues of network selection are analyzed and possible solutions are proposed, including traffic load assignment during network selection, vertical handover decision schemes, etc. Based on all the studies above, we also propose a four-step integrated strategy for MADM-based network selection, which takes advantage of mobility-related factors, uses our efficient weighting method TRUST, combines the analysis on traffic load assignment, and performs VHO decision before handing-over to the best network.

Résumé

R.1. Introduction

R.1.1. Réseaux sans fil hétérogènes et la sélection du réseau

Au vu de la tendance actuelle vers l'ubiquité de communication et la mobilité globale de service, différents réseaux d'accès sans fil peuvent être déployés et utilisés. Ces réseaux ont des propriétés différentes, et donc peuvent être utilisé à supporter les services différents. D'ailleurs, il est également commun d'avoir plusieurs réseaux construits au même endroit pour fournir un meilleur service aux clients, telle que la complémentarité de 3G et WiFi disponible actuellement. Pour distinguer ces multiples réseaux par rapport à un unique réseau, nous les appelons « Réseaux sans Fil Hétérogènes (RFHs) ».

Pour les RFHs, des nouvelles méthodes pour gérer la mobilité sont nécessaires afin de supporter le « handover » vertical. Des nouvelles méthodes de « multihoming » sont aussi nécessaires pour supporter la sauvegarde de transmission et le partage de la bande passante. Mais ce n'est pas suffisant! Il faut encore déterminer le meilleur réseau pour le service actuel, à tous moment et n'importe où. Ce concept est connu comme « Always Best Connected (ABC) en anglais».

Avec la fonctionnalité ABC, MNs pourrait choisir les réseaux d'accès pour des applications diverses pour s'adapter à leurs exigences de qualité de service; MNs peuvent éviter de choisir un réseau à grande trafic donc réaliser l'équilibrage de charge et éviter les embouteillages; MNs peuvent prévoir la disponibilité des réseaux, afin de ne pas se connecter à des réseaux qui disparaissent rapidement; MNs peuvent optimiser le coût de signalisation par la conception de leurs stratégies de la sélection du réseau et handover. En un mot, ABC apporte beaucoup d'avantages aux clients. En outre, les opérateurs bénéficient ABC, aussi. Grâce à la sélection du réseau et l'équilibrage de la charge de la fonctionnalité ABC, les opérateurs peuvent maximiser l'utilisation de leurs réseaux, donc maximiser les recettes; basé sur ABC, les opérateurs peuvent analyser et decider leur stratégie de la déploiement de point d'accès WiFi pour attirer les clients au réseau WLAN. Enfin, ABC est convient de considérer synthétiquement les avantages de clients et opérateurs, afin qu'un partenariat gagnant-gagnant peut être atteint.

Afin de laisser nos terminaux ABC, une première tâche importante considère à définir le « meilleur » réseau. Cette définition est liée à des facteurs très nombreux, tels que le terminal, le client, le service et les propriétés de réseaux. Une autre tâche considère ensuite à appliquer cette définition afin de sélectionner le meilleur réseau, voir la figure R-1.

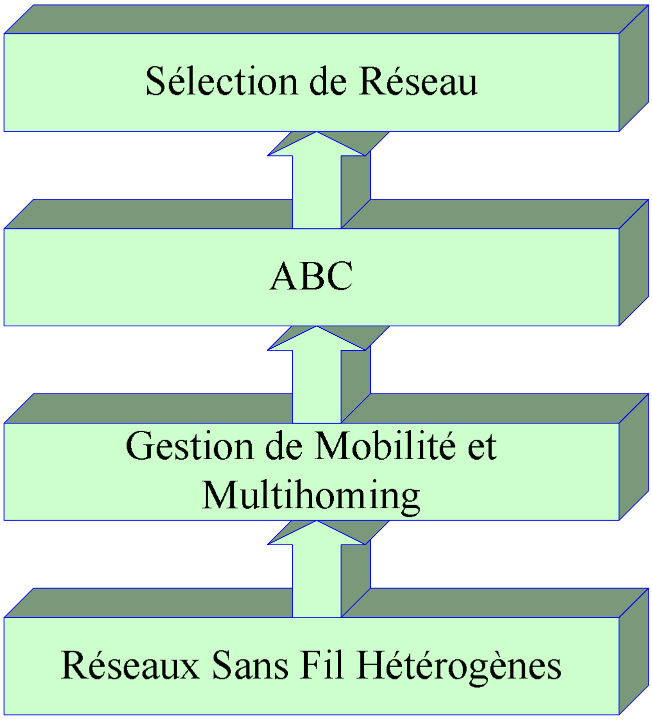


Figure R- 1 Pourquoi nous avons besoin de la sélection du réseau.

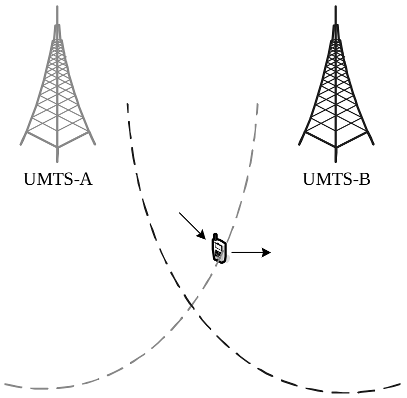
R.1.2. Exemples de la sélection du réseaux

La sélection du réseau est définie comme le choix du meilleur réseau en fonction de multiples facteurs lorsque plusieurs réseaux sont disponibles. Cette procédure est activée quand le client a besoin d'un nouveau service ou quand la communication actuelle devrait changer de réseau.

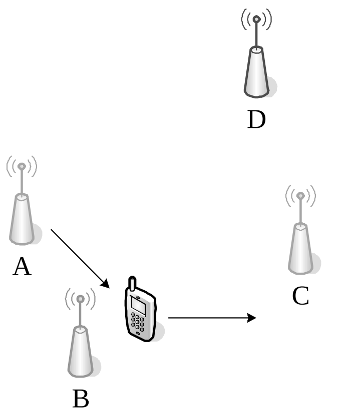
La détermination du handover dans un unique réseau peut être facile, mais la sélection du réseau dans les RFHs est très difficile. Voici deux exemples de détermination du handover dans un unique réseau :

La figure R-2(a) montre la détermination du handover traditionnel entre deux stations de base. Quand la différence de force du signal entre la station de base en bleu et la station de base en orange dépasse un certain seuil, le handover est déclenché. Ainsi, la force du signal est le seul facteur que nous utilisons ici.

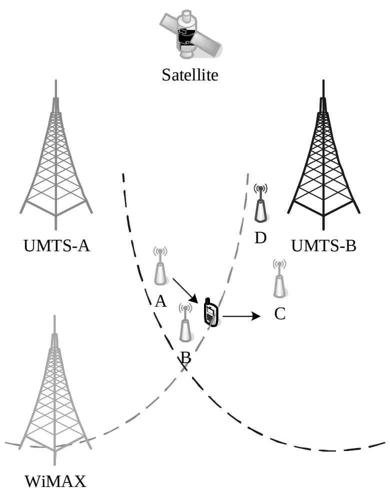
La figure R-2(b) montre la sélection du PA WiFi. Nous supposons que le terminal a utilisé un PA A et l'ajouté dans sa liste de préférence. Ensuite, il trouve d'autres PAs (B, C et D), et B a le plus fort signal. Le terminal sera toujours connecté automatiquement à A s'il est disponible, parce que A est dans la liste de préférence, mais B est inconnu. Cet exemple nous montre que la force du signal n'est pas toujours un facteur décisif. La préférence du client fondée sur son histoire de connexion peut être également importante.



(a) le handover entre deux stations de base



(b) la sélection du PA WiFi



(c) la sélection du réseau dans les RFHs

Figure R- 2 Les exemples de la sélection du réseau.

Par conséquent, si le terminal est sous la couverture des RFHs comme dans la figure R-2(c), la sélection du réseau devient plus compliquée. Plusieurs groupes de facteurs devraient être examinés : les propriétés du réseaux, tels que le prix, la bande passante et la force du signal ; les préférences du client, telle que la liste de préférence dans le dernier exemple ; les exigences des services, parce que les différents réseaux peuvent convenir à des services différents ; les propriétés du terminal, telles que la vitesse et la batterie ; les politiques de l'opérateur, tel que le contrôle du débit par l'opérateur. Ainsi, nous allons expliquer ci-dessous comment cette décision peut être élaborée en fonction de toutes ces facteurs.

R.1.3. Contexte mathématique pour ce sujet

Différentes théories mathématiques ont été utilisées pour ce problème ces dernières années. Notamment,

- la théorie de l'utilité
- L'aide à la décision multiattribut (ADMA)
- la logique floue
- la théorie des jeux
- les modèles NP-complets

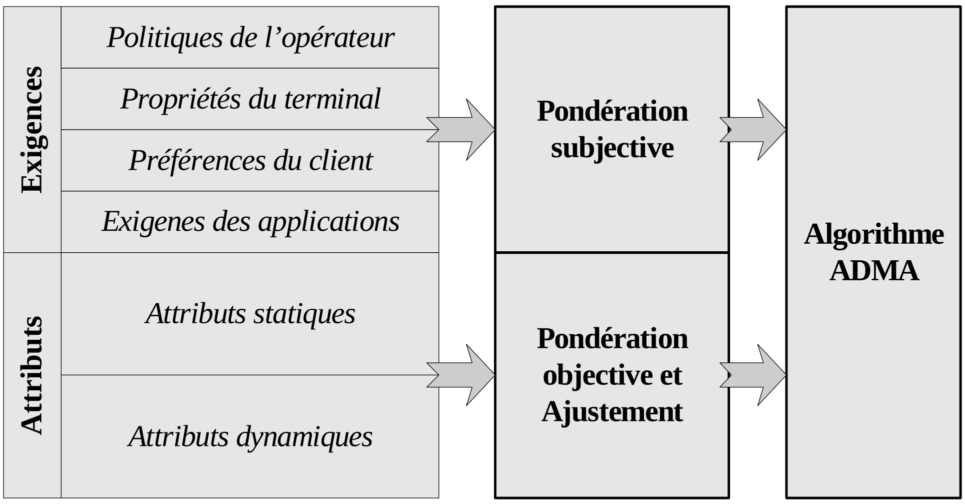


Figure R- 3 Une solution intégrée pour la sélection du réseau.

Parmi toutes ces théories, nous avons écarté les modèles NP-complets, parce que la sélection du réseau devrait être une procédure très rapide, or les modèles NP-complets ont généralement besoin d'une certaine durée pour trouver la solution optimale. Considerons par exemple le modèle du « sac-à-dos », il considère à mettre plus de services dans des bandes passantes limitées. Il est donc approprié pour améliorer l'utilisation des ressources, mais peu apte à garantir toutes les services en ABC, puisque la performance pour les services n’est pas le critère premier. Pour la même raison, nous avons écarté la théorie des jeux, qui peut être appropriée pour la répartition de charge mais pas pour l’ABC. Avec la théorie des jeux, les réseaux (joueurs dans un jeu) sélectionnent ses appropriés services ; mais pour l’ABC, il faut que les services sélectionnent ses meilleurs réseaux.

Nous souhaitons souligner que plusieurs de ces théories pourraient être utilisées ensembles, composant ainsi une solution intégrée. La figure R-3 montre une intégration possible, qui est une manière commune pour la sélection du réseau. Les facteurs subjectifs sont appelés ici les « exigences », et comprennent quatre groupes : les politiques de l’opérateur, les propriétés du terminal, les préférences du client et les exigenes des services. Ces facteurs sont utilisés pour calculer les pondérations subjectives, en utilisant par exemple le « processus hiérarchique analytique (PHA) ».

Il y a aussi des facteurs objectifs, qui sont les attributs des réseaux, y compris les attributs statiques et dynamiques. Ces attributs ont deux usages. Tous d’abord, ils sont utilisés pour calculer les pondérations objectives. Deuxièmement, ces attributs doivent être adaptés pour être combinés ultérieurement. L’ajustement peut être basé sur la normalisation, la logique floue ou la théorie de l’utilité. Enfin, ces attributs seront combinés en fonction de leur poids, en utilisant des algorithmes ADMA pour obtenir un rang.

R.1.4. Problèmes à résoudre

La solution intégrée ci-dessus présente encore beaucoup de problèmes. Quatre problèmes sont identifiés dans cette thèse :

- L’utilisation de facteurs de mobilité (en particulier les propriétés de handover vertical)
- Exigence d’une méthode de pondération automatique et subjective pour calculer les poids subjectifs de tous les paramètres
- La répartition de charge pendant la sélection du réseau dans le modèle MADM
- Décision de handover vertical fondée sur la prévision après la sélection du réseau

Pour les deux premiers problèmes, nous éxposerons nos propositions préciésément. Pour les deux derniers, nous analyserons généralement la possibilité de trouver une solution.

R.2. Propositions

R.2.1. Sélection du réseau basée sur la mobilité

R.2.1.1. Introduction du concept de « permutation »

Comme nous l'avons déjà présenté, un schéma de la sélection du réseau considère généralement plusieurs groupes de facteurs en même temps. Certains facteurs sont liés à la mobilité, tels que le rayon des cellules, la couverture, la vitesse terminale, les propriétés du handover vertical (HOV) ou handover horizontal (HOH), etc.

L'utilisation des propriétés du HOV est complexe. C'est parce que les propriétés du HOV ne dépendent pas seulement de la mobilité du terminal ou de la couverture des réseaux différents, mais aussi de la « permutation » des réseaux. Une *permutation* est un ordre de réseaux qui représente leurs priorités, sans tenir compte de leurs disponibilités. À tout moment et n'importe où, le permier réseau disponible dans la meilleure permutation devrait être utilisé.

Par exemple, dans un environnement avec 3 réseaux (UMTS, WiMAX et WLAN), le schéma choisit le meilleur réseau basé sur les coûts totaux des réseaux, y compris les coût du HOV et des autres coûts. Afin d'expliquer l'idée de permutation clairement, nous ne considerons ici que le coût du HOV. Avec 3 réseaux, il y a 6 permutations, énumérés ci-dessous :

- UMTS > WiMAX > WLAN
- UMTS > WLAN > WiMAX
- WiMAX > UMTS > WLAN
- WiMAX > WLAN > UMTS
- WLAN > UMTS > WiMAX
- WLAN > WiMAX > UMTS

La première permutation correspond à « un coût du HOV nul» parce que l'UMTS est supposé toujours disponibles en raison de son ubiquité. Par contre, la dernière permutation correspond à « un coût du HOV élevé». Par conséquent, 6 permutations correspondent à 6 coûts du HOV, et celui avec le coût du HOV minimum est le meilleur (la première et la seconde dans la liste).

Cependant, il reste un problème évident. Quand l'environnement hétérogène est composé de *N* réseaux, le nombre de permutations est la factorielle de *N*.

Donc, un certaine durée est nécessaire pour trouver la meilleure permutation. En outre, l'évaluation des propriétés du HOV de chaque permutation est également compliquée en raison de la couverture irrégulière des réseaux et des divers modes de mobilité du terminal. Pour résoudre ce problème, nous proposons le schéma de la meilleure permutation en deux étapes. Dans la première étape, nous classifions tous les réseaux en deux groupes. Dans la deuxième étape, nous examinons un cas générique avec *N* groupes de réseaux.

R.2.1.2. Cas de 2 groupes de réseaux

Nous commençons par la modélisation du mouvement du terminal dans un environnement de 2 groupes de réseaux. Dans certaines études récentes, la fonction sigmoïde est recommandée pour ajuster les utilités, au lieu de la fonction linéaire. Nous pouvons concevoir que la fonction sigmoïde à tendance à séparer tous les réseaux en 2 parties. Il est donc raisonnable de diviser les réseaux en 2 groupes. Par exemple, dans cette thèse, les 2 groupes sont les réseaux ubiquitaires (RUB) et les réseaux hotspots (RHS).

La figure R-4 ci-dessous montre le déploiement de *K* points d'accès (PAs) dans une zone carrée comme l'espace de simulation. Les *K* PAs sont distribuées de façon aléatoire. Le déploiement des PAs est supposé indépendant des réseaux ubiquitaires. En outre, *N* terminaux avec un mouvement aléatoire sont uniformément distribués.

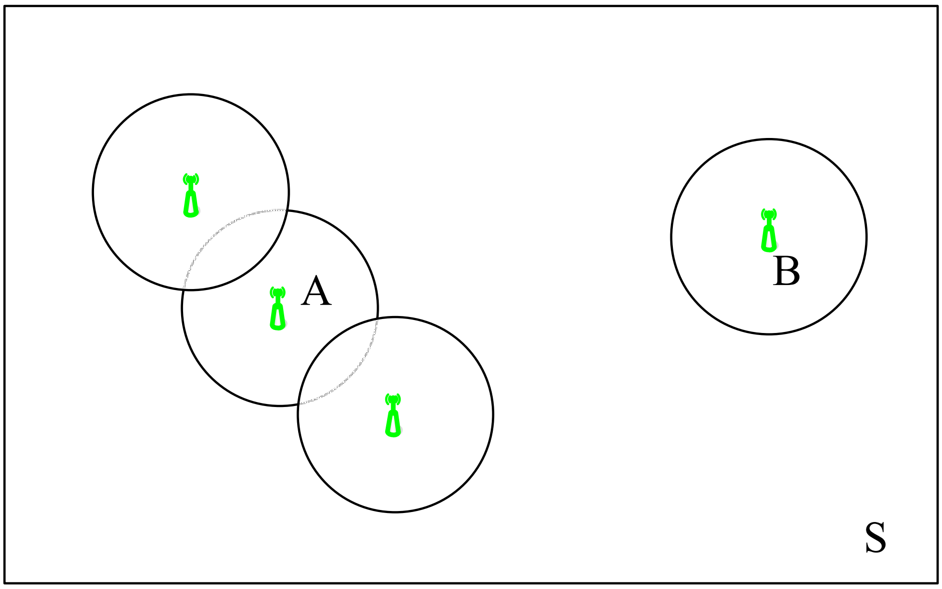


Figure R- 4 Déploiement des PAs.

Quand un terminal avec un mouvement aléatoire sort de l'un de ces PAs, la probabilité de transiter directement à un autre PA est exactement égal au pourcentage constitué par la bordure de la PA délimitée par le recouvrement des autres PAs sur lui. Comme montré dans la figure R-4, la bordure du PA « *A* » est couverte par deux autres PAs, donc la probabilité de transiter directement de ce PA à un autre est grande. Par contre, au moment de quitter le PA « *B* », le terminal n'a aucune chance de transiter directement à un autre PA.

Ainsi, en raison du caractère stochastique de la distribution des PAs, la probabilité de transit P peut être définie comme suit :

$$P \approx 1 - Q_{K-1} \approx 1 - Q \ ,$$

où Q est la couverture des K PAs dans la zone de simulation. La simulation Monte Carlo sera utilisée pour vérifier cette formule.

D'ailleurs, nous avons expliquer « l'effet de la bordure de la simulation Monte Carlo ». C'est à dire que les PAs pourraient être distribués à côté de la bordure de la zone de simulation, causant ainsi le résultat de la simulation imprécis. Lorsque Q est grand, cet effet ne doit pas être ignoré.

Etudions maintenant en détail les transitions entre les RHSs et les UBNs. Nous utilisons « *d* » pour représenter la zone couverte par aucun PA et « *A* » pour la zone d'un PA; « PA_1 » à « PA_K » sont les K PAs; et « *S* » représente toute la zone de simulation; « *Ua* » est le taux de transit d'un PA à l'extérieur; et « *Ud* » est le taux de transit de la zone « *d* » à un PA.

Nous supposons que tous les terminaux sont distribués uniformément et d'un mouvement aléatoire. Nous considerons l'effet de bordure de la zone de simulation comme négligeable. En outre, quand un terminal sort de la zone de simulation par un côté, il entre par l'autre côté (généralement appelé « Wrap Around » en anglais). Par conséquent, la distribution des terminaux est toujours stable. Nouv obtenons :

$$U_d(1-Q) = U_a P \frac{A}{S} K \ .$$

Considérons une vaste zone couverte par un certain nombre de PAs. Si ces PAs ne sont pas densément déployés, Q n'est pas assez grand, nous avons donc le résultat suivant

$$U_d \approx Q U_a \ .$$

Après modélisation de la mobilité du terminal, nous allons utiliser ce modèle pour aider la sélection du réseau. Certain algorithme ADMA doit être utilisé pour combiner le coût moyen de handover avec des autres coûts. Enfin, la permutation avec le coût total le plus bas sera sélectionnée.

Tableau R- 1 Le taux et le coût du HOV et HOH.

Handover	Taux du Handover	Coût par Handover
RUB à RUB	$Ua\ r / R$	X
RHS à RHS	$Ua\ (1 - P)\ Q$	Y
RHS à RUB	$Ua\ P\ Q$	Z_1
RUB à RHS	$Ud\ (1 - Q)$	Z_2

Les taux du handover pour HOH et HOV sont énumérés dans le tableau ci-dessus. Avec deux groupes de réseaux, il y a deux permutations: RUB > RHS ou RHS > RUB. Le coût moyen de handover de chaque permutation est calculé. Pour la première permutation, un réseau ubiquitaire sera toujours utilisé. Ainsi, le coût moyen de handover n'est constitué que du coût HOH entre les cellules du réseau ubiquitaire. Et pour la seconde permutation, le coût moyen de handover est constitué de quatre parties. A la fin, nous pouvons comparer les deux types de coûts totaux pour obtenir un seuil entre les deux permutations ci-dessous

$$W_2 < \frac{\overline{Oth}_{RUB-RHS}}{\overline{Oth}_{RUB-RHS} + \dfrac{\overline{HC}_{RHS>RUB} - \overline{HC}_{RUB>RHS}}{Q}},$$

où $\overline{Oth}_{RUB-RHS}$ est la différence entre les RUBs et les RHSs considérant la combinaison de tous les autre coûts, $\overline{HC}_{RHS>RUB}$ et $\overline{HC}_{RUB>RHS}$ sont les coût moyen de handover pour les deux permutation « RHS > RUB» et « RUB > RHS». Comme nous pouvons voir sur ce seuil, quand tous les facteurs objectifs sont fixés, le poids subjectif du coût moyen de handover W_2 sera décisif pour déterminer la meilleure permutation. Quatre algorithmes seront utilisés dans la comparaison, qui sont

- la pondération additive simple
- la pondération exponentielle multiplicatif

- la technique d'ordre de préférence selon leur similarité à la solution idéale
- la gris analyse relationnelle

Nous définissons la densité de probabilité du terminal comme $f(V)$, et le poids du coût de handover comme $W(V)$. Ainsi, un terminal qui a une vitesse de V_0 ci-dessous va préferer WLAN :

$$V_0 = W^{-1}\left(\frac{\overline{Oth_{UBN-HSN}}}{\overline{Oth_{UBN-HSN}} + \frac{QY + (1-Q)(Z_1 + Z_2)}{Norm}}\right).$$

De plus, le montant de terminals qui préfèrent WLAN peut-être écrire comme

$$n(Q) = Q \cdot N_0 \int_0^{V_0} f(V)dV \ ,$$

où N_0 est le montant total de terminal, et V_0 est comme ci-dessus. En même temps, nous pouvons obtenir le ratio d'augmentation de terminals qui préfèrent WLAN.

R.2.1.3. Une étude générique avec *N* groupes de réseaux

Comme expliquer au début, le nombre de permutations est la factorielle du nombre de réseaux, donc il prend une certaine durée pour calculer les coûts totaux de toutes les permutations. Par conséquent, la première tâche est de trouver la résolution pour réduire cette durée.

Tous d'abord, nous pouvons prendre l'avantage du RUB. En supposant qu'au moins un RUB existe, les réseaux derrière du RUB dans une permutation ne vont jamais être utilisé. Par conséquent, leur ordre n'est pas important. Grâce à cette raison, seulement une partie de permutations doivent être prises en compte.

Plus important que cette idée, nous proposons d'utiliser une méthode combinée. La première étape de notre méthode considère à la sélection du meilleur réseau et le handover au meilleur réseau, qui est rapide et ne perturbe pas les communications actuelles. La deuxième étape considère à la sélection de la meilleure permutation, qui est lente, mais obtient la meilleure permutation, montrer comme la figure R-5. Si le « meilleur réseau » obtenu dans la première étape n'est pas le premier réseau disponible dans la meilleure permutation obtenue dans la deuxième étape, ça cause un HOV supplémentaire. La façon d'améliorer cette méthode est de diminuer la probabilité du HOV supplémentaire. Donc, nous l'améliorons en utilisant un schéma de la sélection du réseau basé sur les permutations dans la première étape. Dans ce schéma, nous ne comparons pas les coûts totaux de différents réseaux. Au lieu de cela, nous comparons les coûts totaux des deux permutations. Par exemple, le coût total de permutation « $A > B > C > D > E$ » et cela de « $B > A > C > D > E$ » sont comparés pour décider si A ou B est le meilleur. De cette façon, nous n'avons besoin que (N - 1) comparaisons, mais les facteurs de mobilité sont pris en compte.

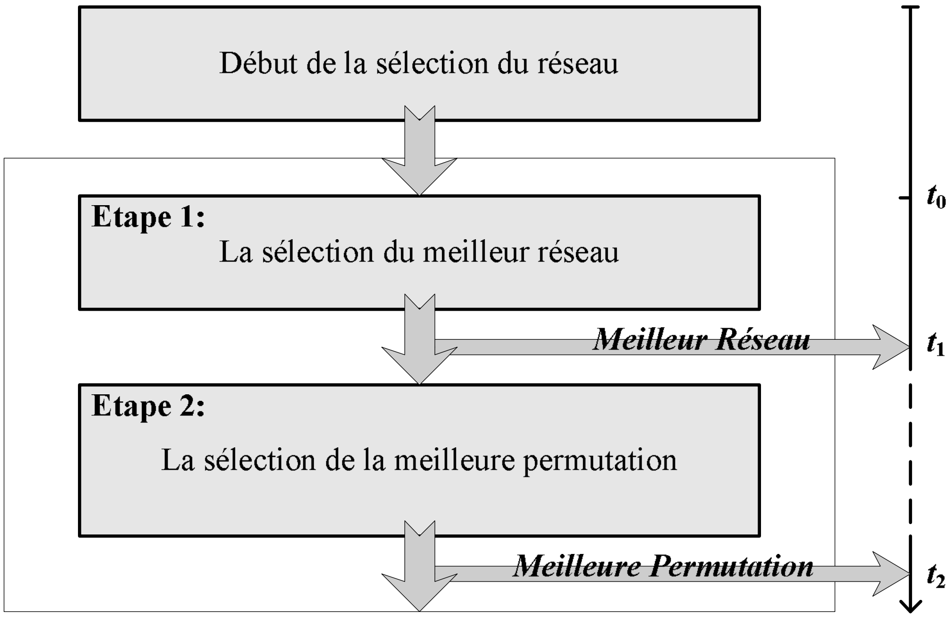


Figure R- 5 Une méthode combinée.

Après avoir résolu le problème de coût du temps, nous avons aussi besoin d'évaluer le coût total de chaque permutation, en particulier le coût moyen de handover. Ici, nous utilisons un cas de trois réseaux pour expliquer comment nous modélisons la mobilité du terminal dans les RFHs. Un état dans la figure ci-dessous est défini comme le séjour d'un terminal dans la zone couverte par les mêmes groupes de réseaux, tel que l'état « *AB* » signifie que le terminal est couvert par le réseau « *A* » et le réseau « *B* ».

En supposant « $A > B > C$ », nous pouvons simplement considérer les transitions causant des HOVs. Nous combinons les états qui ont le même nombre de réseaux et le même meillieur réseau comme un grand état. Par exemple, l'état « *AB* » et l'état « *AC* » forment un grand état. Ainsi, le nombre total des états diminue. Comme nous pouvons voir dans la figure R-

6, les états dans la même rangée ont le même meilleur réseau. Un HOV est nécessaire lorsqu'un terminal transite vers un état qui possède un meilleur réseau (d'en bas à gauche vers en haut à droite) ou un état qui ne possède pas le meilleur réseau précédent (d'en haut à droite vers en bas à gauche). Par conséquent, il n'est pas nécessaire d'examiner les autres transitions, ainsi le nombre de transitions est diminué. Dans le schéma proposé, le coût moyen de handover de chaque permutation est calculé en fonction de ce modèle.

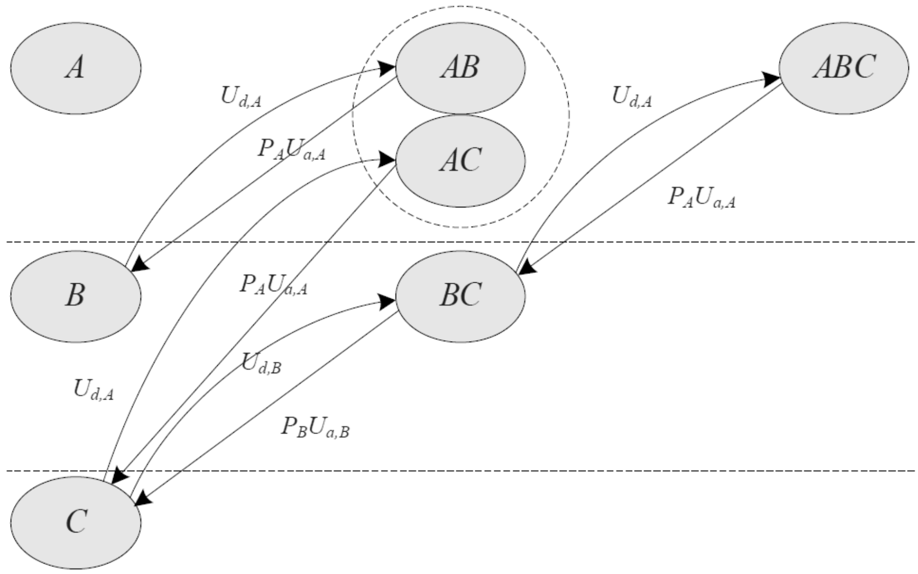


Figure R- 6 Les transitions causant des HOVs dans un exemple de trois réseaux.

Le coût total d'une permutation est calculé en combinant des coûts de handover et des autres coûts. Le coût de handover comporte les coûts HOHs et les coûts HOVs. Pour les coûts HOVs, nous utilisons HC^{+} pour représenter la transition à un nouveau meilleur réseau et HC^{-} pour représenter la transition hors le meilleur réseau actuel. Le coût total est écrit comme

$$TC = (HC_{HOH} + HC^{+}_{HOV} + HC^{-}_{HOV}) \cdot w_{HC} + \overline{Oth}_{RUB-RHS} \cdot (1 - w_{HC}) .$$

où w_{HC} représente le poids du coût du handover total.

Ce nouveau schéma est comparé avec les anciens schémas de la sélection du réseau. Basé sur cettes comparaisons, notre proposition est évidemment préférable pour trois critères suivants : le taux HOV, le coût total et le taux de déclenchement du schéma.

R.2.2. Pondération fondée sur des déclencheurs

Jusqu'à maintenant, nous avons résolu le problème de l'utilisation des facteurs de mobilité, mais ce n'est pas suffisant pour utiliser notre proposition dans les termiaux réels pour sélectionner le meilleur réseau. C'est parce qu'il y a encore un grand problème sans résolution : la pondération subjective.

La pondération de la sélection du réseau est modélisée ci-dessous. Une matrice de décision est utilisée pour contenir toutes les informations d'attributs des réseaux pour *m* réseaux et *n* attributs.

Wo et *Ws* sont utilisés pour représenter les poids objectifs et subjectifs, respectivement. Les poids combinés sont calculés comme

$$\mathbf{W} = \frac{\mathbf{W_O} \cdot \mathbf{W_S^\Delta}}{\mathbf{W_O} \cdot \mathbf{W_S^T}} = \left[w_1 \quad w_2 \quad \ldots \quad w_n \right].$$

Les poids objectifs sont calculés basés sur les différences entre réseaux. Pour obtenir les poids objectifs, la méthode « entropie » et la méthode « variance » peuvent être utilisées.

Les poids subjectifs sont calculés basés sur les sentiments subjectifs du décideur. Pendant le processus de décision, le décideur peut être une personne ou un groupe d'experts. Ils ont des sentiments sur l'importance des attributs en fonction de leurs préférences ou expériences. Pour obtenir les poids subjectifs, la méthode « eigenvector » est vastement utilisée dans les études sur la sélection du réseau. Une autre méthode mentionnée est le « moindre carré pondération ». Tous ces deux méthodes ont besoins de la matrice de comparaison par paire, ce qui est obtenu par le décideur basée sur ses sentiments subjectifs.

La production de cette matrice est totalement subjective. Cependant, dans la procédure de la sélection du réseau, le meilleur réseau ne doit pas être choisi par le client. Par contre, il devrait être choisi automatiquement par le terminal. Donc, nous nous demandons comment une machine pourrait avoir des sentiments subjectifs pour produire cette matrice.

Une façon que nous pouvons imaginer est de stocker toutes les matrices possibles dans le terminal à l'avance, mais il y a vraiment des milliers matrices ou plus. Ainsi, cette façon n'est pas économique. D'ailleurs, ce n'est pas facile pour rechercher la matrice nous avons besoin. Donc, une méthode de pondération subjective pour la sélection du réseau devrait être « automatique », « rapide », « économiques » et bien sûr « précis ».

Dans cette thèse, nous allons proposer une nouvelle méthode, appelée la « pondération subjective et automatique fondée sur des déclencheurs ». Nous expliquons les déclencheurs tous d'abord.

La procédure de la sélection du réseau ne doit pas être périodique, par contre, elle devrait être déclenchée par certains « événements », notamment

- l'avènement d'un nouveau service
- la fin d'un service antérieure
- l'altération évidente de certaines propriétés du terminal
- le changement de la préférence du client ou de politique de l'opérateur
- le changement de la valeur de certains attributs dynamique, tel que le trafic

Nous voyons que ces événements peuvent déclencher la procédure de la sélection du réseau, mais nous ne savons pas pourquoi un événement peut déclencher cette procédure. En d'autres termes, quel est l'effet d'un événement au résultat de la sélection du réseau? Prenant la vitesse de terminal comme un exemple, l'augmentation de la vitesse cause un sentiment subjectif que les attributs de mobilité devraient être plus importants. Par conséquent, les poids subjectifs de ces attributs doivent augmenter, qui peut cause le changement du meilleur réseau (ou de la meilleure permutation). Donc, dans notre proposition, nous allons étudier la relation entre un événement et le changement de poids subjectifs.

Dans cette nouvelle méthode, les poids subjectifs sont calculés basés sur trois paramètres. Un paramètre est la matrice **EA**, qui représente la relation entre les événements et les importances des attributs. Par exemple, un nouveau service de vidéo va renforcer l'importance de la bande passante. Ainsi, la matrice **EA** peut-être écris comme

$$\mathbf{EA} = \begin{bmatrix} c_{11} & c_{12} & . & c_{1n} \\ c_{21} & c_{22} & . & c_{2n} \\ . & . & . & . \\ c_{k1} & c_{k2} & . & c_{kn} \end{bmatrix},$$

où c_{ij} représente l'effet du $i$ème événement sur le $j$ème attribut, et la valeur de c_{ij} sera soit 1 soit 0.

Un autre paramètre est le vecteur $\mathbf{W_E}$, qui représente le poids de ces événements. Par exemple, les services Streaming sont plus difficiles à desservir que les services Background, ainsi le poids de Streaming est beaucoup plus grand que cela de Background.

Les poids de ces événements ne sont pas facile de calculer, mais ils sont déjà plus facile que la calculation des poids des attributes. C'est parce que les poids de ces événements ne changent pas beaucoup. Dans ce cas, les poids de tous les événements seraient calculés en avance et gardés par le terminal mobile. Nous écrivons $\mathbf{W_E}$ comme

$$\mathbf{W_E} = [we_1 \quad we_2 \quad \ldots \quad we_k],$$

et il sera calculé ci-dessous.

Une hiérarchie de tous les événements est formée comme montrer dans la figure ci-dessous. Il y a deux niveaux dans cette hiérarchie. Pour le niveau dessus, nous avons

$$\mathbf{W_{E1}} = [we1_1 \quad we1_2 \quad \ldots \quad we1_{k1}],$$

et pour le niveau dessous, nous avons

$$\mathbf{W_{E2i}} = [we2_{i,1} \quad we2_{i,2} \quad \ldots \quad we2_{i,k2}],$$

où *i* représente le *i*ème groupe dans la figure.

Les poids intégrés peuvent-être écris comme

$$we_{i,j} = we1_i \cdot we2_{i,j},$$

où $we_{i,j}$ rerpésente le *j*ème événement dans le *i*ème groupe.

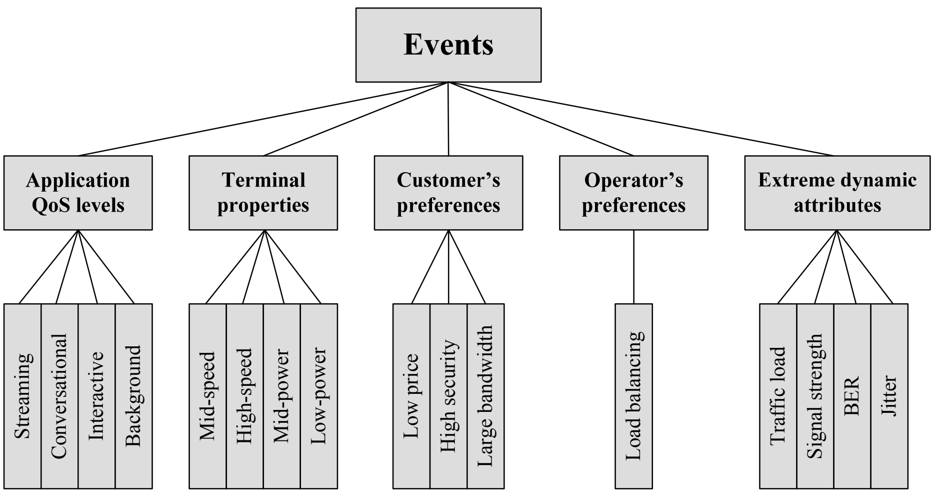


Figure R- 7 La hiérarchie des événement.

En outre, ces événements ci-dessus ne sont pas seulment événements, mais nous pouvons dire qu'ils sont aussi les états, en raison qu'ils durent pendant une certaine période. En détail, pour déclencher la procédure de la sélection du réseau, ils sont les « événements », mais après, ils sont les « états ». Ainsi, à un moment, il est possible que plusieurs de ces états peuvent être vrais. Lorsque la procédure est déclenchée, tous les états sont vérifiés et une matrice **TF** est obtenue, représentant le vrai ou faux de chaque état, écris comme

$$\mathbf{TF} = \begin{bmatrix} tf_{11} & 0 & . & 0 \\ 0 & tf_{22} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & tf_{kk} \end{bmatrix},$$

où tf_{ii} représente le vrai ou faux de état i. Pour une application, tf_{ii} représente combien de cett type d'applications sont encore utilisées.

Enfin, les poids subjectifs sont calculés à partir de la matrice **TF**, la matrice **EA** et le vecteur **WE**, indiqué comme suit :

$$\mathbf{W_S} = [ws_1 \quad ws_2 \quad \ldots \quad ws_n] = \mathbf{WE} \cdot \mathbf{TF} \cdot \mathbf{EA},$$

où $ws_j = \sum_{i=1}^{k} we_i \cdot tf_{ii} \cdot c_{ij}$ est le poids subjectif pour le *j*ème attribut.

Cette nouvelle proposition est simple, mais il correspond bien les quatre buts d'une pondération subjective :

- automatique : parce qu'il n'a pas besoin de comparaison par paire
- rapide : parce que le vecteur **WE** et la matrice **EA** sont obtenues à l'avance, la matrice **TF** exige seulement de vérifier les *k* événements
- économique : parce que nous n'avons pas besoin de stocker les nombreuses grandes matrices. Par contre, nous stockons juste la matrice **EA** et le vecteur **WE**
- précis : cette nouvelle proposition est comparée à la méthode d'eigenvector. Basées sur ces comparaisons, les poids calculés par les deux méthodes pourraient être très proches

R.2.3. Une étude simulaire sur la sélection du réseau basée sur le modèle MADM

La selection du réseau a déjà étudiée depuis plusieurs années, et differents modèles mathématiques sont utilisés pour ce sujet. Nous choisissons le modèle MADM, parce que la selection du réseau est vraiment un problème où nous considérons multiple attributes pour prendre une decision ; La procedure de la selection du réseau dois completer dans quelques millisecondes, le modèle MADM correspond cette exigence. Par contre, les autres modèles, comme NP-complet, prennent trop long temps. Ainsi, le modèle MADM est choisi comme le noyau de notre schéma.

Nous établissons un simulateur configuré ci-dessous :

Attributs : multiples attributs seraient utilisés ensemble, par exemple, le coût monétaire, le débit, la consumation d'energie, le niveau de sécurité, le trafic, la force de signal, le taux d'erreur, la gigue, le rayon cellule, la pourcentage de couverture, etc.

Exigences : deux caractéristiques de terminal (la condition d'energie et la vitesse) et quatre niveaux de QoS (vocal, vidéo, interactif et background).

Réseaux : WPAN, WLAN, WMAN et WWAN.

Pondération : la méthode eigenvector est utilisée pour la pondération basée sur une 9 * 9 matrice de pair-à-pair comparaison.

Ajustement : normalisation et fonction d'utilité sigmoïde.

Ordre de réseaux : cinq algorithms MADM

Nous avons simulé les scenarios ci-dessous :

- l'effet d'exigences de terminal ;
- les coûts totals de différents algorithms MADM
- les résultas de la sélection de différents algorithms MADM
- allocation du trafic et répartition de charge

Nous trouvons qu'il y a encore beaucoup de problèmes irrésolus :

- eigenvector n'est pas une bonne méthode pour la pondération pendant la sélection du réseau ;
- les paramètres de mobilité ne sont pas utilisés correctement dans les schémas traditionnels ;
- la décision de handover vertical n'est pas simple quand il y a plusieurs réseaux ;
- c'est possible d'ulitiser le trafic comme un paramètre dans le modèle MADM, mais le poids de ce paramètre peut faiblir les importances des autres paramètres.

R.2.4. Répartition de charge vs. sélection du réseau

Comme montré dans la section R.1.3., tous les attributs de réseau doivent être combinés. Le « trafic » est aussi un attribut de réseau, mais sa utilisation est totalement différente. Tout d'abord, cet attribut est dynamique. Deuxièmement, il a une caractéristique spéciale comme suit : prenons par exemple un cas avec deux réseau, les trafics des réseaux sont 0,1% et 10%. Après la normalisation, les deux valeurs sont près de 1% et 100%. Basées sur ces valeurs normalisées, nous sommes quasiment certaine qu'un nouveau trafic soit mis dans le premier réseau. Cependant, le trafic réel du deuxieme réseau est seulement 10%, nous devons donc examiner des autres attributs pour élaborer cette décision. Sinon, l'utilisation de « trafic » sera immodérée, qui minimise l'importance des autres attributs.

Ce problème est à cause de deux raisons, l'une est la normalisation, l'autre est appelée « équilibre immodéré » dans cette thèse. Nous voullons utiliser cet attribut de la même façon que les autres, alors nous avons décidé de changer sa fonction de l'utilité. Comme indiqué ci-dessous, nous utilisons une fonction sigmoïde spéciale

$$U(x) = x^{\eta} / (1 + x^{\eta}),\ (\eta \geq 2),$$

où x est la valeur de la circulation et η est une constante de l'expérience.

R.2.5. Décision HOV basé sur la prévision

Comme nous avons expliqué dans la section R.2.1.1., il est nécessaire de considérer les coûts du HOV, mais ce n'est pas suffisant parce que ces coûts HOV sont calculés sur la moyenne. Lorsqu'un réseau est sélectionné grâce à son petit coût du HOV, il y a encore une probabilité que le terminal sorte de la couverture de ce réseau rapidement. Si cela peut être prédit, ce terminal ne devrait pas transiter à ce réseau. C'est exactment pour cette raison, nous avons besoin de la décision HOV après la sélection du réseau.

Comme le mouvement d'un terminal est irrégulière, il est assez difficile de prédire leur mouvement pendant une longue période. Ainsi, la prévision la plus croyable est dans une courte période. Par conséquent, une durée limitée T pour la prédiction devrait être envisagée. Voyons quelques exemples :

- un meilleur réseau est seulement un peu mieux que l'actuel. Ce n'est pas suffisant pour rejeter le handover à ce réseau. Nous devons considérer si ce petit avantage dure une longue période ou pas.

- un meilleur réseau peut disparaître dans une durée courte. Ce n'est pas suffisant pour rejeter le handover à ce réseau. Nous devons examiner s'il est beaucoup mieux que l'actuel ou pas.
- un réseau beaucoup mieux peut être disponible dans une courte durée. Par exemple, nous utilisons *C* et nous trouvons un réseau *B* qui est meilleur que *C*. Mais, nous prédisons qu'un réseau *A* beaucoup mieux que *B* sera disponible dans une courte durée. Si nous faisons handover à *B* en ce moment, nous prenons le risque de faire le deuxième handover à *A* dans une courte durée.
- en outre, si nous prévoyons pour *N* réseaux, il sera plus compliqué.

R.2.6. Coût de mobilité et sélection du MAP

Une autre contribution de cette thèse est un nouveau schéma pour la sélection du MAP pour les réseaux basés sur le protocole « mobile IPv6 hiérarchique » et l'evaluation du coût de la signalisation pour la mobilité. Un MAP est une agence de mobilité, qui est appelé « Mobility Anchor Point » en anglais.

Dans le schéma de la sélection du réseau basée sur la mobilité, le coût moyen du HOH et HOV sont utilisés comme les attributs importants, mais ils sont en réalité très complexe à évaluer. L'évaluation de ces coûts devraient être au moins liés aux choses suivantes :

- structure de réseau, par exemple, les handovers intra-MSC (mobile switching center en anglais) et les handovers inter-MSC
- manière de couplage, par exemple, la couplage fort et la couplage lâche
- gestion de la localisation, par exemple, MIPv6, l'optimisation du routage (OR), HMIPv6, et PMIPv6
- schéma de handover, par exemple, « soft handover » et « fast handover »

Avec HMIPv6, un LCoA et un RCoA sont obtenus par le terminal. Le LCoA est enregistré au MAP sélectionné, et le RCoA est enregistré au agent famille et aussi noeud correspondant pour la mode OR.

Il existe deux types de signalisation : la signalisation locale entre le terminal et sa MAP et la signalisation mondiale à l'extérieur de la couverture du MAP. D'ailleurs, il peut être nécessaire d'évaluer le coût moyen de « tunneling » des paquets, parce que l'en-tête de « tunneling » coûte également un débit supplémentaire. Par conséquent, si je veux comparer un réseau HMIPv6 avec d'autres réseaux, il est nécessaire d'évaluer le coût de la signalisation de HMIPv6 et cette évaluation peut être faire comme ci-dessous :

Tous d'abord, nous savons qu'il existe différents types de mobilité, principalement

- presque statique
- mouvement tout droit avec une certaine vitesse
- mouvement aléatoire (direction aléatoire)
- mouvement principalement dans une région relativement petite

Le dernier type est intéressant, parce que les clients de terminaux mobiles ont généralement l'habitude de visiter et de séjourner dans les mêmes lieux, telles que la maison, le bureau, le campus, la centre d'affaires, le café, l'aéroport, le KFC, etc. Pour évaluer le coût de mobilité, un nouveau paramètre est utilisé, nous l'appelons « *le taux de localisation* », ce qui est défini comme le nombre moyen des endroits où un client visite en temps unité. Quand un client se déplace beaucoup mais seulement dans une petite région ou entre plusieurs endroits fixés, la vitesse peut être grande mais le taux de localisation peut être petit.

Basée sur cette analyse, le taux de la transition d'entre MAPs est calculée en fonction du taux de localisation. Le coût total de signalisation est composée de 3 parties : le coût local, le coût mondial, et le coût de « tunneling ». Un avantage de ce schéma est le groupe de seuils que nous avons finalement obtenus ci-dessous, ce qui simplifie largement la sélection du meilleur MAP :

$$Z = \frac{U_{AR}(L_{BU} + L_{BACK}) + U_S P L_{PT}}{U_{Location}},$$

où L_{BU}, L_{BACK} et L_{PT} représentent le coût de « binding update », le coût de « binding acknowledgement » et le coût de « packet tunneling ». U_{AR}, U_S et $U_{Location}$ représentent le taux de la transition entre des routeurs accès, le taux de l'arrive du service et le taux de transition entre des locations. La règle est « *plus Z est élevé, moins la couche de MAP devrait être sélectionnée* ».

R.2.7. Stratégie intégrée pour la sélection du réseau

La figure R-8 montre notre stratégie intégrée pour la sélection du réseau. Après la procédure est déclenchée, la première étape est de vérifier les informations disponibles. Certaines informations seront utilisées pour la pondération, et certains pour l'ajustement. La méthode de pondération proposée sera utilisée pour calculer les poids subjectifs, et la fonction de l'utilité suggérée sera utilisée pour ajuster les valeurs de « trafic ». Ensuite, le schéma de la sélection du réseau basé sur les permutations sera utilisé pour trouver le meilleur réseau et le schéma de la sélection de la meilleure permutation sera utilisé pour trouver la meilleure permutation. Après, le compromis HOV sera utilisé pour décider s'il vaut le coût de faire handover au meilleur réseau ou pas. À la fin, le HOV sera réalisé si nécessaire.

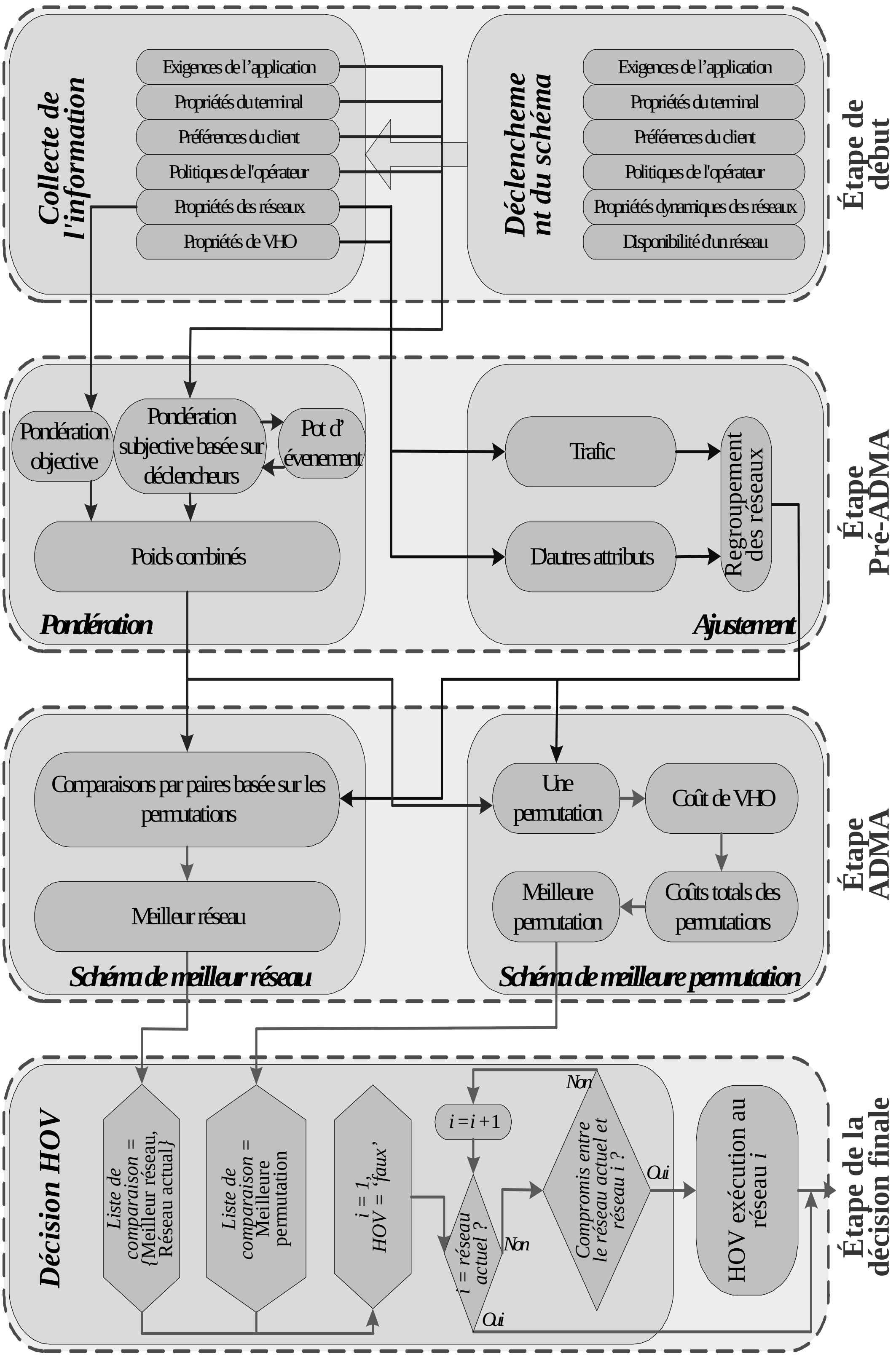


Figure R- 8 Solution intégrée pour la sélection du réseau.

R.2.8. Sélection du réseau pour un réseau mobile

Un réseau mobile (REMO) est un réseau consistant un nombre de terminaux déplaçant ensemble, par exemple des terminaux dans un bus. La sélection du réseau pour un REMO n'a pas été étudiée dans la littérature, mais il y a vraiment des problèmes.

Tout d'abord, la façon de recueillir toutes les informations doivent être étudiée. Pour un REMO, cette procédure est différente que celle d'un terminal à cause de la transparence de la mobilité aux terminaux intérieurs. Il y a deux équipements possibles pour faire la sélection : sois par le routeur mobile, sois par des terminaux intérieurs. Toutefois, les informations du client et des services sont sur des terminaux, mais les informations de la mobilité et du « multihoming » sont traitées par le routeur mobile et ils sont généralement transparentes pour des terminaux intérieurs.

Nous pensons qu'il est préférable d'utiliser le routeur mobile, ainsi le schéma ne brise pas l'architecture des protocoles de la mobilité et du « multihoming ». Avec cette façon, toutes les informations de l'extérieur seront réunis par le routeur mobile. Toutes les informations du client ou des services seront transféré au routeur mobile aussi. Il y a principalement deux informations intérieurs à transférer :

- les préférences du client, qui peut être indiquée par l'adresse de source dans l'en-tête d'IP
- l'exigence des services, qui peut être envoyée au routeur mobile aussi dans l'en-tête d'IP

Avec cette méthode, il y a encore un problème évident. À l'origine, lorsque nous parlons de la sélection du réseau, cette sélection est pour un service ou un terminal. Mais avec cette méthode, il faut toujours regarder les en-tête d'IP pour envoyer les paquets, donc la sélection du réseau est changé à une procédure du traitement sur les paquets.

R.2.9. Un simulateur en Matlab

Nous établissons un simulateur en Matlab pour évaluer la performance de différents schemas. Ce simulateur peut simuler plein de cas et peut montrer plein de résultas de différents simulations.

Les parameters que ce simulateur peut configurer comportent :

- *Modules*

  - L'algorithmes MADM, par exemple, SAW, MEW, GRA, TOPSIS, ELECTRE, etc.
  - Les methods pondérations, par exemple, Equality, Entropy, AHP, TRUST, etc.
  - La normalisation
  - La decision de handover vertical, etc.

- *Paramètres*

  - Les attributs de réseaux ;
  - Les préferences d'utilisateur ;
  - Les exigences d'application ;
  - Les caractèristiques de terminal, etc.

Les simulations que ce simulateur peut faire comportent :

- Réseau vs. Attribut
- Coût total vs. Attribut
- L'effets d'une série d'événement, etc.

R.3. Contributions

Dans cette thèse, nous étudions la sélection du meilleur réseau. Les contributions sont énumérées ci-dessous :

- présentation de l'état de l'art sur la sélection du réseau
- établissement d'un simulateur utilisant Matlab
- simulation du modèle ADMA pour la sélection du réseau
- analyse de l'utilisation des attributs de mobilité pour la sélection du réseau
- proposition d'un schéma pour trouver la meilleure permutation de réseaux
- modélisation de la mobilité dans les RFHs
- proposition d'une nouvelle méthode de pondération fondée sur des déclencheurs

- evaluation du coût de signalisation de la mobilité dans les réseaux HMIPv6
- proposition d'un schéma de la sélection du MAP basé sur le taux de localisation
- analyses sur des autres problèmes :
    - usage de « trafic »
    - compromis HOV basé sur la prévision
    - intégration de toutes les propositions ci-dessus
    - sélection du réseau pour un REMO

Les problèmes qui ressentent une étude plus approfondie abordant au chapitre 7, notamment :

- une proposition générique pour compromis HOV sur les prévisions
- l'équilibrage de charge de trafic vs. ABC
- une evaluation analytique pour la sélection de MAP basée sur les nouveaux modèles de gestion de localisation
- la sélection de réseau pour REMO et pour handover en groupe
- la bordure équivalent dans les simulations Monte Carlo pour les réseaux WLAN

Table of Contents

Acknowledgement i
Abstract ii
Résumé iv
R.1. Introduction iv
R.1.1. Réseaux sans fil hétérogènes et la sélection du réseau iv
R.1.2. Exemples de la sélection du réseaux v
R.1.3. Contexte mathématique pour ce sujet vii
R.1.4. Problèmes à résoudre viii
R.2. Propositions ix
R.2.1. Sélection du réseau basée sur la mobilité ix
R.2.2. Pondération fondée sur des déclencheurs xvi
R.2.3. Une étude simulaire sur la sélection du réseau basée sur le modèle MADM xix
R.2.4. Répartition de charge vs. sélection du réseau xxi
R.2.5. Décision HOV basé sur la prévision xxi
R.2.6. Coût de mobilité et sélection du MAP xxii
R.2.7. Stratégie intégrée pour la sélection du réseau xxiv
R.2.8. Sélection du réseau pour un réseau mobile xxvi
R.2.9. Un simulateur en Matlab xxvi
R.3. Contributions xxvii
Table of Contents xxix
List of Figures xxxii
List of Tables xxxv
Glossary xxxvi
1. INTRODUCTION 1
1.1. Background 1
1.2. Problem statement 4
1.3. Contributions 7
1.4. Organization of this dissertation 11
2. STATE OF THE ARTS 13
2.1. Background 13
2.1.1. Evolutionary of HWNs 13

2.1.2. IP mobility and multihoming....18
2.1.3. Always best connected....24
2.2. Information gathering....28
2.2.1. Required information....28
2.2.2. IEEE 802.21 standard....30
2.2.3. IEEE P1900.4....32
2.2.4. Context-aware information gathering....33
2.2.5. Pilot channel....35
2.3. Network ranking schemes....36
2.3.1. Cost/Utility function....36
2.3.2. MADM....41
2.3.3. Fuzzy logic....47
2.3.4. Markovian model....51
2.3.5. Game theory....53
2.3.6. NP hard....55
2.3.7. Power saving....57
2.3.8. User/operator combined strategy....58
2.3.9. Integrated solution....59
3. MADM-BASED NETWORK SELECTION....61
3.1. Configuration of simulator....61
3.2. Effects of terminal-side requirements....63
3.3. Coefficients of various MADM algorithms....64
3.4. Selection results of various MADM algorithms....66
3.5. Traffic load assignment for MADM-based network selection....67
3.6. Important observations and issues....67
4. MOBILITY-BASED NETWORK SELECTION....69
4.1. Introduction....69
4.2. Mobility-based scheme considering two groups of networks....71
4.2.1. Mobility modeling....71
4.2.2. Evaluation of HHO and VHO costs....78
4.2.3. Performance evaluations....79
4.2.4. Coverage of small-scale networks vs. network selection....81
4.3. Mobility-based scheme considering *N* groups of networks....83
4.3.1. Evaluating total costs of permutations in a generic scenario....84

4.3.2. Methods to get the best permutation..88
4.3.3. Performance evaluations 90
5. TRUST: A NOVEL WEIGHTING METHOD FOR NETWORK SELECTION...95
5.1. Modeling the weighting issue for network selection 95
5.2. Objective weighting methods 96
5.3. Subjective weighting methods...98
5.4. Inappropriateness of the eigenvector method............................99
5.5. TRUST 101
5.6. Performance comparisons 104
6. MOBILITY SIGNALING COST EVALUATION AND MAP SELECTION...108
6.1. Mobility signaling cost evaluation...108
6.1.1. Background on HMIPv6 signaling 108
6.1.2. Defining a new parameter 'location rate'...............................109
6.1.3. Mobility signaling cost..110
6.2. MAP selection in HMIPv6 networks.....................................112
6.2.1. Related work...112
6.2.2. Location history based MAP selection scheme.......................115
6.2.3. Performance evaluation...117
7. OTHER ISSUES ON NETWORK SELECTION.........................125
7.1. Traffic load assignment during network selection...................125
7.2. Vertical handover decision scheme 127
7.3. An integrated strategy of network selection............................131
7.4. An analysis of network selection for NEMO 132
8. CONCLUSIONS...134
APPENDIX 137
A. Network selection simulator..137
REFERENCES 145

List of Figures

Figure R- 1 Pourquoi nous avons besoin de la sélection du réseau. v

Figure R- 2 Les exemples de la sélection du réseau. vi

Figure R- 3 Une solution intégrée pour la sélection du réseau. vii

Figure R- 4 Déploiement des PAs. x

Figure R- 5 Une méthode combinée... xiv

Figure R- 6 Les transitions causant des HOVs dans un exemple de trois réseaux. xv

Figure R- 7 La hiérarchie des événement. xviii

Figure R- 8 Solution intégrée pour la sélection du réseau. xxv

Figure 2- 1 Relationship between various wireless networks.................... 13

Figure 2- 2 The evolutionary of WWAN technologies [LP08]. 14

Figure 2- 3 The evolution of WLAN standards. 15

Figure 2- 4 Development and key features of WiMax standards [Erxn]. ... 16

Figure 2- 5 Interworking of HWNs. 18

Figure 2- 6 Host identity protocol stack. 21

Figure 2- 7 Mapping with changing locators in shim6.............................. 21

Figure 2- 8 Searching Binding Entries with MONAMI6. 23

Figure 2- 9 Sub-optimal route and HAHA tunnel with Global HAHA protocol. 24

Figure 2- 10 A series of scenarios for ABC... 26

Figure 2- 11 Technical components of an ABC solution [GE03].............. 26

Figure 2- 12 IEEE 802.21 architecture. 30

Figure 2- 13 EMIH architecture. 31

Figure 2- 14 IEEE P1900.4 functional architecture. 33

Figure 2- 15 An architecture of context-aware network selection. 35

Figure 2- 16 A detailed architecture of context-aware network selection... 35

Figure 2- 17 CPC-based network selection scheme. 36

Figure 2- 18 A utility-based network selection strategy focusing on transfer time cost. 40

Figure 2- 19 Compensatory and non-compensatory integrated network selection. 43

Figure 2- 20 Combined procedure of AHP and GRA. 45

Figure 2- 21 An example of an AHP hierarchy of criteria. 46
Figure 2- 22 Membership function of ‘signal strength’. 47
Figure 2- 23 The procedure of fuzzy logic based VHO decision. 48
Figure 2- 24 Fuzzy MCDM network selection scheme. 49
Figure 2- 25 Fuzzy MADM with imprecise information. 49
Figure 2- 26 Network specific and user specific combined network selection. 50
Figure 2- 27 Network selection combining fuzzy logic and neural network. 51
Figure 2- 28 Terminal-controlled network selection scheme. 58
Figure 2- 29 A synthetic strategy of network selection..............................60

Figure 3- 1 Total costs for different terminal properties and applications (from left to right: WPAN, WLAN, WMAN and WWAN)........................63
Figure 3- 2 Coefficients’ changes with respect to certain criterion’s weight. 65
Figure 3- 3 Selection results of MADM algorithms. 66
Figure 3- 4 Load balancing feature. 67

Figure 4- 1 Utilities of different networks through sigmoidal function. 72
Figure 4- 2 Distribution of hotspots. 72
Figure 4- 3 Experimental result of α_2. 73
Figure 4- 4 Monte Carlo simulation results. 73
Figure 4- 5 Transition of an MT in an area with K hotspots. 75
Figure 4- 6 Transition of an MT within 2 states. 75
Figure 4- 7 Monte Carlo simulation results. 77
Figure 4- 8 Coefficients of various networks vs. Weights of average handover cost. 82
Figure 4- 9 Coefficients of various networks vs. Coverage percentage.......82
Figure 4- 10 Examples of f(V) and W(V). 83
Figure 4- 11 Consumer increment rate (CIR) of WLAN. 83
Figure 4- 12 Transitions in a 3-group scenario. 86
Figure 4- 13 Procedure of Enhanced Besper. 90
Figure 4- 14 Vertical handover rate comparison. 92
Figure 4- 15 Total cost comparison..93

Figure 4- 16 Trigger rate comparison... 93
Figure 4- 17 Time cost vs. Number of groups of networks........................ 93
Figure 4- 18 Time cost vs. Weight of handover cost (with 4 groups of networks). 94

Figure 5- 1 Hierarchy of trigger events. 103
Figure 5- 2 Comparison results between TRUST and Eigenvector method. 106

Figure 6- 1 A three-layer MAP model for HMIPv6 networks...................109
Figure 6- 2 Cost comparison with SMR = 0.2. 121
Figure 6- 3 Handover delay comparison in unit movement with SMR = 0.2. 122
Figure 6- 4 Cost comparison with Residence Factor = 0.5. 123
Figure 6- 5 Handover delay comparison with Residence Factor = 0.5. 124

Figure 7- 1 Utility function for traffic load. 127
Figure 7- 2 VHO tradeoff scenario 4..129
Figure 7- 3 VHO tradeoff scenario 5..130
Figure 7- 4 VHO tradeoff scenario 6..130
Figure 7- 5 An integrated strategy of network selection. 132
Figure 7- 6 MR-side network selection. 133

Figure A- 1 Main-board of our simulator. 137
Figure A- 2 Framework window...138
Figure A- 3 Network properties window..139
Figure A- 4 Customer preferences window. 140
Figure A- 5 Application QoS requirements window. 140
Figure A- 6 Terminal properties window. 141
Figure A- 7 Network vs. factor simulation. 141
Figure A- 8 Simulation results of Simulation_NF_gui. 142
Figure A- 9 Coefficient vs. factor simulation. 142
Figure A- 10 Simulation results of Simulation_TF_gui............................143
Figure A- 11 Event series simulation. 143

List of Tables

Tableau R- 1 Le taux et le coût du HOV et HOH..................................... xii

Table 2- 1 Details of a series of scenarios for ABC...................................26

Table 2- 2 Examples of several common utility functions..........................38

Table 2- 3 Examples of normalization operations. 42

Table 2- 4 The first round of an example of a game between networks......55

Table 3- 1 Weights of various scenarios (%)...62

Table 3- 2 High weights corresponding to subjective requirements............63

Table 4- 1 Simulation parameters. 76

Table 4- 2 Handover rates and costs. 78

Table 4- 3 Symbols for a generic scenario. 86

Table 5- 1 Relationship between trigger events and subjective weights....102

Table 5- 2 Comparisons between TRUST and Eigenvector method. 107

Table 6- 1 Symbols in the proposal...111

Table 6- 2 Costs of RO and Non-RO modes for one movement...............112

Table 6- 3 Main simulation parameters..120

Glossary

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
AAA	Authentication, Authorization and Accounting
ABC	Always Best Connected
AHP	Analytic Hierarchy Process
B3G	Beyond Third Generation
Besnet	Best network scheme
Besper	Best Permutation scheme
CDMA	Code Division Multiple Access
CS	Consumer Surplus
DM	Decision Maker
DSL	Digital Subscriber Line
EDGE	Enhanced Data rates for GSM Evolution
ELECTRE	ELimination and Choice Expressing Reality
EMIH	Enhanced Media Independent Handover
ETSI	European Telecommunications Standard Institute
FDD	Frequency-Division Duplexing
FMIPv6	Fast Handover for Mobile IPv6
GPRS	General Packet Radio Service
GPS	Global Positioning System
GRA	Grey Relational Analysis
GSM	Global System for Mobile communications
HC	Handover Cost
HHO	Horizontal Handover
HIP	Host Identity Protocol
HMIPv6	Hierarchical Mobile Internet Protocol version 6
HSDPA	High-Speed Downlink Packet Access
HSN	HotSpot Network

HSUPA	High-Speed Uplink Packet Access
HWN	Heterogeneous Wireless Network
HyperLAN	High-Performance Radio Local Area Network
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
ISM	Industrial Scientific Medical
ISP	Internet Service Provider
MADM	Multiple Attributes Decision Making
MAP	Mobility Anchor Point
MCDM	Multiple Criteria Decision Making
MEW	Multiplicative Exponent Weighting
MIH	Media Independent Handover
MIPv6	Mobile Internet Protocol version 6
MNN	Mobile Network Node
MMKP	Multiple choice Multiple dimension Knapsack Problem
MMT	Multi-homed Mobile Terminal
MN	Mobile Node
MODM	Multiple Objectives Decision Making
MONAMI6	MObile Nodes And Multiple Interfaces in IPv6
MR	Mobile Router
MT	Mobility Terminal
NAT	Network Address Translation
NEMO	Network Mobility
NP hard	Non-deterministic Polynomial-time hard
PDC	Personal Digital Cellular
PDF	Probability Density Function
PMIPv6	Proxy Mobile Internet Protocol version 6
QoS	Quality-of-Service
RAN	Radio Access Network
RSS	Received Signal Strength
RWMM	Random Walking Mobility Model
SAW	Simple Additive Weighting

SIP	Session Initiation Protocol
SMR	Session Mobility Ratio
SMT	Single-homed Mobile Terminal
TDD	Time-Division Duplexing
TDMA	Time Division Multiple Access
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TRUST	TRigger-based aUtomatic Subjective Weighting
UBN	UBiquitous Network
UMTS	Universal Mobile Telecommunications System
UWB	Ultra-Wide Band
VHO	Vertical Handover
WG	Working Group
WiMax	World-wide Interoperability for Microwave Access
WiBro	Wireless Broadband
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WMC	Weighted Markov Chain
WPAN	Wireless Personal Area Network
WWAN	Wireless Wide Area Network

1. INTRODUCTION

This chapter provides an overview of this dissertation, including background of the study, statement of the problems we are going to tackle and a summary of our contributions.

1.1. Background

The recent development of wireless technologies has totally revolutionized the world of communications. Multiple technologies are evolving simultaneously towards providing customers high-quality services of broadband access and seamless mobility. On the one hand, wireless wide area networks (WWANs) evolve from GSM to UMTS and beyond 3G [LP08], providing wide coverage and good mobility support. On the other hand, a series of standards of wireless local area networks (WLANs) [MG02], including IEEE 802.11b, IEEE 802.11g, IEEE 802.11a, IEEE 802.11n, etc., have been proposed for local-area high bandwidth wireless access. To complement these networks, wireless personal area networks (WPANs) [LA07], e.g. Bluetooth and zigbee, and wireless metropolitan area networks (WMANs), e.g. WiMax [PD06], are developed for short-range and metropolitan coverage, respectively. All the above networks could be deployed with coverage overlapping, hence form a hybrid environment for wireless access, which is usually called heterogeneous wireless networks (HWNs).

To access the Internet through HWNs, current terminals (e.g. laptops, cellphones, etc.) are usually installed with multiple wireless access network interfaces. In the scope of this dissertation, two types of terminals with multiple interfaces are considered. One type of terminals widely used nowadays is those with multiple interfaces but no functionality to support IP mobility or multihoming, called multi-mode terminals. The other type of terminals is with IP mobility and multihoming functionalities, called multi-homed mobile terminals (MTs). Multihoming means that one terminal or network has multiple IP connections to one or multiple service providers simultaneously [GH05] [KH06]. A multi-homed MTs has multiple addresses through one or multiple interfaces, while MTs with multiple interfaces are quite common in the context of HWNs. Multi-homed MTs could use multiple interfaces for load sharing for the same session, and could support session continuity with low (or no) packet loss during mobility or link break. By contrast, multi-mode MTs could only select and use one interface for certain session at a time.

Meanwhile, internet protocol (IP) is considered today as the basic transport technique for HWNs. Since IPv4 address pool might be exhausted in a couple of years [HT05] [v4ad], the wide deployment of IPv6 becomes inevitable, so we focus on the latter in this dissertation, although our studies and proposals may be also suitable for IPv4 networks. To support location management and session continuity among HWNs, mobile IPv6 (MIPv6) has been specified a few years ago [JD04]. However, this solution does not perform well and has many issues to solve. Therefore, route optimization (RO) and fast handover for MIPv6 (FMIPv6) [KR05] were proposed to improve its performance, e.g. decreasing tunneling cost and packet loss, etc. Furthermore, hierarchical MIPv6 (HMIPv6) [SH05] was specified to decrease signaling cost of location management when the mobile node (MN) is far away from its home network (HN).

Unfortunately, even with the above extensions of MIPv6, there is still one important function that could not be supported. That is, now that multiple networks are ready for usage, a customer might want to share the traffic load of one session among more than one network. To do this, IETF MONAMI6 working group specified an extension to MIPv6, in order to register a MN's multiple care-of-addresses (CoAs) to its home agent (HA) [WR07]. This solution does not only support load sharing, but also enhance MIPv6 for seamless mobility. Briefly speaking, packet loss could be extremely decreased when the two interfaces (before and after handover) are available for transmission simultaneously [MH03] [LH07]. Furthermore, this solution could be easily extended to HMIPv6, which could register a MN's multiple on-link care-of-addresses (LCoAs) to its mobility anchor point (MAP).

Seen from another angle, the reason that MIPv6 and its extensions could support mobility and multihoming is their feature of splitting IP address's identification function from its location function. Detailed speaking, IP address is traditionally used both to identify the MN and to indicate its location. When a MN attaches itself to another network and changes its IP address, this breaks all the on-going sessions which are identified by its old IP address. Therefore, MIPv6 and its extensions use home address (HoA) for the MN's identification and CoAs for its location. Hence, when the MN moves among HWNs and changes its CoAs, its HoA never changes, which continues its on-going sessions. Similarly, some IETF working groups work specifically on identification/location separation of IPv6 address, e.g. host identity protocol (HIP) [MR07] and level 3 shim for IPv6 (shim6) [NE07]. By inserting a sub-layer within the IP layer to split identification and location functions, these solutions are considered easily to have mobility and multihoming features. However, compared with MIPv6, the requirement of new protocol installation on correspondent nodes (CNs) might affect their deployment.

Up to now, there are billions of MNs used on this planet without having any of the above protocols supporting mobility or multihoming. It is a big challenge but could be very interesting to provide them mobility and multihoming features without doing any change on themselves. Proxy MIPv6 (PMIPv6) [GS08] is a network-based localized mobility management protocol, which requires no participation of the MN to mobility related procedures. Similar to HMIPv6, the above mentioned MONAMI6 working groups' solution could also be extended to PMIPv6 to register a MN's multiple proxy care-of addresses (proxy-CoAs) to its local mobility anchor (LMA). Seen from the angle of identification/location separation, there is another network-based solution, called locator/ID separation protocol (LISP) [FD09].

To sum up, many solutions have been proposed to support mobility and multihoming for the HWN environment in recent few years. Based on where these new protocols should be deployed, they could be classified into three groups: 1) on both MNs and CNs, e.g. HIP and shim6; 2) on only MNs not CNs, e.g. MIPv6, HMIPv6, FMIPv6 and MONAMI6; 3) on neither MNs nor CNs, e.g. PMIPv6 and LISP. Any group of these protocols could work separately, but they might also be deployed cooperatively to support mobility and multihoming in the near future.

However, with all the above solutions, there are still obvious questions that has no been well answered. When a MN obtains multiple IP addresses from its wireless access network interfaces, the question is 'which IP address should the MN use for a certain application'. Similarly, when a MN registers multiple IP addresses (CoAs in MIPv6, LCoAs in HMIPv6, Proxy-CoAs in PMIPv6 or locators in HIP/shim6), the question is 'which one of its multiple IP addresses should the HA (or MAP, LMA, etc.) use to transfer packets to the MN through the tunnel, or which one should a CN use to initiate a session with it'. To answer these questions, a new and important concept should be introduced, well known as 'Always Best Connected (ABC)' [CY07] [GV05].

ABC, as an important concept to provide high-quality services, became a popular research topic in recent few years. ABC is to select and always connect to the most appropriate network when multiple networks are available. It contains many necessary components [GE03], such as network discovery, network selection, handover execution, authentication, authorization and accounting (AAA), mobility management, profile handling, content adaptation, etc., in which network selection is a key component and will be precisely discussed in this dissertation.

Before talking about problems and contributions in later sections, there are three declarations to make clear. The first one is about the layer in the OSI seven-layer model that mobility and multihoming would belong. We can see that all the protocols mentioned above could be classified into one single group, i.e. IP layer solutions. By contrast, there are also transport layer solutions [AM05], application layer solutions and link layer solutions. This dissertation will not try to give a conclusion on which layer should mobility and multihoming belong. In this dissertation, I consider only IP layer solutions as references on network selection issues, but most of the proposals are generic and are not limited to mobility and multihoming protocols.

The second declaration is about the type of mobility that the MN is concerned with. Mobility can be classified into different types, e.g. personal mobility, terminal mobility, session mobility and service mobility [PI03]. In this dissertation, I focus on terminal mobility among HWNs, but the proposals could be easily extended to more generic scenarios. For example, when a session is moving from a customer's mobile phone to its laptop (known as session mobility), this session movement event could trigger the network selection procedure and a new network might be selected because of the change of terminal properties. All we need to do for extending my proposals is to identify and use more terminal properties (e.g. screen size) in the decision procedure.

The third declaration is about the access technologies we considered. Although we mainly consider UMTS, WiMax, IEEE 802.11g and Bluetooth 2.0 in our analysis and simulations, the proposals in this dissertation are not limited to certain group of networks. In the near future, if some new access technology has been developed, the proposals can still be used to select the best network. Moreover, only wireless technologies are considered in this dissertation, but the proposals are also suitable to take wired access technologies, e.g. asymmetric digital subscriber line (ADSL), into consideration. The only thing that needs to change is the matrix recording network properties.

1.2. Problem statement

Network selection has been studied a lot in recent few years, but there are still many problems that have not been solved. Some problems result in sub-optimal network selection results; while some problems are serious enough to make a network selection scheme unable to work. These problems are described as follows:

- A network selection scheme usually considers multiple groups of factors simultaneously, including network attributes, operator policies, terminal properties, customer preferences, application QoS levels, VHO properties, etc. Some factors are mobility-related, such as cell radius, coverage percentage, terminal velocity, HHO and VHO properties, etc. These factors are gathered (e.g. by an MIH information server [802.21]) and used to represent MNs' mobility features and networks' mobility support capabilities, and are important for network selection. For example, according to these factors, high speed MTs should not select a network with small cell radius; otherwise, live applications will be severely disturbed by frequent handovers.

However, these factors are not well used in related works. To the best of our knowledge, before our work, nobody realized and raised that the usage of mobility-related factors is quite complicated than other factors. Mobility-related factors are difficult in usage because VHO properties (e.g. signaling cost) are corresponding to the ordering of networks. Hence, when a scheme is trying to find the solution with minimum VHO signaling cost, it corresponds to an ordering of networks (called the 'best permutation' in this dissertation), not a network. Moreover, the number of permutations is the factorial of the number of networks, so it may require too much time to calculate total costs and to find the best permutation. For further information of this study, please refer to chapter 4.

- Besides attributes, another important part in a network selection scheme is the weights of these attributes. Weights decide the relative importance of different attributes, while this relative importance usually decides the best network. Therefore, correct weights are necessary, while wrong weights usually lead to wrong decision. However, weighting in network selection is quite complicated. Various factors should be considered, as mentioned in the above paragraphs. And, there are both objective and subjective weights.

  Entropy method can be used to obtain the objective weights which denote the relative differences of candidate networks respecting to various attributes. However, it is not enough to use only objective weights to represent these attributes relative importance because most factors for deciding the weights are subjective information which requires subjective weighting methods.

  AHP was usually considered for calculating subjective weights in recent research papers. However, this method has an obvious problem in the network selection procedure. As we know, pair-wise comparison matrices in this method are given by the decision maker (DM) based on his subjective feelings on the attributes, and the DM is usually human beings in most decision making processes. However, for the network selection issue, an automatic method is required because customers usually do not have the basic knowledge to construct the pair-wise comparison matrix. Moreover, the matrix changes in different situations (e.g. different applications and terminal properties), so customers do not want to be involved in the complicated pair-wise comparison process for each situation, even though they know how to do it. To sum up, when designing a network selection scheme, we should not suppose the customers to be the DMs who could provide pair-wise comparison matrix to calculate subjective weights. This means the subjective weights should be calculated automatically by the MT (or a network-side entity). However, MTs cannot be DMs, because machines do not have subjective feelings on the attributes. Moreover, this weighting procedure should be fast enough to avoid disturbing VHO which is usually performed after network selection.

In a word, evaluating the subjective weights in the network selection procedure is a tough work. A trigger-based automatic subjective weighting method is going to be proposed in chapter 5, which satisfies all the above requirements.

- Another problem is the consideration of traffic load during network selection because a network with limited resource should not be selected. As we will explain in chapter 2 and 3, game model, knapsack model and bin packing model can balance the load, but they focus too much on the capacities of networks and seriously compromise other network attributes.

  Traffic load can be used as a network attribute in the MADM model, and it will be combined with other attributes based on certain combination rule. However, different from most attributes, traffic load is dynamic, so the information of traffic load should be gathered from multiple networks.

  Furthermore, when all the networks have enough resource, the usage of traffic load in MADM will compromise other attributes. This sometimes obviously affects the selection of the best network. For more information, please refer to the discussion of traffic load assignment in section 6.1.

- After the best network is selected, handover to that network is usually performed. However, I would like to argue that it is not always right to handover to the best network.

  Network ranking schemes do not consider prediction of the future events that the MT might encounter. Detailed speaking, when we find a better network with good QoS support capability, there is still the possibility that the MT is about to leave the coverage of this network. If the MT could predict that the better network's availability lasts only a little time, it should not handover to this network. Furthermore, when a better network is found, the MT may predict that a much better network is going to be available in a little time, so it may be not a good idea to handover to the current better network. By contrast, waiting for the much better network and handing-over directly to it could avoid performing two handovers within a short period.

In a word, prediction is usually not considered in network ranking schemes, but it is necessary to take it into account before handover. In this dissertation, this issue is studied as ‘VHO tradeoff’, which is used before handing-over to the selected best network. For more information, please refer to the analysis in section 6.2.

1.3. Contributions

The goal of the study in this dissertation is to achieve ABC in HWNs. Mobility and multihoming solutions for movement in HWNs have been surveyed. They could be used to support session continuity and location management for terminal mobility and emergencies (e.g. link breaks down), but that is not enough for ABC. The main feature of ABC is to always connect to the best network at any time anywhere, for any customer, any terminal, any session, and no matter what happens (e.g. movement or link broken). In order to achieve this, the premier thing is to select the best network. However, after surveying the network selection schemes raised in recent years, we have to conclude that they are not sufficient to complete this work. Therefore, many problems are identified as shown in section 1.2.

The contributions of this dissertation are listed as follows:

- State of the arts
    - State of the arts is described for providing ABC service to customers. Two important components are studied and summarized: one is mobility and multihoming solutions; the other is information gathering and network selection schemes.
    - A simulation platform is established by Matlab for simulations on network selection. The platform is powerful to simulate vast number of scenarios, and it will be used frequently during this study. Plus, the simulation platform has a friendly GUI which makes it easy to operate.
    - By simulations on carefully designed scenarios (including terminal state change, different applications, different weights, different MADM algorithms, traffic load change, etc.), MADM model is shown suitable for network selection. More important, we define many unsolved problems which greatly affect the performance of network selection schemes. These problems are about the usage of mobility-related factors, the evaluation of subjective weights, traffic load

assignment, VHO tradeoff, etc. They will be studied one by one in this dissertation.

- Mobility-based network selection
    - The usage of mobility-related factors is analyzed. It is shown that they are more complicated than you could imagine beforehand. That is because VHO properties depend on not only features of terminal mobility and network coverage, but also the ordering of networks. Therefore, how to use these factors becomes an issue.
    - To use VHO properties in MADM model, we propose a totally new concept – best permutation selection (in order to be distinguished with traditional best network selection). Permutation here means the ordering of networks, no matter they are currently available or not. Hence, the first available network in the best permutation is selected as the best network. However, the selection of the best permutation is not easy because the number of permutations is the factorial of the number of networks. Hence, the selection procedure takes too much time to find the best permutation, which might disturb real-time applications (time complexity problem).
    - To solve the above problem, one proposal is to divide all the networks into two groups: hotspot network group and ubiquitous network group. This division of networks is proved reasonable by analysis and simulations. Cost function of each group is defined, and a closed-form of the threshold between the two groups is derived.
    - Another proposal to solve the time complexity problem is to do ‘best network selection’ before best permutation selection. The ‘best network selection’ used here is different from traditional best network selection schemes. A permutation-based best network selection scheme is suggested. It takes VHO properties into account, but the selected ‘best network’ is not precisely the best network because the comparison between permutations is not ergodic. However, this imprecise ‘best network selection’ scheme is as fast as traditional best network selection schemes.
    - Besides, a simplification method is proposed to further decrease the time cost of best permutation selection. Since ubiquitous networks cover the whole area (including hotspot networks), it is not necessary to do comparison between

hotspot networks when a ubiquitous network is supposed to have higher priority than them.

- Based on the above analysis of best permutation selection, cost functions of permutations are defined. Not only VHO properties, but also all kinds of network attributes are considered in the defined cost functions. Simulation results prove that best permutation selection out-performs best network selection in many aspects, including total cost, VHO rate, scheme trigger rate, etc.
- The effect on network selection by deploying more hotspots is analyzed. A velocity threshold is derived to decide whether a terminal could select hotspot networks or not, in section 4.1.3. It shows the trend that more customers prefer hotspot networks when deploying more hotspots or evolving from 802.11b to 802.11g.

- Subjective weighting method
  - Weighting methods are surveyed, including both subjective and objective methods. Objective weights denote the relative difference between networks respecting to certain attribute, and subjective weights denote subjective feelings of a decision maker (DM) on these attributes' importance. The problem is on the calculation of subjective weights, which is clearly explained in section 5.4.
  - Based on the analysis on the relationship between scheme trigger events and network selection results, a novel subjective weighting method is proposed, called TRigger-based aUtomatic Subjective weighTing (TRUST). It can efficiently calculate the subjective weights of various network attributes based on both terminal-side and network-side subjective requirements in the network selection issue. Trigger events of the selection procedure are specifically considered by this method, so the subjective weights are calculated based on the effects that these events bring to the selection results.
  - The proposed method is compared in extensive scenarios with eigenvector method, which is widely considered in the study of network selection in recent years. The obtained weights from the two methods are quite close to each other. The difference is that TRUST could calculate these weights automatically and fast, but eigenvector method requires a DM to provide his/her subjective feelings (slow and not automatic).

- Mobility signaling cost evaluation and MAP selection
    - Since mobility is a key factor to distinguish between wireless networks, it becomes an important task to evaluate mobility signaling cost. In previous proposals (i.e. mobility-based network selection schemes), HHO and VHO costs are assumed to be known. In chapter 6, mobility signaling cost in HMIPv6 networks are evaluated, so that it can be used as a key factor in network selection schemes.
    - Moreover, a scheme for selecting the mobility anchor point (MAP) that corresponds to the minimum total cost (including mobility signaling cost and packet tunneling cost) is proposed, and this scheme is compared with many other schemes, showing its advantages in many aspects.
- Traffic load assignment, VHO tradeoff, integrated strategy and network selection for NEMO
    - Traffic load assignment during network selection procedure is analyzed. By using traffic load as a network attribute in MADM model, load balancing feature is obtained. However, if linear of sigmoidal utility function is used for this attribute, the problem is that it might greatly compromise other attributes. Therefore, a specific utility function for the attribute 'traffic load' should be used.
    - VHO tradeoff is analyzed. Many scenarios are considered and tradeoff threshold are found for each scenario.
    - Based on all the study above, an integrated network selection strategy is proposed. MADM is used as the core, while utility theory and fuzzy logic are used to process the values of attributes. Multiple groups of network attributes are considered, including traffic information for load balancing. The strategy uses TRUST as its subjective weighting method, and combines best network selection and best permutation selection. Moreover, VHO tradeoff is considered at the end of this strategy.
    - Finally, network selection for NEMO is analyzed. The difficulty is to gather all the information because network attributes are transparent to customers in a NEMO. Due to this reason, the network selection functional entity should not be on the terminals. Meanwhile, information of the terminal-side (e.g. customer preference and application QoS level) is difficult to be delivered to a network-side functional entity (e.g. ABC server). One suggestion is to gather all the

information to the mobile router of the NEMO, and do network selection there.

1.4. Organization of this dissertation

This chapter provides an overview of this dissertation, including background, problem statement and contributions. The reminder of this dissertation is organized as follows:

Chapter 2 provides the state of the arts, including features of wireless networks, mobility and multihoming, always best connected, information gathering strategies and mathematical models for network selection.

Chapter 3 shows our simulation study on MADM-based network selection schemes. Effects of terminal-side requirements, coefficients and selection results of various MADM algorithms, traffic load assignment feature of MADM-based schemes are simulated. And, some important observations and existing issues are described at the end of this chapter.

Chapter 4 proposes a mobility-based network selection scheme, in two steps: two network clusters and *n* network clusters. In the scenario with two network clusters, we study the usage of mobility-related factors based on sigmoidal utility function, calculate average costs of horizontal and vertical handovers, and derive a threshold between the two clusters. In the scenario with *n* network clusters, we formulize the total handover cost, and propose methods (*Bespers*) to obtain the best permutation of networks rapidly. Simulation results demonstrate the performance of our proposals.

Chapter 5 proposes a trigger-based automatic subjective weighting (TRUST) method, which considers the relationship between trigger events and their effects on subjective weights. Compared with analytical hierarchy process (AHP) subjective weighting, TRUST is a quite efficient method to obtain similar subjective weights. Finally, we suggest combine the subjective weights obtained by TRUST and the objective weights obtained by Entropy method as the hybrid weights used in the network selection procedure.

Chapter 6 analyzes the signaling cost of mobility in a HMIPv6 network. Then, a MAP selection scheme in HMIPv6 networks is proposed, which location history of an MT is used to decide the optimal MAP in a multi-layer hierarchy of MAPs. The mobility signaling cost of using the best MAP can be used as a key attribute in the proposed mobility-based network selection schemes.

Chapter 7 describes studies and suggestions on some other network selection issues, including traffic load assignment, vertical handover tradeoff, etc. Based on all the study above, an integrated network selection strategy is proposed. Besides, some suggestions on how to do network selection for NEMO are given at the end of this chapter.

Chapter 8 finally summarizes the whole work and contributions, and points out some potential directions for further research activities.

Appendix A, besides, describes a network selection simulator, which is established by Matlab and used for simulations during all the studies in this dissertation.

2. STATE OF THE ARTS

This chapter provides the state of the arts, including background of the topic, information gathering schemes and most important a survey of network selection schemes mainly based on mathematical models used for this issue.

2.1. Background

Before presenting our study on network selection, we firstly provide, in this section, three indispensable preliminaries: evolution of HWNs, IPv6 mobility and multihoming protocols, and ABC components.

2.1.1. Evolutionary of HWNs

Nowadays, multiple wireless networks are being developed simultaneously, and these networks have different characteristics and could complement each other. The relationship between these networks with respect to several main characteristics is shown in figure 2-1. In our study, a large number of characteristics of these networks should be considered, including monetary cost, power consumption, mobility support capability, bandwidth, bit error rate, and so on. In this part, we are going to provide a brief description of various wireless networks, including WWANs, WMANs, WLANs, WPANs and satellite networks.

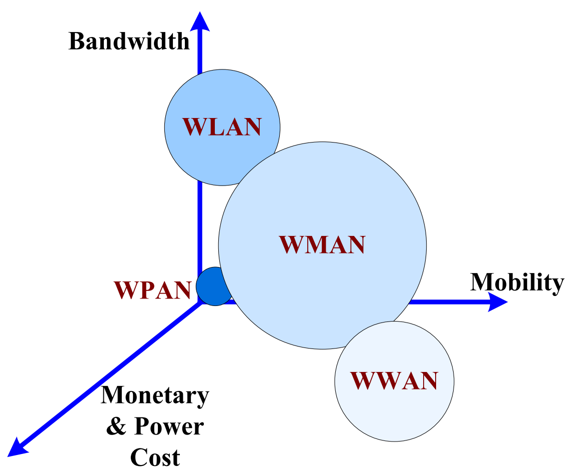


Figure 2- 1 Relationship between various wireless networks.

Wireless WAN: Emerging in the mid of 1980s and experiencing an exponential growth from the early 1990s, the WWAN cellular technologies have been considered as a worldwide success. At the moment of writing this dissertation, both UMTS and cdma2000 achieves several hundreds of millions of subscribers on this planet.

Figure 2-2 shows the evolutionary path of WWAN technologies. During the past twenty years, WWAN has been developing from 2G to the current widely deployed 3G mobile communication systems, and are now evolving towards the future beyond 3G heterogeneous all-IP networks.

There are two main families of WWAN systems: mobile application part adopted by 80% of the subscriber base and IS-41 holding the remaining 20%. They are, respectively, standardized by two main organizations working on standardization of WWAN communication systems, i.e. 3GPP and 3GPP2. Along with the evolvement, the most noticeable feature of these systems in both families is the greatly increasing downlink data throughput.

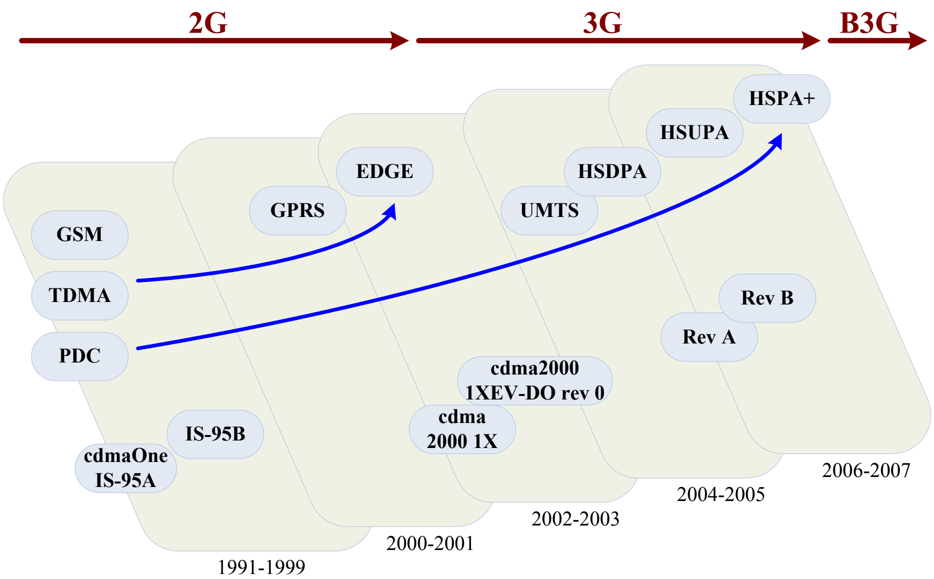


Figure 2- 2 The evolutionary of WWAN technologies [LP08].

Another feature of wireless communication systems is the radio spectrum, which is a critical and complex issue, due to the scarce of resource and historical reasons. GSM was initially built on the 900MHz frequency band; while other systems were assigned to use frequency band around 2000MHz.

Wireless LAN: There are two main organizations which produce standards for WLAN systems (i.e. IEEE and ETSI). IEEE 802.11 series are the most popular used standards for WLAN. The development of WLAN can be traced back to the mid of 1990s when the first IEEE 802.11 standard and

HyperLAN standard were defined. The evolution of WLAN standards during the past decade is shown in figure 2-3.

Due to the large bandwidth of WLAN technologies, it has many applications for both indoor and outdoor customers. Various services, e.g. VoIP, VoD, video conferencing, etc., can be supported by WLAN networks.

Nowadays, this technology is used to set up some hotspot access in both public and private places, e.g. coffee houses, airports, offices, etc. Seen from large scale coverage, the deployment of hotspots is relatively random, which might be modeled as a Poisson point process. Nomadic usage is common for WLANs due to their small cell radius.

By contrast, according to the recent study on IEEE 802.11 wireless mesh networks, WLAN is likely to be deployed for city-wide ubiquitous coverage, but IP-layer handover is probably processed frequently. As we know, mobile IP requires the new care-of address (CoA) being registered to the home agent (HA) which might be far away from the visited network, so certain scheme, e.g. fast handover, may be used.

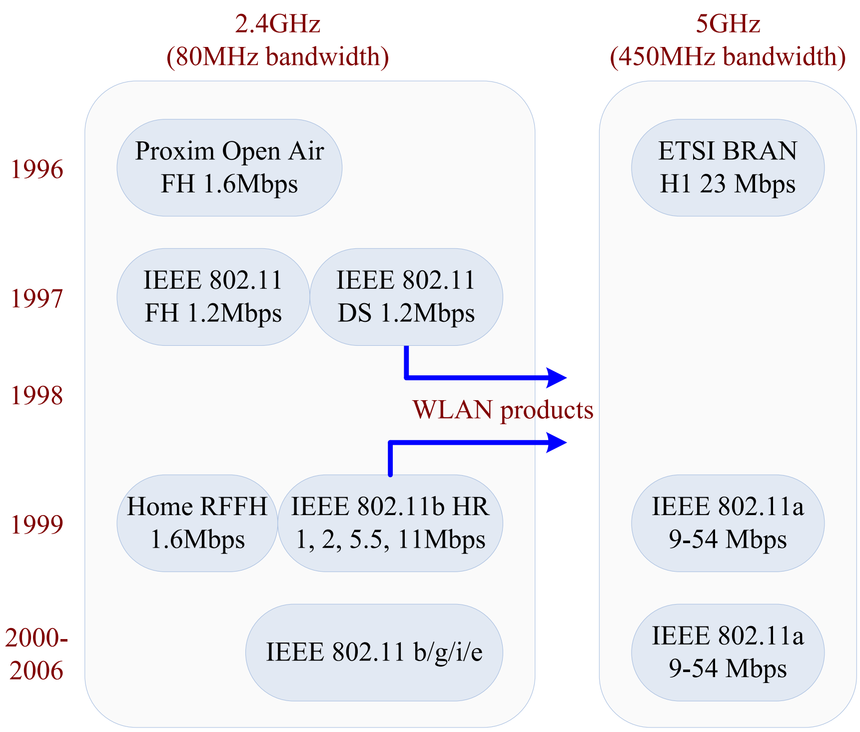


Figure 2- 3 The evolution of WLAN standards.

Wireless MAN: As a technology complementing the gap between WWAN and WLAN, there have been many standards, e.g. WiBro in Korea and ETSI HiperMAN in Europe. Among all the standards, the Worldwide Interoperability for Microwave Access (WiMax), developed by IEEE 802.16 working group, is the most popular. According to recent studies, WiMax has now mainly two types of usage models: fixed WiMax (IEEE 802.16-2004) and Mobile WiMax (IEEE 802.16e), see figure 2-4.

IEEE 802.16-2004 is designed to serve as a wireless DSL replacement technology, to compete with the incumbent DSL or broadband cable providers or to provide basic voice and broadband access in under-served areas where no other technology exists. It is also a viable solution for wireless backhaul of WiFi access points or potentially for cellular networks. Moreover, it can be used as a T1 replacement option for high-value corporate subscribers. By contrast, IEEE 802.16e is intended to offer portability and eventually full-scale mobility.

Unlike WWAN with a large coverage but small throughput or WLAN with a large throughput but small coverage, WiMax technology could reach a theoretical 30 miles coverage radius and achieve data rates up to 75 Mbps. However, there is surely a tradeoff between the coverage and the throughput.

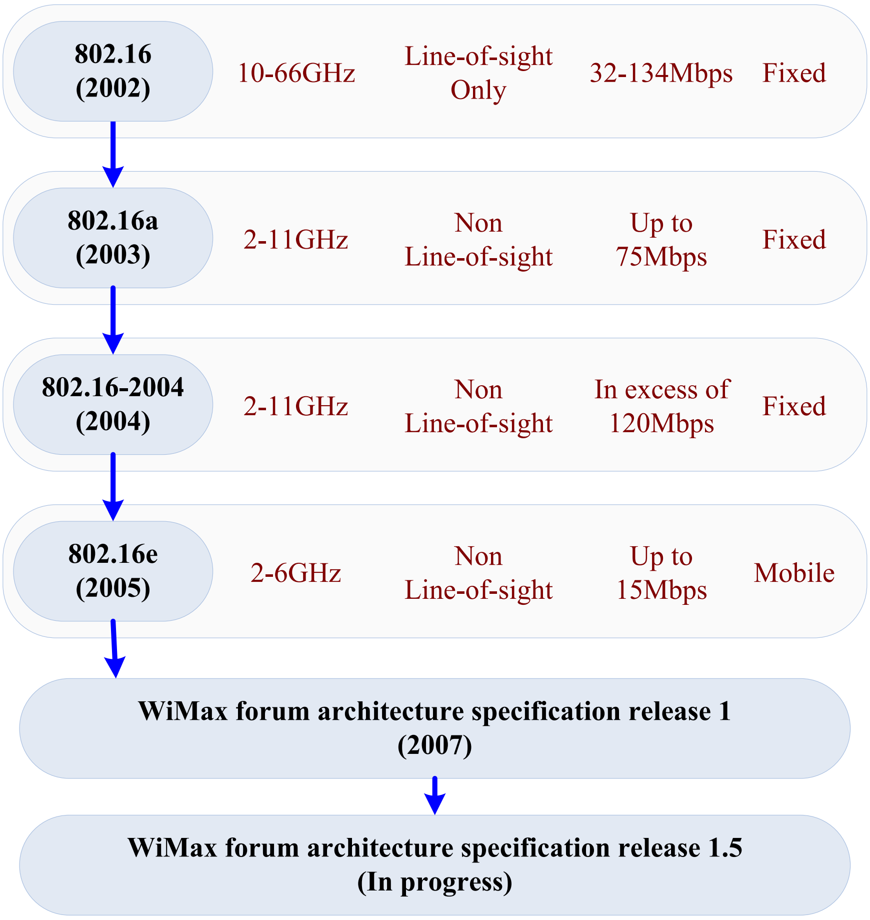


Figure 2- 4 Development and key features of WiMax standards [Erxn].

Wireless PAN: WPAN was firstly studied in the 1990s, and it was designed for low-power short-range connectivity. Bluetooth is one of the most typical

WPAN versions. The maximum transmission range is as large as 100 meters, but its general applications are only around 10 meters. Bluetooth works on unlicensed Industrial-Scientific-Medical (ISM) band 2.4GHz.

Considering the short coverage range of Bluetooth, its distribution could be similar to that of WLAN networks. Compared with WLAN, WPAN has a better power saving property, but the maximum data rate of traditional WPAN is not attractive. WPAN with UWB technology might provide a large data rate in the near future.

Up to now, the main usage of Bluetooth technology is as a replacement of low-rate personal area cables, but there is the possibility to use it for establishing short-range access networks.

Satellite networks: Satellite communications have been used a lot in the past decades. Although the uplink bandwidth is usually low and the communication quality might be largely affected by e.g. moisture in the sky, it still has many useful applications, e.g. positioning. Besides, satellite network is an important complement to the other wireless networks. That is because a satellite could cover an extremely large area with even billions of population. Many applications can choose to be transmitted through satellite networks, such as emails, broadcasting, positioning, and so on.

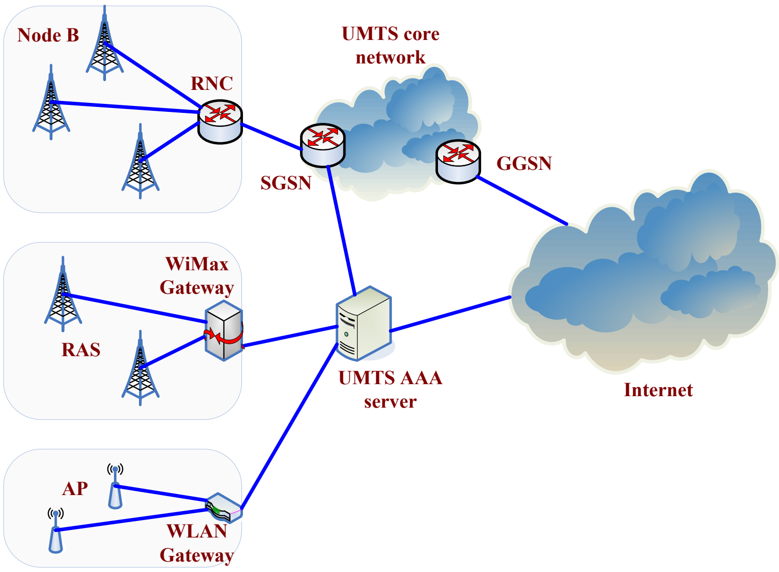
Satellite technology has many advantages over terrestrial communication technologies, including

- large coverage area,
- point to point broadcast, and
- rapid development as compared to erecting ground relay towers or laying cables over long distances or difficult terrain.

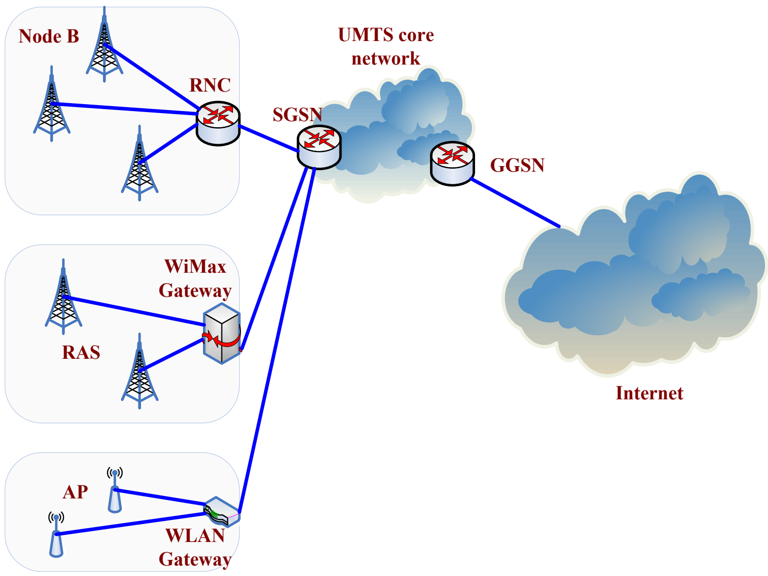
Interworking of HWNs: The future HWNs require a flexible architecture to combine QoS with resource management and handover strategies. Various networks described above might interwork together to provide different types of services to customers using multi-mode or multi-homed terminals. And, when new access technology is deployed, it could be easily combined with previous interworking networks. In recent research and standardization process, the internetworking of WLAN and 3G has been studied and many proposals, e.g. loose coupling and tight coupling, have been proposed. Loose coupling interconnects different networks independently, by utilizing a common subscription, as shown in figure 2-5(a). By contrast, tight coupling suggests transfer data and signalling through a single core mobile network, as in figure 2-5(b).

2.1.2. IP mobility and multihoming

Therefore, in this section, we are going to provide a brief description of several main protocols on IPv6 mobility and multihoming, which is an important preliminary of our work on network selection.



(a) Loose coupling



(b) Tight coupling

Figure 2- 5 Interworking of HWNs.

Mobile IPv6 [JD04]: IP mobility support for IPv4 and IPv6 have been specified by RFC 3344 in 2002 and RFC 3775 in 2004, respectively. Mobile IPv4 (MIPv4) has been deployed on a wide basis (e.g. cdma2000 networks), by contrast, the deployment of Mobile IPv6 (MIPv6) is still rare (e.g. HP-UX MIPv6) [HPUX]. According to research on IPv6-related techniques and support of IPv6 on operating systems in the past dozen of years, MIPv6 seems have more chance to be utilized in the near future than before.

MIPv6 is one of the main protocols for IP macro mobility of MTs in the future HWN environment [SD04]. The basic idea of MIPv6 is to use a home agent (HA) in the home network of the MT for delivering packets between the MT and its correspondent nodes (CNs). An MT with MIPv6 has one home address (HoA) assigned by its HA. When the MT moves to a visited network, it gets a care-of address (CoA) from the visited network and registers the CoA to its HA as a binding between its HoA and CoA. Once the binding registration is finished, the MT could communicate with its CNs through the HA. That is, the MT packs packets using an MIPv6 header destined to the HA. The HA unpacks the packets and checks the bindings in its cache to deliver these packets to the CNs. Reversely, a CN sends packets destined to the MT's home network. The HA catches these packets and delivers to the MT by packing it with an MIPv6 header. Briefly speaking, MIPv6 establishes a tunnel between the MT and its HA to deliver packets when the MT is not in the home network.

Obviously, the basic MIPv6 operation has a triangular routing problem (dog-leg problem). That is, even the MT is close to the CN, all the packets have to pass through the HA which might be far away. In order to solve this problem, Route Optimization (RO) mode was designed, in which the MN registers its CoA to not only the HA but also all its CNs, so that packets between the MN and its CNs don't have to route through the HA (except the first packet, which is used to indicate its CoA to the CN). Besides, a type-2 routing header is defined to carry the MT's HoA. However, RO mode requires the CN-side terminal to support the MIPv6 protocol, which is probably an obvious disadvantage for its popularity.

Besides, MIPv6 does not support the registration of multiple CoAs to the HA or CNs, so a selection procedure should be used by the MT itself to select the primary CoA and register it to HA and CNs. Thus, the existence of multiple CoAs is transparent to the MT's HA or CNs.

Hierarchical Mobile IPv6 [SH05]: When a MN with RO mode moves frequently in a VN, it has to send binding update (BU) messages to the HA (and CNs in RO mode) any time it updates its CoA, which leads to a large traffic load, especially when the MN is far away from its CNs. To decline the BU cost, several micro-mobility protocols were proposed [CA02]

[AI04], including HAWAII, cellular IP, Hierarchical Mobile IPv6 (HMIPv6), etc. In all these protocols, HMIPv6 is paid more attention than others in recent years.

In HMIPv6, a Mobility Anchor Point (MAP) which serves as a local HA is used for micro-mobility. When a MN moves into a VN, it first gets an on-link CoA (LCoA) which belongs to the prefix of the attached router (AR). Then, the MN registers this LCoA to a MAP and binds it with a Regional CoA (RCoA) which belongs to the prefix of the MAP. After local registration, only the RCoA is registered to the HA and CNs, while the LCoA is totally transparent to the outside of this MAP region. When the MN moves within the coverage of a MAP, it changes its LCoA and registers it to the MAP, which means no global BU registration is needed. Since one MAP usually covers a large group of ARs, most handovers are achieved by local BU registration, while only those handovers between MAPs should still require global BU registration. Therefore, HMIPv6 greatly decreases BU cost by changing global BU registrations into local ones.

Proxy Mobile IPv6 [SH05]: As a micro mobility protocol, HMIPv6 could decrease signalling cost when the MN is not in its home network, but this protocol requires modification on the terminal-side. Considering that, in this world, there are billions of MTs without IP mobility support, to improve their IP mobility functionality with these new protocols could be inconvenient. Therefore, it can be better if the change is only on the network-side, and a new protocol, called proxy mobile IPv6 (PMIPv6), satisfies this requirement.

PMIPv6 is a protocol for network-based mobility management. It enables IP mobility for a host without requiring its participation in any mobility-related signalling. The network is responsible for managing IP mobility on behalf of the host. The mobility entities in the network are responsible for tracking the movements of the host and initiating the required mobility signalling on its behalf.

Detailed speaking, PMIPv6 uses mobile access gateway (MAG) and local mobility anchor (LMA) to manage mobility. MAG is a function on an access router that tracks a MN’s movement and manages the mobility-related signalling taking the place of the MN which is attached to its access link. LMA is as a HA in the PMIPv6 domain, so it manages binding updates sent from MAGs in this domain.

When a MN enters a PMIPv6 domain and attaches to a MAG, the MAG assigns a proxy care-of address from the home network prefix to this MN. Since the home network prefix for the MN does not change during the movement of the MN in the PMIPv6 domain, this proxy care-of address does not have to change and the MN feels itself always connecting to the same link. That is why the participation of MN for mobility support is not required.

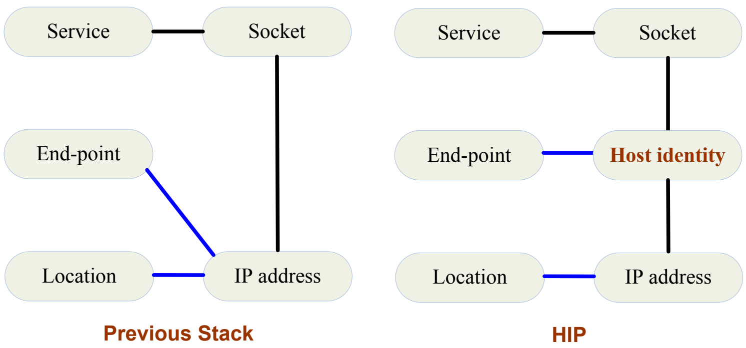


Figure 2- 6 Host identity protocol stack.

HIP [MR06]: host identity protocol (HIP) is specified by IETF HIP working group. The idea is to separate the functionalities of IP address on identity and location. Hence, a new identity, called end-point host identity, is defined, as shown in figure 2-6. The IP address continues to be the locator, while host identifier is used above IP layer as a host identity. Worth mentioning that this host identity is different from an interface identity and a single host identity can be reachable through multiple interfaces.

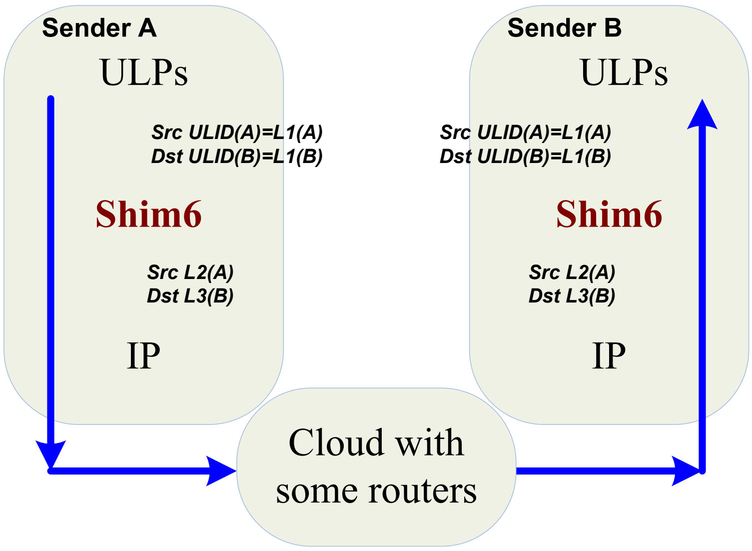


Figure 2- 7 Mapping with changing locators in shim6.

Shim6 [NE07]: Shim6 is specified by IETF site multihoming by IPv6 intermediation working group. Similar to HIP, shim6 is a layer 3 protocol for providing locator agility below the transport protocols, so that multihoming can be provided for IPv6 with failover and load sharing properties.

As shown in figure 2-7, shim6 layer is placed within the IP layer, above IP routing sub-layer and below IP endpoint sub-layer and Upper Layer Protocols (ULPs), e.g. TCP and UDP. Upper layers use upper layer identifier (ULID) which is mapped to different locators by the shim layer. To perform this mapping, the shim6 layer maintains a ULID-pair context state per ULID pair, so that upper layers see packets sent with ULIDs from end to end.

LISP [FD09]: Proposals on separating locator and identifier are widely discussed, and Shim6 and HIP are just two of them. Another mentionable proposal is Locator/ID separation protocol (LISP). Compared with Shim6 and HIP, the best part of LISP is that no change is required on the host-side.

Detailedly, IP addresses, called Endpoint Identifiers (EIDs) in LISP, are still used for tracking sockets, connections, and for sending and receiving packets as they are now. Between sender and ingress tunnel router (ITR) and between egress tunnel router (ETR) and receiver, routers continue to forward packets based on IP destination addresses. However, between ITR and ETR, packets are transmitted by LISP encapsulation using a LISP header. The address used to deliver the packets with LISP headers is called a routing locator (RLOC). Therefore, protocols on hosts stay the same as before, while the change is the LISP tunnel part and the mapping between EIDs and RLOCs on both ITRs and ETRs.

Monami6 [WR07]: A MN may have multiple CoAs because of multiple interfaces or multiple prefixes on single interface. Instead of registering the first CoA to the HA and CN, a new MIPv6 extension was defined by MONAMI6 IETF working group to register multiple CoAs.

A new identification number called Binding unique IDentification (BID) number was proposed for each binding cache entry to accommodate multiple bindings' registration. The BID is assigned to either interface or prefix bound to a single HoA of a MN to distinguish between multiple bindings. The MN notifies the BID to both its HA and CNs by containing a newly defined BID sub-option in the BU message. As shown in figure 2-8, the BID is recorded into the HA and CNs' binding caches as a search key. When one of the CoAs changes, the MN sends a BU message including the new CoA and the corresponding BID. HA and CNs search their entries according to the HoA and the BID, and update the new entry. If a MN decides to act as a regular MIPv6 MN, it just sends a BU without a BID sub-option, so that the receiver of this BU could delete all the other bindings for this MN registered with BIDs.

Besides, HMIPv6 can have a similar extension as this MONAMI6 solution. The only difference is that multiple LCoAs could be registered to the MAP instead of multiple CoAs to the HA. Similarly, the case where multiple proxy CoAs assigned by multiple MAGs in PMIPv6 networks has been discussed in the scope of IETF Multiple interfaces (MIF) working group [BC09].

Binding1:	HoA	BID1	CoA1
Binding2:	HoA	BID2	CoA2
Binding3:	HoA	BID3	CoA3

Figure 2- 8 Searching Binding Entries with MONAMI6.

Network Mobility [DV05]: A NEMO is an entire network, moving as a unit, which dynamically changes its point of attachment to the Internet. It is composed of one or multiple IP-subnets and is connected to the global Internet via one or multiple Mobile Routers (MRs). A NEMO is composed of MRs and mobile network nodes (MNN) which can be fixed in the NEMO or mobile.

A basic mobility approach of NEMO is for each MR to have a HA, and use bidirectional tunneling between the MR and HA to preserve session continuity while the NEMO moves. The MR will acquire a CoA from its attachment point much like what is done for MTs using Mobile IPv6. This approach allows nesting of NEMOs, since each MR will appear to its attachment point as a single node. This mobility approach is an extension of MIPv6 to network's mobility, so it is backward compatible with MIPv6. It is worth mentioning that a NEMO compliant HA can operate as a MIPv6 HA.

HAHA [TP06] [DV06] [WR04]: Although the mobility of a NEMO is similar to that of a host, Route Optimization (RO) is only suitable for mobile host, not for NEMO. That is because a NEMO may have hundreds of CNs leading to large traffic cost for BU updates in the RO mode. Therefore, it is necessary to find another method to avoid the overhead of triangular routing bypass HA between CN and MN. Global HAHA protocol was proposed to achieve a relatively optimal route from MN to CN by deploying multiple HAs globally.

In the Global HAHA protocol, a proxy located at a nearby site of the MN is introduced, which acts as a HA for the MN and a MN for the original HA. Specifically, this proxy terminates the MN-HA tunnel and the associated encryption, extracts packets and re-encapsulates them to the other side of the tunnel, as shown in figure 2-9.

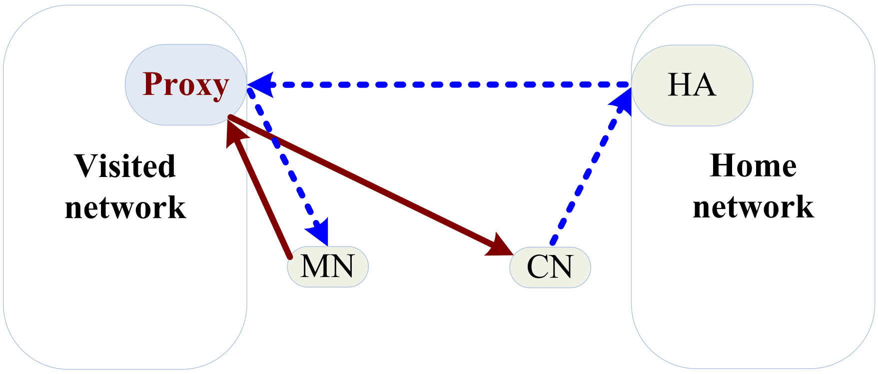


Figure 2- 9 Sub-optimal route and HAHA tunnel with Global HAHA protocol.

The MN can find its nearest proxy by DHAAD mechanism. Then, it registers to this proxy with ordinary BU message. After that, the proxy will perform BU to the MN's primary HA pretending to be the MN. The proxy can find out the MN's primary HA, because it could receive Router Advertisement (RA) messages from other agents. With this kind of BU registration, packets sent by MN could be transferred by the proxy directly to its CN, which is supposed as a relatively optimal RO mode, but packets from CN to MN still have to pass through the primary HA and the proxy.

To further optimize the route, the proxy pretending the MN with RO mode could send BU to all the CNs. Thus, CNs will send packets directly to the proxy because they think the updated address of the proxy is the MN's CoA. Since the proxy is usually on the top of the hierarchy of the visited network, the MN does not have to update its CoA while moving within the visited network.

It is worth mentioning that this proxy is similar to a gateway MAP in HMIPv6 with RO mode.

2.1.3. Always best connected

During the 1990s, the widespread deployment of the 2G WWAN system brought to us the notion of being *always connected*. Along with the development of communication technologies in the past two decades, multiple wireless access technologies came out and complement with each other, as explained in section 2.1. In an environment of heterogeneous wireless access networks, the previous concept of being always connected becomes *always best connected*. This refers to being not only always connected, but also being connected in the best possible way, combining the usage of various networks' resources.

ABC brings plenty of advantages to customers. With ABC functionality, MNs could select appropriate access networks to fit for various applications' QoS requirements; MNs could avoid selecting a network with high traffic load hence avoiding congestion; MNs could predict networks' availability so that they do not connect to networks which disappear soon; MNs could minimize signaling costs by using network selection and handover decision strategies specifically for this purpose. Moreover, ABC benefits operators. Since ABC has the feature of assisting the assignment of traffic load to multiple networks, operators could maximize the utilization rate of the resource of the networks they operated, hence maximizing revenue; according to network selection strategies operators could analyze and decide the number of WiFi access points they should deploy to attract customers to WLANs. Finally, ABC is suitable to synthetically consider customers' and operators' benefits, so that a win-win partnership can be achieved.

During our studies, we designed a series of scenarios which could be used to explain the requirements of ABC for customers' daily communications, as shown in figure 2-10 and further explained in table 2-1. In this series of scenarios, Caro holding a multi-homed MT and Will holding a multi-mode MT work together in the same company. One morning, they surf the Internet for some project-related information, using their MTs. During this period, wireless networks, including WWAN, WMAN, WLAN and WPAN, are all available. Then, they attend a meeting in their company's meeting room, where there are a lot of colleagues using WPAN. After the meeting, they take coffee in a coffee house, where WLAN is free only for those who buys staffs from the coffee house. After having their coffee, they take a taxi to a university to give a presentation during a conference. However, on the way to the university, they realize that the conference has been started, so they have to use video conferencing software to attend it in the taxi. In one hour, they arrive and give their presentation in the auditorium of the conference, where WLAN is totally free for attendees. Unfortunately, their MTs are about to be out of power, but they cannot charge them during their presentation.

During the above scenarios, there is more than one available network in each scenario. For some scenarios, we have the feeling that certain networks can offer a better connectivity quality than others; but for some ambiguous scenarios, we really cannot intuitively tell our preference. That is because every network has some reasons to be selected and the final selection depends highly on how we evaluate the relative importance of these reasons.

ABC service means to being always connected to the best network in any specific scenario, and a good ABC solution should be able to select appropriate networks even in those ambiguous scenarios.

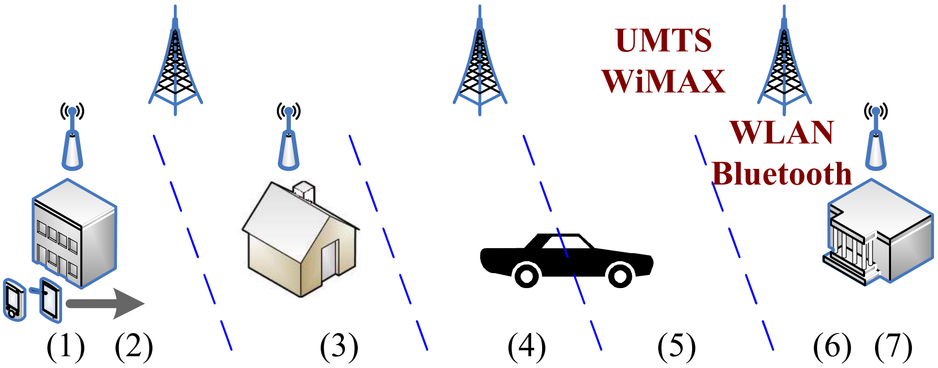


Figure 2- 10 A series of scenarios for ABC.

Table 2- 1 Details of a series of scenarios for ABC.

N.	Location	Changes of Apps	Changes of Networks & MTs
1)	office	+ WWW	–
2)	meeting room	–	WPAN High-traffic
3)	coffee house	–	WLAN NOT Totally Free
4)	taxi	–	WPAN Low-traffic
5)	taxi	+ Video Conf.	–
6)	auditorium	–	WLAN Free
7)	auditorium	–	Battery Low

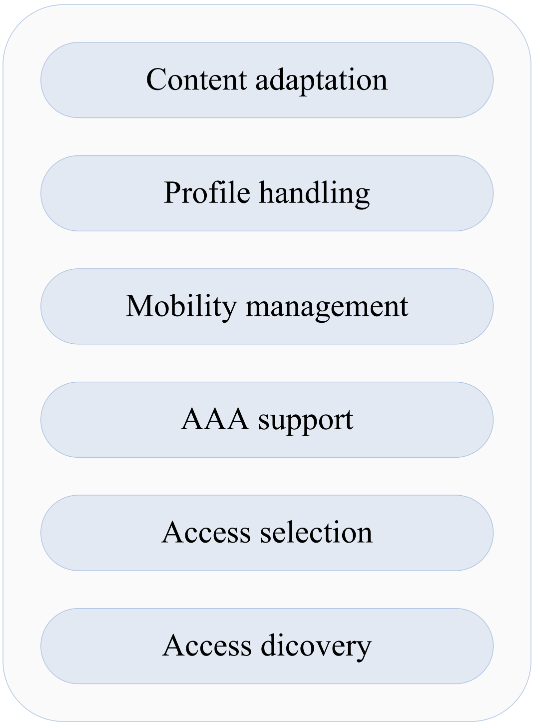


Figure 2- 11 Technical components of an ABC solution [GE03].

In order to build an ABC solution, we need many functional blocks: network discovery, network selection, AAA, mobility, management, profile handling, content adaptation, etc., as shown by figure 2-11. In the following, we are going to briefly explain these blocks one by one.

Network discovery: Before using any scheme to select the best network, it is necessary to firstly discover all the available networks. For a terminal-side network selection scheme, this process is for the MT to know the key features of these networks. By contrast, if the network selection entity is on the network-side, there should be a similar step to transfer all the network and terminal information to the network selection entity.

There are many issues for this step. One is to define the set of parameters that should be utilized for describing these networks' key features, e.g. monetary cost, bandwidth, coverage, etc. The usage of set of parameters largely decides the quality of the ABC service provided to the customers.

Another issue is how to gather all the information of networks, because this ABC service might increase a lot of signaling between access operators, ABC service provider and MTs. Moreover, the service provider may be independent from the network operator, so some dynamic information, e.g. traffic status of networks, may be difficult to be obtained.

A third issue is when to do network discovery. This procedure can be triggered either periodically or by a specifically designed set of trigger events. Many functional entities in an ABC solution could sense the change of condition and provide related information to trigger a content adaptation. An obvious entity that could provide these triggers is the MN. Once the MN has selected an access network or the session is transferred to a new personal device, trigger should be sent to application servers to adapt the current applications. Another entity is the ABC server which gathers all the information and analyzes whether an adaptation is required. The third entity is access providers, but this entity might not sense the change on the customer-side.

Besides, the gathered information could be incomplete or even incorrect, so some effort might be required to process incomplete or incorrect information.

Network selection: Network selection is to selection the best network and access technology based on information of a large number of parameters collected during the network discovery step. In the current literature, there are plenty of proposals on how to select the best network among multiple options.

Network selection scheme is the main contribution of this dissertation, which will be further explained in latter chapters.

AAA: Since ABC service is provided by a HWN environment, authentication, authorization and accounting for various services to the customer is required to be designed for collaboration of these networks. The detailed AAA scheme depends on the interworking strategy of multiple networks. For loose coupling, each network should have an AAA server, and signalling between these AAA servers should be designed to support ABC service. For tight coupling, it is possible to use a single AAA server for all the networks, but AAA scheme must be much more complicated than that of loose coupling, due to its management of multiple networks.

Mobility management: Mobility management includes handover and location management. Since ABC service is provided within a HWN environment, we require a macro-mobility solution (e.g. Mobile IP) for location management when the MN is moving among these networks. Similarly, handover here is mainly about VHO among different networks. My work in this dissertation is highly related with terminal mobility within HWNs. For more information, please refer to section 2.2.

Profile handling: AAA is about the logon information concerning a single network; by contrary, an ABC customer should provide his or her ABC profile before using this service, which is called profile handling. Profile handling should also include the delivery of the customer's preference on network selection. Probably also include the terminal and application properties.

Content adaptation: Due to the difference between various networks and between various personal devices concerning session mobility, the application should be able to adapt its key features, e.g. size of the video picture and quality of music, to the current condition. When a multi-homed terminal uses multiple access networks to share the traffic of a single application, it is necessary to consider transmissions through all the networks in order to perform a precise content adaptation.

2.2. Information gathering

2.2.1. Required information

Among all the components of ABC, network discovery has a quite close relationship with network selection. That is because the design and the performance of a network selection scheme is highly related with both the available information used in the scheme and how the information is gathered before usage. Due to related works, the information is usually classified into various groups to facilitate its usage. Therefore, here we first describe our classification of information for network selection.

Network selection information is called in general 'factors' in this dissertation. Factors are the basis of selecting the best network for an MT or a traffic flow of a multi-homed MT. A network selection scheme should synthetically consider multiple factors. In fact, a large number of factors should be considered, including not only network attributes but also terminal properties. In related work, a large number of factors have ever been used by one or multiple proposals, but there is little specific discussion on why these factors are chosen and why they are used in such ways. As we know, using certain essential factor in a wrong way may result in sub-optimal selection results, so we emphasize that the choice and the usage of factors are quite important in the network selection issue.

In our categorization, we first classify all the factors into two groups: network-side attributes and terminal-side requirements. The former group includes static network properties, dynamic network properties and VHO properties; while the latter group includes terminal properties, customer preferences and application QoS levels. Now, these groups of factors are further explained as follows:

Static and dynamic network properties: these attributes represent the candidate networks' properties, including monetary cost, bandwidth, power consumption capability, security level, bit error rate, jitter, horizontal handover (HHO) signaling cost and latency, signal strength, traffic load information, etc. Among these attributes, some are static, such as monetary cost; some are dynamic, such as signal strength; while some attributes are semi-dynamic, such as bit error rate. Bit error rate is dynamic because it might change due to the change of wireless channel condition, but bit error rate (BER) of a network does not change a lot compared with BER of other networks, so we called it a semi-dynamic attribute. In our later research, we will use specifically mobility-related attributes as dynamic attributes for mobility-based network selection. By contrast, static and semi-dynamic attributes are both supposed with stationary values (i.e. average values).

VHO properties: they are not network attributes because they are not properties of a certain network but that of an ordering of networks. This group includes two main factors: VHO signaling cost (e.g. binding update cost and packet tunneling cost of MIPv6, see chapter 7) and VHO latency. VHO signaling cost is dynamic, while VHO latency is considered as semi-dynamic in our research. We use this group for both network ranking and VHO tradeoff.

Terminal properties: battery state and MT velocity are two main factors in this group. Battery state is an important factor to decide whether a network with good power saving feature is strongly required, while MT velocity is highly related with mobility-related network attributes and VHO properties.

Customer preferences: this group includes several options, e.g. low monetary cost, high bandwidth, high security level, etc. Customers should have the right to select one or multiple of the above options while purchasing the service or through their user-interface software. Similar to customer preferences, there are actually also operator preferences (i.e. policies), which might be considered by a network selection scheme.

Application QoS levels: applications can be divided into the following four levels based on their QoS requirements [3gpp]: conversational, streaming, interactive and background. Applications of different levels have different requirements and prefer different networks. For example, video-streaming requires bandwidth; Mobile VoIP requires low jitter and low handover cost; while E-mail requires security.

After explaining our categorization of all the factors, we then survey the most important related work on information gathering. Recently, many works have been performed for this purpose, including IEEE 802.21 working group, IEEE P1900.4 working group, and numerous proposals on context-aware network selection architectures, which will be described one by one in the following sub-sections.

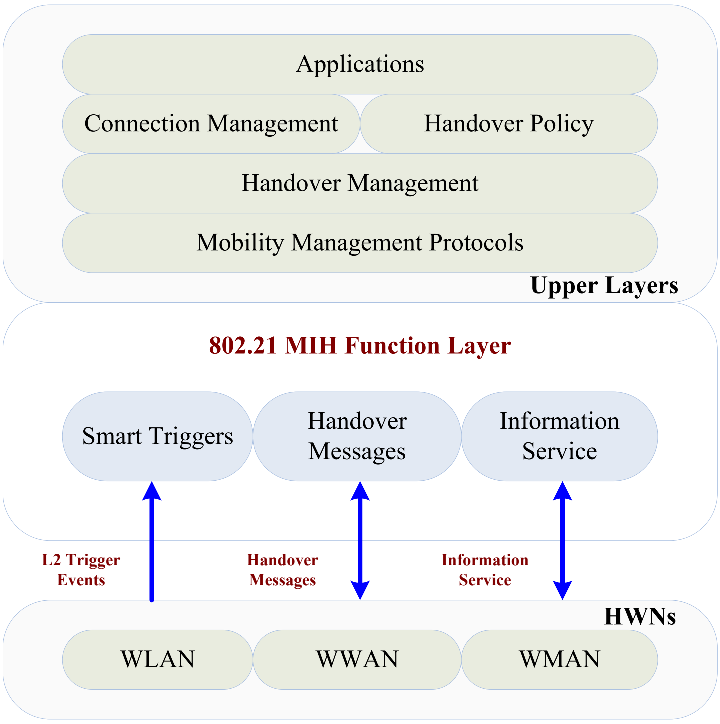


Figure 2- 12 IEEE 802.21 architecture.

2.2.2. IEEE 802.21 standard

In IEEE 802.21 standard [802.21], a media independent handover (MIH) framework is presented to facilitate handover with measurements and triggers from link layers. IEEE 802.21 defined a new layer between link layer and IP layer, which is called 802.21 MIH function layer. This new layer is used to handle handover triggers, handover messages and handover related information initiated by link layer. Figure 2-12 shows the service architecture of IEEE 802.21 standard. IEEE 802.21 specification defines the interaction between the 802.21 MIH function layer and the link layers of various wireless networks.

However, network selection decision making requires not only information of link layers. In [WY08] and [WY08-2], an enhanced media independent handover (EMIH) framework is proposed to collect more information from application layers and user context information.

The architecture of EMIH is shown in figure 2-13. The motivation is to collect and make full use of available information in both client side and network side to optimize network selection. EMIH uses many trigger events and collects the static and dynamic information. The following entities are defined to help gathering the information:

Client-side entities:

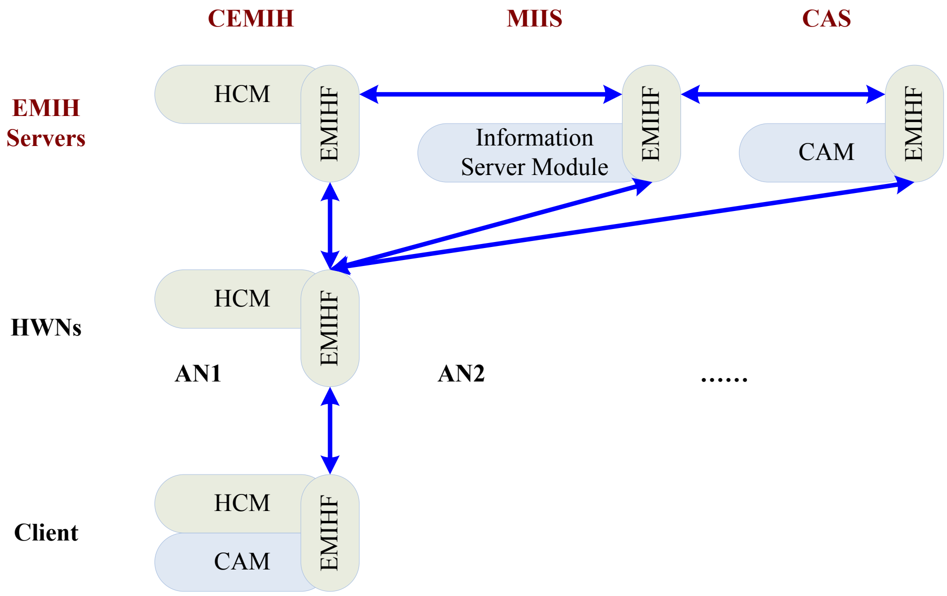


Figure 2- 13 EMIH architecture.

- EMIHF (EMIH function): to provide link layer intelligence and offer a unified interface between different access techniques and upper layer applications;
- CAM (context-aware module): to identify information of the mobile terminal, to generate trigger events, to transfer event and related information to HCM;
- HCM (handover control module): to support the MT controlled handover.

Network-side entities:

- AN (access networks);
- MIIS Server: including EMIHF and information service module. Network features collected in MIIS can be accessed by other entities;
- CAS (Context-aware server): to identify network context, to generate trigger events, to transmit these events to subscribers;
- CEMIH (control EMIH): to collect trigger events, to initiate a handover, to control handover signaling to pass core network, and to do network selection.

2.2.3. IEEE P1900.4

The objective of the IEEE P1900.4 [1900.4] is to define standardized protocols and corresponding reconfiguration management system architecture for the optimization of resource management, in order to provide improved capacity, efficiency and utility within heterogeneous wireless networks wherein devices support multiple air interfaces, with multihoming and dynamic spectrum access capabilities in licensed and unlicensed bands.

More specifically, the scope of IEEE P1900.4 consists in:

- providing protocols carrying information between network resource managers and device resource managers supporting wireless terminal and network reconfiguration management, including the context of heterogeneous networks;
- providing corresponding reconfiguration management functionalities of the wireless system for the support of efficient optimization of resource usage;
- providing corresponding management functions and standardized rules to allow the multi-mode and/or dynamic spectrum access capable devices making decisions in a distributed fashion whilst providing operators with fair and effective exploitation of network resources thanks to an exhaustive set of rules to be followed by user equipments.

The functional architecture defined in the IEEE P1900.4 standard is depicted in figure 2-14. This architecture specifies seven entities and six interfaces. The interfaces ensure interoperability of equipment from different manufacturers covering different parts of the 1900.4 system. Four entities are defined on the network side with functions described as follows:

- Operator spectrum manager (OSM) is the entity that enables operator to control dynamic spectrum assignment decisions of the NRM;
- RAN measurement collector (RMC) is the entity that collects RAN context information and provides it to the NRM;
- Network reconfiguration manager (NRM) is the entity that manages the wireless networks and terminals for network terminal distributed optimization of radio resource usage and improvement of QoS;
- RAN reconfiguration controller (RRC) is the entity that controls reconfiguration of RANs based on requests from the NRM.

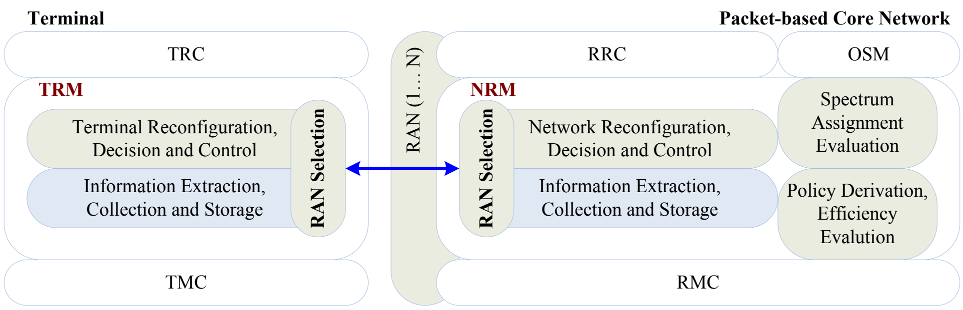


Figure 2- 14 IEEE P1900.4 functional architecture.

On the terminal side, three entities are defined for radio resource optimization:

- Terminal measurement collector (TMC) is the entity that collects terminal context information and provides it to the TRM;
- Terminal reconfiguration manager (TRM) is the entity that manages the terminal for network-terminal distributed optimization of radio resource usage and improvement of QoS within the framework defined by the NRM and in a manner consistent with user preferences and available context information;
- Terminal reconfiguration controller (TRC) is the entity that controls reconfiguration of terminal based on requests from the TRM

2.2.4. Context-aware information gathering

The concept of context-aware network selection can be defined as a network selection procedure that selects a target network based not only on the signal quality or explicit advertisements sent by network-side entities, but also on the knowledge of the context information of MN and networks, in order to take an intelligent decision.

In a HWN environment, the selection of the best network is much more complicated than the handover decision between base stations of a traditional homogeneous wireless network. Context information of both terminal-side and network-side are important for making this final decision. Therefore, a context-aware network selection strategy is necessary, as long as it is with an acceptable complexity.

To design a context-aware network selection strategy, a key issue is how to gather all the information, including network context, terminal context, user and application information, etc. All the information should be gathered as soon as possible, because the user cannot wait for a long time for the final decision. Moreover, some information might change frequently, which is an important feature of the information that a context-aware network selection strategy should consider.

Therefore, a context-aware network selection strategy requires first an architecture of context information management to assure that the information can be available in time. Second, the exchange of information between networks and the terminal should be minimized to save wireless resources. Third, after the gathering of the information, a scheme is required to combine all the information together for making the final decision. In this subsection, we focus on information gathering, and the third step will be discussed in subsection 2.

[BS04] provided a context-aware model which described both static and dynamic detailed information that should be gathered.

In [WQ06], an architecture was proposed for context-aware network selection, as shown in figure 2-15. Context information is stored in context information repositories, such as *Location information server (LIS), network traffic monitor (NTM) and user profile repository (UPR)*. Moreover, a *handover manager* is introduced to filter and process handover-related context information collected from various context repositories. A *Service deployment server (SDS)* is used to manage and install the service modules needed on both network-side and terminal-side entities.

A detailed architecture of context-aware network selection was given by [AT06]. As shown in figure 2-16, this architecture uses the gathered context-aware information for network selection with the following five steps:

- taking user inputs;
- mapping limit values from discrete preferences;
- assigning scores to available networks;
- calculating network ranking based on AHP method;
- session management.

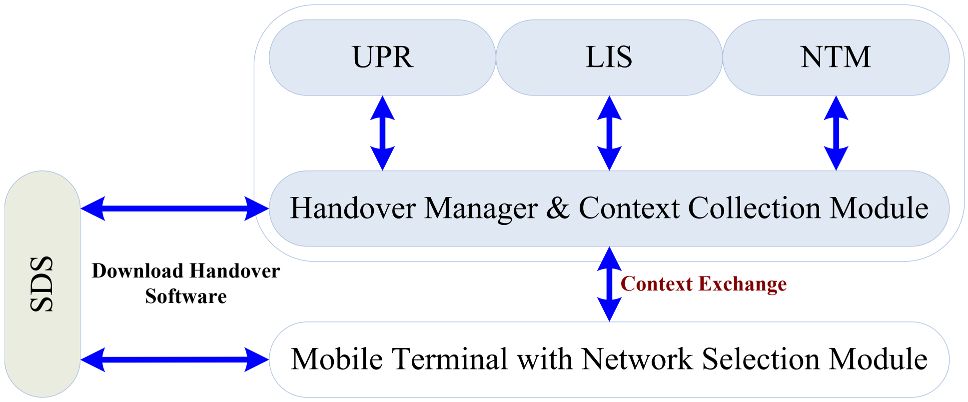


Figure 2- 15 An architecture of context-aware network selection.

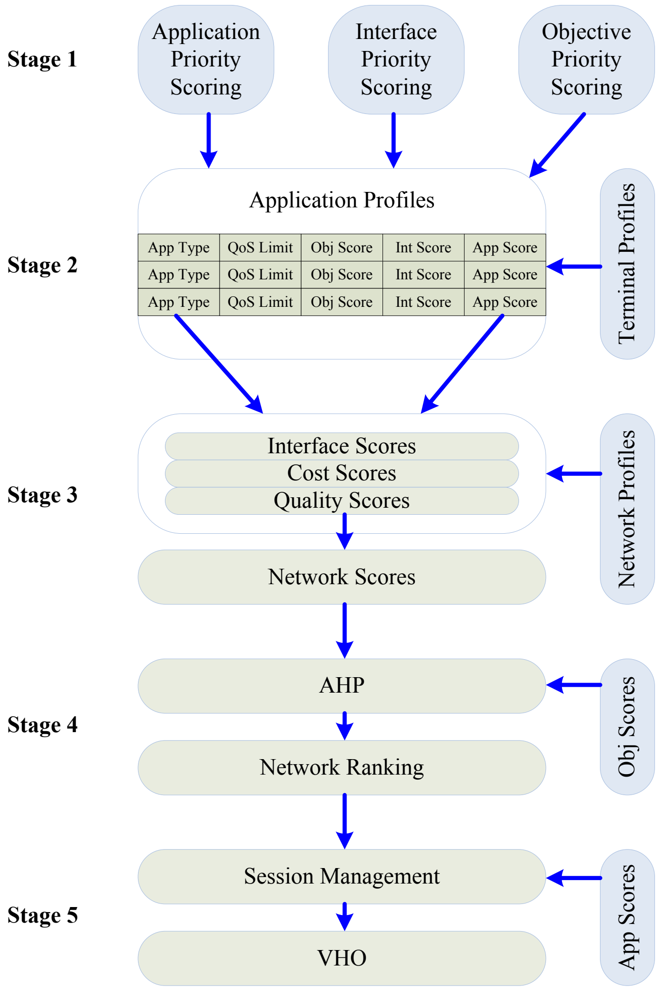


Figure 2- 16 A detailed architecture of context-aware network selection.

2.2.5. Pilot channel

Pilot channel is another way to gather required information for network selection.

[HP06] proposed to use a common pilot channel to deliver necessary information to initiate its connection. When a MT is switched on, it has neither information about the available networks in its area, nor information on the current spectrum allocation. In order to avoid the scanning of all the spectrum range and to facilitate the initial connection to the network, this terminal could listen to this pilot channel for broadcasted information to initiate its connection. Information of the common pilot channel should be received everywhere, so a wide-range access technology (e.g. UMTS) is suitable for this task.

In [YJ07], a cognitive pilot channel is used for gathering the information of networks. Cognitive pilot channel is being investigated within the European project E2R. It is a channel that would bring information to the terminal in order to facilitate its initial connection and handover to the surrounding networks. The terminal would benefit from lower power consumption, because of avoiding a scanning process of the whole frequency range. This pilot channel assisted network selection procedure is shown in figure 2-17.

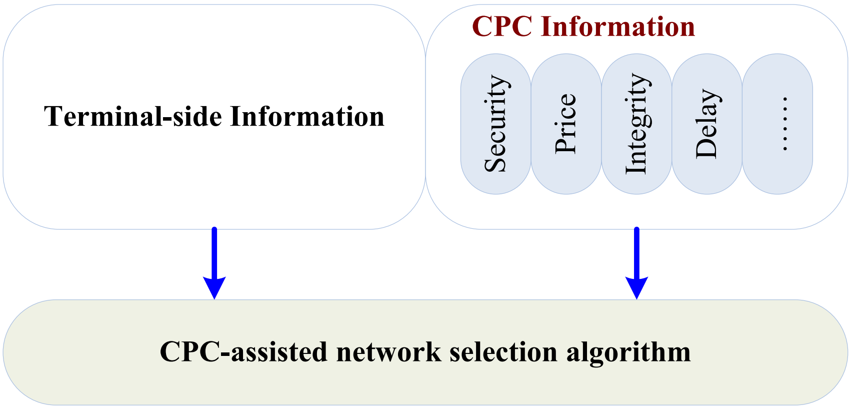


Figure 2- 17 CPC-based network selection scheme.

2.3. Network ranking schemes

2.3.1. Cost/Utility function

The traditional handover decision policies are based on received signal strength (RSS) [PK00, PG96], e.g.

- Handover if $RSS_{new} > RSS_{old}$;
- Handover if $RSS_{new} > RSS_{old}$ and $RSS_{old} < T$, where T is a threshold to avoid frequent handovers;
- Handover if $RSS_{new} > RSS_{old} + H$, where is called a hysteresis, in order to avoid ping-pong effect;
- Handover if $RSS_{new} > RSS_{old} + H$ and $RSS_{old} < T$; etc.

For HHO, it is common to use RSS as the criterion to make a handover decision, but this is not sufficient for VHO decision because RSSs of different networks cannot be compared directly, and moreover, RSS cannot reflect network features adequately.

As early as 1999, [WH99] proposed a cost function based network selection scheme which considers bandwidth, power consumption and monetary cost.

Cost function is a measurement of the benefit obtained by handing over to a particular network. A general form of cost function in the network selection issue is given in [MJ04], which integrates all the QoS values and their corresponding weights, and furthermore, they combine a network elimination factor with the function, given by

$$f^n = \sum_s \left(\prod_i E_{s,j}^n\right) \sum_j f_{s,j}(w_{s,j}) N(v_{s,j}^n), \tag{2-1}$$

where $v_{s,j}^n$ represents the QoS parameters, such as bandwidth, jitter, etc. *n* in this equation is to represent the *n*th network. *s* is the index of services, and *j* represents the *j*th network. $N(v_{s,j}^n)$ represents the normalized value of QoS parameters, $f_{s,j}(w_{s,j})$ represents their weights. And, the network elimination factor $E_{s,j}^n$ is used to reflect whether current network conditions are suitable for requested services. For example, if a network cannot guarantee the delay requirement of certain real-time service, its corresponding elimination factor will be set to infinite. Hence, the corresponding cost becomes also infinite, which eliminates this network.

In micro-economics, utility means the ability of a good or service to satisfy a human need. An associated term is utility function which relates to the utility derived by a consumer from a good or service. Different consumers with different user preferences will have different utility values for a same product. Thus, the individual preferences should be taken into account in the utility evaluation.

When evaluating the utility of an access network, we should distinguish between the upward and downward criteria associated with a network. The criteria of the higher preference relation is in favor of the higher value are called upward criteria. Conversely, the downward criteria encompass various costs.

Given a criteria, its utility could be calculated based on its utility function. And, the utility function of one criterion could be different from those of other criteria. Some common utility functions are listed in table 2-2. Note that it is important to select the suitable utility function for each criterion. The authors of [NV08] pointed out that sigmoidal utility functions are suitable for the network selection issue. And, a few modifications on the sigmoidal function is made to fit for further requirements.

Table 2- 2 Examples of several common utility functions.

Utility form	Generalized formula	Increasing & Differentiability	Concavity	Convexity
Linear piecewise	$u(x)=\begin{cases} 0 & x<x_{\min} \\ \dfrac{x-x_{\min}}{x_{\max}-x_{\min}} & x_{\min}\le x\le x_{\max} \\ 1 & otherwise \end{cases}$	Yes	No	No
Logarithm	$u(x)=\ln(x)$ or $u(x)=\ln(1+ax)$ $(a>0)$	Yes	Yes	No
Exponential 1	$u(x)=e^{(x-M)}\ \ (0\le x\le M)$	Yes	No	Yes
Exponential 2	$u(x)=1-e^{-ax}\ \ (a>0)$	Yes	Yes	No
Sigmoidal 1	$u(x)=\dfrac{1}{1+e^{\zeta(x_m-x)}}\ \ (\zeta,x_m>0)$	Yes	Yes	Yes
Sigmoidal 2	$u(x)=\dfrac{(x/x_m)^{\zeta}}{1+(x/x_m)^{\zeta}}\ \ (x_m>0,\zeta\ge 2)$	Yes	Yes	Yes

In the network selection issue, we should consider multiple criteria together, so the utilities of multiple criteria should be combined as a total utility. The authors of [NV08] pointed out that a valid form to combine these criteria together should satisfy the following requirements:

$$\frac{\partial U(\mathbf{x})}{\partial u_i}\ge 0 \tag{2-2}$$

$$sign(\frac{\partial U(\mathbf{x})}{\partial x_i})=sign(u_i'(x_i))$$

$$\lim_{u_i\to 0}U(\mathbf{x})=0\quad \forall i=1...n$$

$$\lim_{u_1,...,u_n\to 1}U(\mathbf{x})=1$$

where *U(***x***)* is the total utility of the criterion vector **x** and u_i is the utility of criterion *i*. Based on the requirements, the authors of [NV08] also pointed out the multiplicative exponent weighting (MEW) function is the one that satisfies all the requirements.

In the literature, there are many works on cost/utility function based network selection scheme. In [SW08] and [SW07], a cost-function-based network selection strategy was proposed, which considers from a system's perspective, and also considers a user's needs. Signal strength and network resource are specifically considered, so that the system could decide to whether accept or block an originating call / a handover call. Besides, the authors also provide a theoretical analysis method to evaluate the performance of the system, e.g. blocking probability, average received signal, etc.

In [WT08], a cost-aware handover decision algorithm was proposed for cooperative cellular relaying networks. Two cost functions, namely the triggering and priority decision cost functions are exploited, which involves the signal transmission quality, handover signaling cost, handover latency and interference estimation.

The triggering cost function is used in a triggering step, given by

$$f_1 = \alpha S - \beta P + \gamma T \ , \tag{2-3}$$

where S denotes the signaling cost induced by the handover, P denotes received power and T denotes HO latency time. α, β and γ are weights with a sum equals 1.

The priority decision cost function is used in a priority decision step, given by

$$f_2 = (1-\eta)SIR_{rel-cur} - \eta G \ , \tag{2-4}$$

where η is the weight factor which is changing with the different weight of the two parameters. G is defined as a resource release gain to describe the channel resource utilization difference before and after HO. $SIR_{rel-cur}$ denotes the relative SIR gain between the current link and the link with the highest SIR in the network candidate list.

In [OO05], a network selection scheme was proposed, which is mainly about to meet the users' data transfer requirements. Since the nature of radio links is unreliable, the scheme needs to predict the data rate of each available network and make the decision based on those predictions.

In [OO06] and [OO06-2], an intelligent utility-based network selection strategy was proposed, which focuses on the monetary aspects of various networks. A number of utility functions are examined which explore different user attributes to risk for money and delay preferences related to their current application.

Generally speaking, users will seek value for money and will have a patience limitation on their willingness-to-wait for mobile system response to their requests. In this strategy, users aim to maximize their positive consumer surplus while meeting their transfer completion time deadline. Consumer surplus (CS) is a term of microeconomic, representing the difference between the monetary value of the obtained data to the user and the actual price charged by the network operator.

The strategy is shown in figure 2-18. Tc1 is used to represent the user's expectation on the transfer time cost, and Tc2 to represent the maximum limitation. That is to say the user is not willing-to-pay if transfer time cost is larger than Tc2. All the networks will be checked one by one. If the cheapest network has a transfer time cost smaller than Tc1, it will be selected directly. Otherwise, all the networks with transfer time cost smaller than Tc2 will be added into an eligible network list. Then, the list will be ranked based on CS and the network with maximum CS will be selected. In a special case when all networks' transfer time costs are larger than Tc2, the cheapest network is selected.

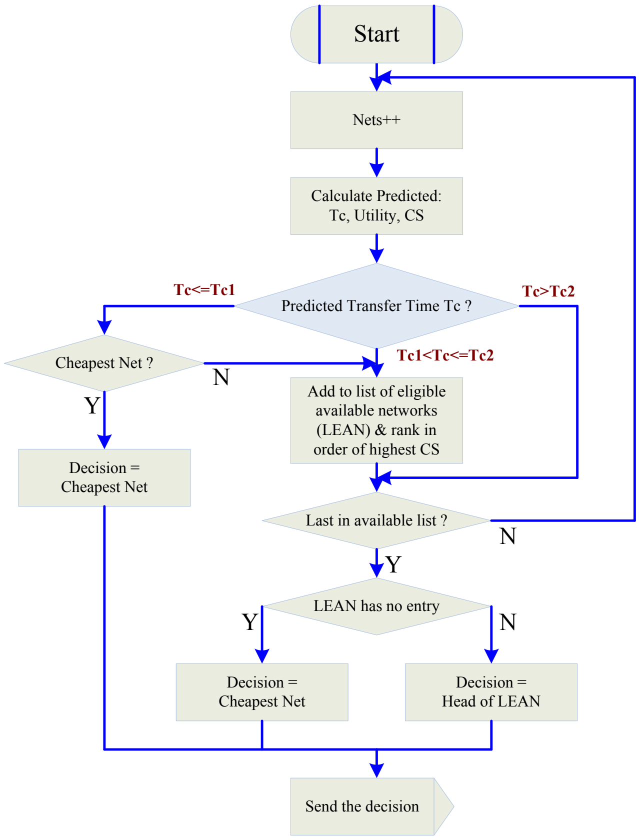


Figure 2- 18 A utility-based network selection strategy focusing on transfer time cost.

Moreover, in [RE06], the network selection issue was studied with the goal of getting the minimum RTT delay during communications.

Besides the transmitting time, another important issue is to reduce unnecessary handovers. Since network-side and terminal-side contexts change frequently, it is common that the best network changes a lot from time to time. If handover is performed at any time a new network is found better than the current network, it is possible that many unnecessary handovers are performed. In [CX07], three cost-function-based algorithms were proposed to reduce unnecessary handovers.

Algorithm 1: Handover is triggered when

$$F_{best} - F_{current} \geq \Delta \, , \qquad (2\text{-}5)$$

where $F_{current}$ is the current used network, F_{best} is the new best network, and Δ is a pre-defined threshold which should be carefully selected.

Algorithm 2: Handover is triggered when F_{best} is better than $F_{current}$ for at least D units of time. The key issue is to deicide the value of D, and the authors suggested that D should be at least larger than the handover latency.

Algorithm 3: Handover is triggered when the tendency of F_{best} is to be larger than $F_{current}$ particularly in the most recent instants. The tendency can be tracked with the help of an exponential moving average function.

2.3.2. MADM

Multiple attributes decision making (MADM) refers to making preference decision over the available alternatives that are characterized by multiple (usually conflicting) attributes. MADM is a branch of multiple criteria decision making (MCDM) which also includes multiple objectives decision making (MODM). MODM problems involve designing the best alternative given a set of conflicting objectives. For example, automobile manufacturers with to design a car that maximizes riding comfort and fuel economy and minimizes production cost. The alternatives are created by the design process. MADM problems are considered have several common characteristics [HY81]:

Alternatives: a finite number of alternatives are screened, prioritized, selected and/or ranked for making the final decision. The term 'alternative' is synonymous with 'option', 'policy', 'action' or 'candidate', etc.

Multiple Attributes: the DM should consider multiple attributes of these alternatives. The term 'attribute' can also be referred to as 'goal', 'criterion', 'property', characteristic', etc.

Incommensurable Units: Attributes have different units of measurement, so some adjustment (e.g. normalization) is required before combining them together.

Attribute Weights: Different DMs might focus on different aspects (i.e. attributes) when ranking the alternatives, so attribute weights must be used to represent their relative importance.

Decision Matrix: a MADM problem can be concisely expressed in a matrix format, where columns indicate attributes and rows indicate alternatives. Thus, a typical element $x(i, j)$ of the matrix indicates the value of the ith alternative, with respect to the jth attribute.

Normalization: different attributes have different measurement units, so normalization is treated as a necessary step of network selection. There are several methods of normalization, which are compared in table 2-3. In this table, N represents the number of networks, v_i represents the value of the ith criterion, and p_i represents its normalized value. The third method categorizes all the network-side criteria into three sub-groups, i.e. LB, SB and nominal-the-best (NB), so $NB(v_i)$ represents the nominal value of the ith criterion. The difference between the first and the third method is that the first one does not consider the NB group. For usages of these four normalization methods, please refer to [BF07-3] [BF07-2] [BB07] [OM07], respectively.

Table 2- 3 Examples of normalization operations.

$p_i = \frac{v_i - \min(v_i)}{\max(v_i) - \min(v_i)}$
$p_i = \frac{v_i}{\sqrt{\sum_{i=1}^{N} v_i^2}}$
$p_i = \frac{v_i}{\sum_{i=1}^{N} v_i}$
$p_i = \begin{cases} 1 - \frac{\lvert v_i - \max(v_i)\rvert}{\max(v_i) - \min(v_i)} & \text{for LB} \\ 1 - \frac{\lvert v_i - \min(v_i)\rvert}{\max(v_i) - \min(v_i)} & \text{for SB} \\ 1 - \frac{\lvert v_i - NB(v_i)\rvert}{\max\{\max(v_i) - NB(v_i), NB(v_i) - \min(v_i)\}} & \text{for NB} \end{cases}$

MADM algorithms can be divided into compensatory and non-compensatory ones. [BF07] described a comprehensive decision making process to rank candidate networks for service delivery to the terminal, which is based on a unique decision process that uses compensatory and non-compensatory multi-attribute decision making algorithms jointly to assist the terminal in selecting the top candidate network.

The network selection procedure was suggested as shown in figure 2-19. Non-compensatory algorithms (dominance, conjunctive, disjunctive or sequential elimination) are firstly used to separate the given alternatives into acceptable and unacceptable groups. Acceptable alternatives are those that satisfy the minimum cutoff criteria. Then, compensatory algorithms are used to select a network from the acceptable group.

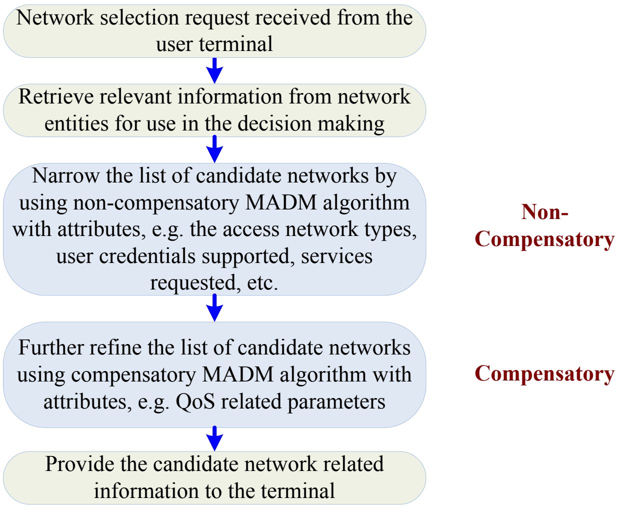


Figure 2- 19 Compensatory and non-compensatory integrated network selection.

MADM algorithms that have been used for network ranking include SAW, MEW, GRA, TOPSIS, ELECTRE, AHP, etc. The first four algorithms rank networks based on their coefficients (such as total costs or total utilities) calculated by combining adjusted values of all the criteria, while the last algorithm use pair-wise comparisons among all the networks, which is a totally different procedure.

In SAW and MEW, the coefficients are calculated separately by additive and multiplicative operations [SN06]:

$$C_{SAW} = \sum_{j=1}^{M} w_j v_{i,j} \text{, and} \tag{2-6}$$

$$C_{MEW} = \prod_{j=1}^{M} v_{i,j}^{w_j} \text{,} \tag{2-7}$$

where w_j represents the weight of the jth criterion, $v_{i,j}$ represents the adjusted value of the jth attribute of the ith network.

Equation (2-7) can be modified as

$$C^*_{MEW} = \ln(C_{MEW}) = \sum_{j=1}^{M} w_j \ln(v_{i,j}) \text{.} \tag{2-8}$$

Considering the characteristic of the natural logarithm, the attribute whose cost being close to 0 has large impact on the total cost than others. For example, Bluetooth is more often selected by MEW than by other algorithms due to its low monetary and power costs, as shown in our later simulations.

The application of TOPSIS to network selection was described in [BF07-2], which contains the following steps:

- normalizing of all the attributes;
- evaluating weights of all the attributes;
- defining the best and worst reference networks, which are formed by the best and worst values of each attribute, respectively;
- calculating the Euclidean distances of each candidate network to the best and worst reference networks;
- calculating the coefficient of TOPSIS as

$$C_{TOPSIS} = \frac{D_{w,i}}{D_{b,i} + D_{w,i}} \tag{2-9}$$

where $D_{w,i/b,i} = \sqrt{\sum_{j=1}^{M} w_j^2 (v_{i,j} - R_{w,j/b,j})^2}$.

- selecting the network with the largest C_TOPSIS.

The application of ELECTRE to network selection was described in [BF07-3], which contains the following steps:

- identifying attributes of different networks as a decision matrix;
- defining an ideal network;
- calculating the absolute difference between each network and the ideal network;
- normalizing the absolute difference;
- multiplying weights of attributes;
- calculating concordance and discordance matrices;
- making decision based on concordance and discordance matrices according to certain rules.

In step 6, concordance and discordance matrices are calculated based on *CSet* and *DSet*, which are obtained by comparing attributes of two networks. *CSet*(i, j) contains the attributes on which network i is better than network j, and *DSet*(i, j) is inverse. Therefore, we have

$$CSet(1, 2) + DSet(1, 2) = \{\text{all the attributes}\}. \tag{2-10}$$

Then, the elements in concordance and discordance matrices are calculated as follows:

$$c_{kl} = \sum_{j \in CSet_{kl}} W_j \text{ , and} \tag{2-11}$$

$$d_{kl} = \frac{\sum_{j \in DSet_{kl}} \left| (\mathbf{NW}_{norm})_{kj} - (\mathbf{NW}_{norm})_{lj} \right|}{\sum_{j} \left| (\mathbf{NW}_{norm})_{kj} - (\mathbf{NW}_{norm})_{lj} \right|} . \tag{2-12}$$

Finally, the best network will be decided based on the above matrices. The authors also proposed two approaches to rank the networks based on the concordance and discordance matrices.

Network selection schemes combining AHP and GRA was proposed in [SQ05], [SQ05-2], [CD08] and [OM07]. As shown in figure 2-20, AHP is to decide the relative weights of attributes according to various kinds of subjective information, while GRA is to rank the network alternatives by combining both the values and weights of multiple attributes.

For more details, AHP is carried out with the following steps:

- Structuring the weighting issue as a decision hierarchy of all the criteria. A hierarchy used by [SQ05] is shown in figure 2-21.
- Comparing criteria pair-wise on each level in the hierarchy to obtain several matrices of relative priorities. For example, criteria on Level-1 will be pair-wise compared to get a 5*5 matrix; while criteria on Level-2 will be pair-wise compared to get two 3*3 matrices.
- Calculating the weights of criteria on each level as the eigenvector of each matrix.
- Synthesizing the above results as an overall vector of weights of all the criteria. That means eigenvectors of these sub-matrices should be multiplied by their parent weights to obtain their overall weights.

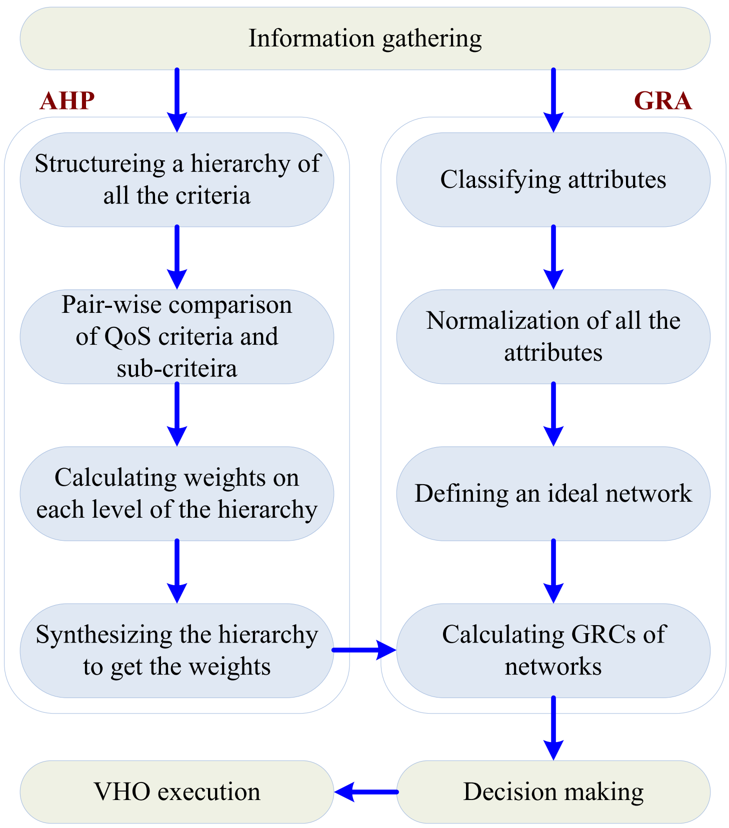


Figure 2- 20 Combined procedure of AHP and GRA.

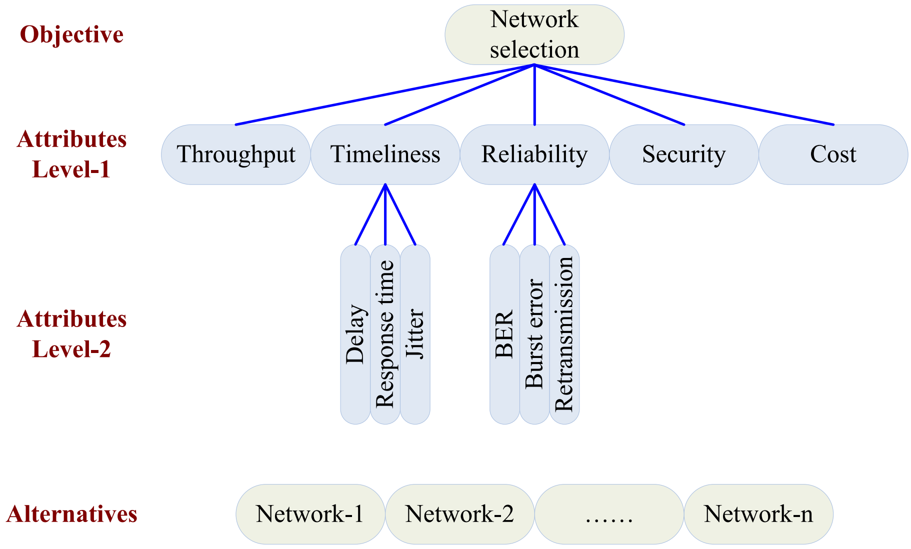


Figure 2- 21 An example of an AHP hierarchy of criteria.

One of the key characteristics of AHP is the subjectivity of those pair-wise comparisons, that's why adjustment is required when the consistency ratio (CR) of the matrix of overall priorities is too large (e.g. >10%).

GRA is one of the popular MADM algorithms, which is based on comparisons of grey relationships between elements of two series. Every alternative will be compared with a pre-defined ideal reference alternative to get its preference. The procedure is as follows:

- classifying the attributes;
- defining the lower, upper and moderate bounds of each attribute;
- normalizing attributes;
- defining an ideal network with best values for all the attributes;
- calculating GRCs (grey relational coefficients) of networks by comparing with the ideal network, given by

$$C_{GRA} = \frac{1}{\sum_{j=1}^{M} w_j \left| v_{i,j} - R_j \right| + 1}, \tag{2-13}$$

  where R_j represents the ideal value of the jth criterion.

- the network with the maximum GRC is the best network.

Note: If we firstly inverse all the 'larger-the-better' criteria into 'smaller-the-better' ones, the operation of calculating the absolute value in the above equation is eliminated. Thus, we can see that GRA should have similar performance with SAW.

[SN06] provided a comparison of the performance between four vertical handover decision algorithms, i.e. SAW, MEW, TOPSIS and GRA. Results show that SAW, MEW and TOPSIS have similar performance to all four traffic classes, while GRA provides a slightly higher bandwidth and lower delay for interactive and background traffic classes.

Generally, the utility of an attribute is monotonically increasing or decreasing, but it is possible to define the utility of certain attribute to be non-monotonic. Take the attribute 'delay' as an example, the network with minimum delay may not have the maximum utility. Instead, the network with a large delay could have a large utility for web browsing applications. [BF07-4] studied the scenarios when some attributes have non-monotonic utilities and argued that GRA is more suitable than TOPSIS or ELECTRE in these scenarios.

2.3.3. Fuzzy logic

Fuzzy logic theory is an important mathematic model used for network selection. In the literature, there are many studies on combining fuzzy logic with MADM algorithms for network selection. Here, we present some of the key contributions.

In [CP02] and [CP01], a segment selection scheme based on fuzzy multiple objective decision making algorithm is presented. The use of fuzzy logic allows a sensitivity analysis to be performed in order to determine which factors are critical to the efficient utilization of the network. Two specific groups of attributes are given particular attention: charging model and quality of service.

A fuzzifier is used to adjust the attributes before combining them together. For different attributes, the fuzzifier could be different. An example of a membership function of the fuzzifier is given in figure 2-22.

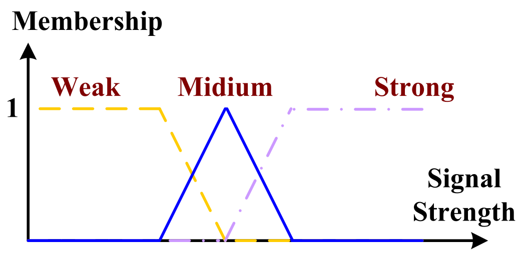


Figure 2- 22 Membership function of 'signal strength'.

In [HJ06], a novel fuzzy logic based decision making algorithm was proposed, which is capable of combining the merits of both immediate VHO and dwell VHO to achieve excellent handover in terms of packet transfer delay. The FL is used to handle the exploited uncertain and conflicting decision metrics.

Three input fuzzy variables are identified: the probability of a short interruption, the failure probability of handover to radio, and the size of unsent messages. Two input fuzzy sets are defined for the first and the second fuzzy variable, i.e. high and low. And, two input fuzzy sets are defined for the third fuzzy variable, i.e. large and small.

Figure 2-23 shows the procedure of the proposed fuzzy logic based VHO decision. A singleton fuzzifier and a largest-of-maximum defuzzifier are employed. In the fuzzy inference engine, an algebraic product operation is used to fuzzy implication. At the beginning, the fuzzy variables are fuzzified and converted into fuzzy sets. Then, the fuzzy inference engine maps the input fuzzy sets into output fuzzy sets. Finally, the out fuzzy sets are defuzzified into a crisp decision point, which has two waiting-time options corresponding to the two VHO algorithms.

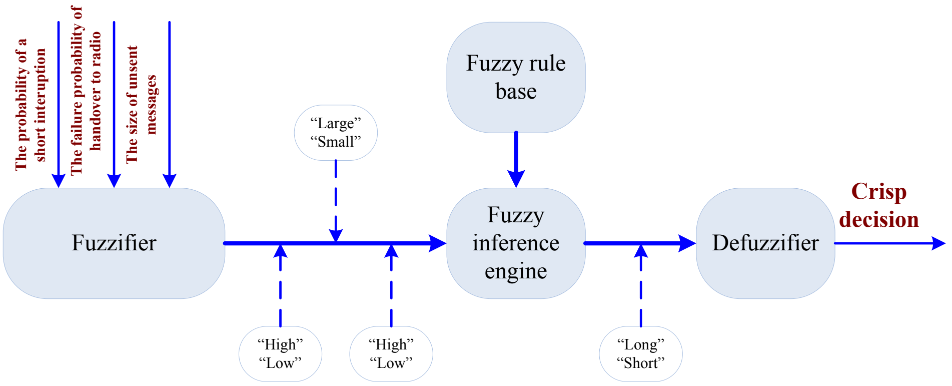


Figure 2- 23 The procedure of fuzzy logic based VHO decision.

In [AM07], a combined fuzzy logic control and MCDM scheme for network selection was proposed. The first step of the scheme is to use small fuzzy logic based subsystems for representing input criteria. Fuzzy logic is used to overcome the complexity and fuzziness associated with the heterogeneous wireless environments and their services. The second step uses a MCDM technique that takes the first layer's output as its input. MCDM ensures that all the different characteristics and view points are taken into account for the decision. The details of the scheme are shown in figure 2-24. For each criteria, fuzzy logic based subsystem is performed based on a group of fuzzy rules composed of IF/THEN statements.

During the gathering process of network selection information, it is possible that some information is imprecise due to, for example, the frequent change of some dynamic parameters. Classic MADM schemes cannot make the decision based on these attributes, so there are some contributions on combining fuzzy logic with MADM to solve this problem.

In [ZW04], a fuzzy MADM network selection scheme was proposed, in which fuzzy logic is applied to deal with the imprecise information of some criteria and user preference. According to the data type of the alternative's performance, fuzzy MADM scheme can be categorized into three groups: data are all fuzzy, all crisp, and either fuzzy or crisp. [ZW04] suggested that the scheme in [CS92] will be used to convert imprecise linguistic terms to crisp numbers, while SAW or TOPSIS will be applied for the final ranking.

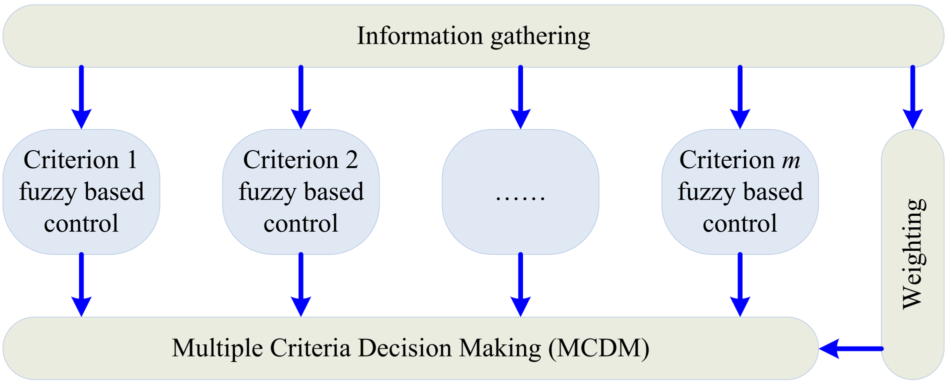


Figure 2- 24 Fuzzy MCDM network selection scheme.

[BF07-5] provided a comprehensive approach towards network selection mechanism that leverages parameter estimation techniques, fuzzy theory, MADM algorithms, and so on. The main idea is to estimate missing data with forecasting techniques before MADM algorithm is performed, as shown in figure 2-25.

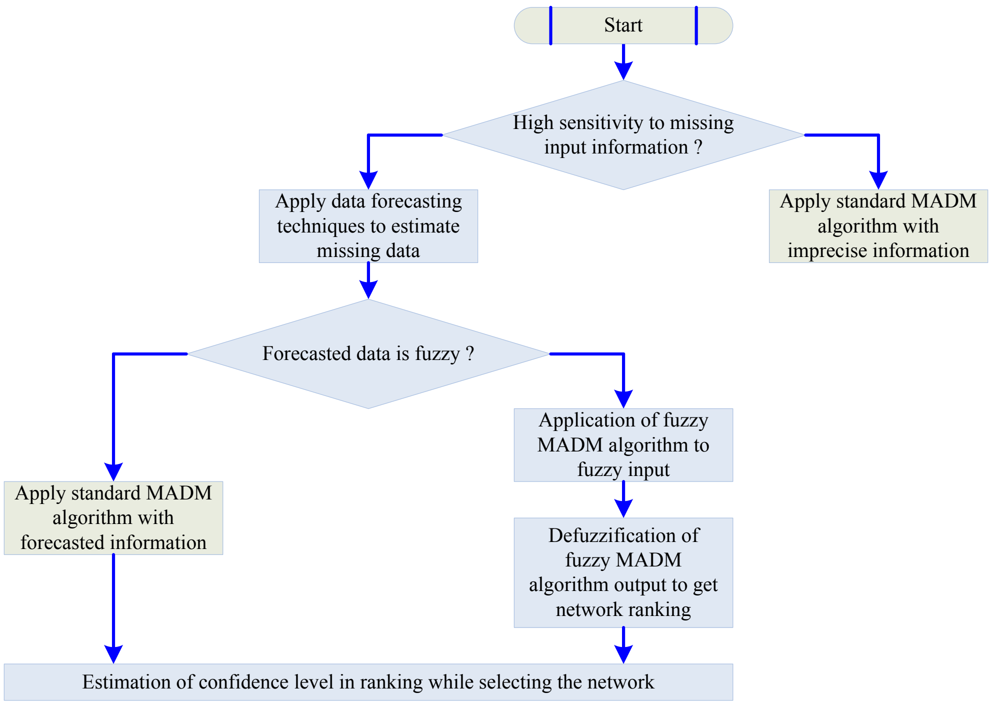


Figure 2- 25 Fuzzy MADM with imprecise information.

In [KS05], a network selection scheme combining network-specific and user specific in modern peer-to-peer systems was proposed, as shown in figure 2-26. The network operator provides network attributes to the fuzzy modules. The fuzzy module composes of two schemes that rank the network as per service, or as per user demand. The fuzzy decision system in each of the schemes fits a non-linear function to the data set in hand, and derives a rank for the network. Thus, the networks are ranked as *Rn*. Then, this rank is used as input for the user-specific ranking scheme, and a new rank *Ru* is generated to fit a user's particular requirements.

Since some dynamic factors change frequently, it might be not sufficient and not timely to detect the current state of these factors. Therefore, to make a better decision, some recursion is used to combine the current detected states with the previous ones or the previous ranking results, e.g. the combination of rank (*Ru*) with user-specific information in figure 2-26. Another idea is to use neural network for learning of some specific parameters, as shown in figure 2-27.

[GQ05] combined fuzzy logic technique with a modified neural network for network selection. Fuzzy inference system is adopted for adjusting the crucial criteria as the input variables and makes the decision based on a defined rule set. Elman neural network is used to predict the number of users using certain network after the selection.

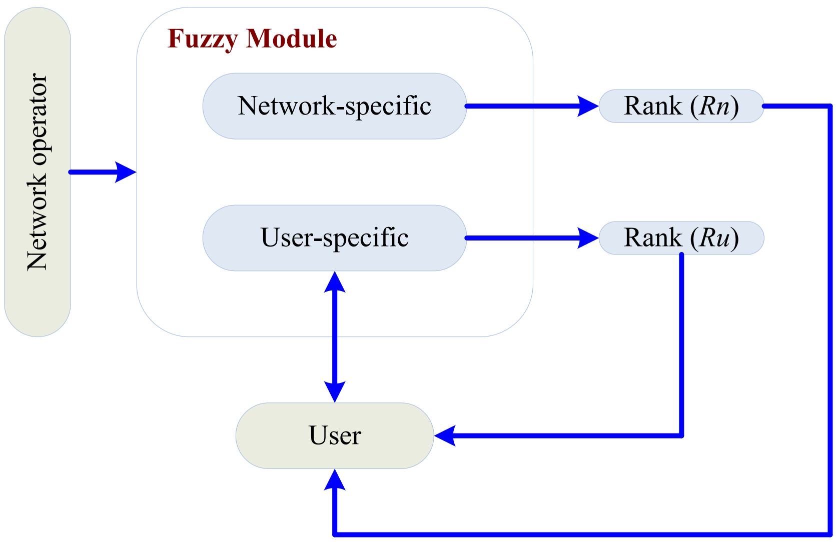


Figure 2- 26 Network specific and user specific combined network selection.

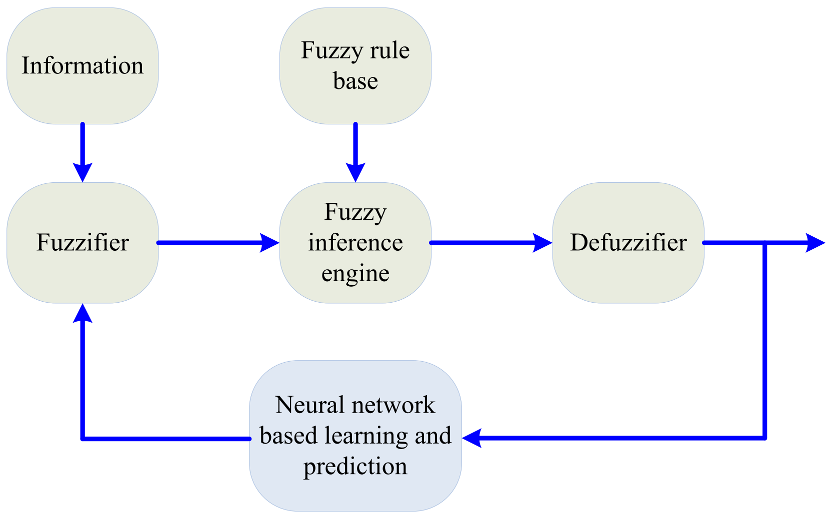


Figure 2- 27 Network selection combining fuzzy logic and neural network.

[SH07] also proposed a network selection scheme combining fuzzy logic and neural network. The functionalities of the two techniques in the scheme are similar, but the attributes used in the neural network module for learning different.

Instead of neural network, [VB07] used online kernel learning to dynamically adapt tradeoffs between the various attributes. This combines the maximum margin idea from kernel methods such as support vector machines with stochastic approximation for online optimization. The procedure of the proposed scheme is given in the figure below.

2.3.4. Markovian model

Markov decision process is a common mathematic model for decision making. In the literature of network selection, many studies using Markov decision theory have been proposed. Here, we present several most important proposals on Markov decision theory based network selection.

In [SN08] and [SN07], the vertical handover decision issue was modeled as a Markov decision process, in which the objective is to maximize the total expected reward per connection. The network resources that are utilized by the connection are captured by a link reward function, a signaling cost is used to model the signaling and processing load incurred on the network when vertical handover is performed, and the value iteration algorithm is used to compute a stationary deterministic policy. Besides, this model could take into account the connection duration of various networks, which is an important feature for making the vertical handover decision.

An obvious advantage of this model is the integration of all the above parameters together. In other words, the decision could be made based on just one final decision function which considers all the above parameters.

But, according to my understanding, there might be some difficulty in its implementation and testing, because any adjustment of the model will lead to different final decision function. That is not convenient for an engineer to derive again and again different final decision functions.

In [SC08], the vertical handover decision issue was further modeled as a constrained Markov decision process (CMDP). Their objective is to maximize the expected total reward of a connection subject to the expected total access cost constraint.

A benefit function is used to assess the quality of the connection, and a penalty function is used to model signaling and call dropping.

According to the authors' evaluations, this algorithm outperforms the MDP algorithm in [SN08], thanks to the usage of user's velocity information. Detailedly, an MT's velocity is assumed to be correlated in time and can be modeled by a discrete Gauss-Markov random process. The following recursive realization is used to calculate the transition probability of the MT's velocity:

$$v' = \alpha v + (1-\alpha)\mu + \sigma\sqrt{1-\alpha^2}\phi, \qquad (2\text{-}14)$$

where v is the MT's velocity at the current decision epoch, v' is the MT's velocity at the next decision epoch, α is the memory level, μ and σ are the mean and standard deviation of v, and ϕ is an uncorrelated Gaussian process with zero mean and unit variance which is independent of v.

In order to combine multiple attributes together, [WY08] proposed two network selection approaches based on rank aggregation and using weighted Markov chain (WMC).

The procedure of the proposed network selection schemes are as follows:

- based on each attribute, a ranking list of all the networks is obtained as its corresponding ordering;
- construct the transition matrix of weighted Markov chain, given by

$$\mathbf{MC} = \begin{pmatrix} 0 & 0 & . & 0 \\ 0 & . & . & 0 \\ 0 & 0 & . & . \\ 0 & 0 & . & 0 \end{pmatrix}_{N\times N,} \qquad (2\text{-}15)$$

  where $mc(i, j)$ represents the transition probability from network i to network j;

- for each attribute and its corresponding ordering, the transition matrix **MC** is updated as follows (this step is repeated for all attributes):

  - WMC scheme 1:

    if network j is not preferred to network i , we have

$$mc_{ij} = mc_{ij} + \frac{w_q}{\tau_q(p_i)} \tag{2-16}$$

    where w_q is the weight of attribute *q* and $\tau_q(p_i)$ denotes the position the network *i* in the corresponding ordering of attribute *q*.

  - WMC scheme 2:

    if network i is preferred to network j, we have

$$mc_{ij} = mc_{ij} + \frac{w_q}{N} \tag{2-17}$$

    else if network *j* and network *i* have the same performance, we have

$$mc_{ij} = mc_{ij} + \frac{N - \tau_q(p_i) + 1}{N} w_q \tag{2-18}$$

    where *N* is the number of available network.

- compute the stationary distribution vector **SD** = {sd_1, sd_2, … sd_n} as

$$\mathbf{SD} = \mathbf{SD} \times \mathbf{MC} \tag{2-19}$$

  where sd_n is the element for network *n*.

- the network with maximum sd_n value is the best one.

Besides, [SQ08] modeled the network selection issue as a semi-Markov decision process.

2.3.5. Game theory

Game theory is related to the actions of decision makers who are conscious that their actions affect each other. The essential elements of a game are players, actions, payoffs, information, etc. These elements are explained as follows:

- Players are the individuals who make the decision. Each player's goal is to maximize his own utility by a choice of actions.

- An action is a choice of a player as his one-round strategy in the game. For a certain player, he must have an action combination as his strategies during the whole game.
- Payoff means the utility that a player can receive by taking certain action while all the other players' actions are decided. Therefore, the payoff of one action of one player could change if other players' actions have any change.
- For each player, there should be a strategy set which contains all the strategies he might choose. In each round, the player chooses one strategy from the set.
- Information, including that of the player himself and that of other players, is important in a game.
- An equilibrium is a strategy profile consisting of a best strategy for each of the players in the game. The equilibrium strategies are those which lead to the maximum payoffs. A Nash equilibrium is a solution of a game, in which no player has any more payoff to gain by changing only his own strategies.

Game theory techniques can be easily adapted for use in resource management mechanisms in a heterogeneous wireless network environment. [JA07] provided a gaming model for network selection by defining a game between access networks, given by

$$G = (N, R, S(i), V(i, j)), \qquad (2\text{-}20)$$

where N denotes the set of players, i.e. the set of access networks;

R denotes the set of game resources, i.e. the set of service requests;

$S(i)$ denotes the set of strategies for player (i); and

$V(i, j)$ denotes the payoff related to player (i) when choosing resource (j).

The aim of this game is to maximize the total payoff by choosing different resources for different players. To solve the game, several rounds should be taken. Taking the first round shown in table 2-4 as an example, the best choice is the left corner with 6 + 6 payoffs. Therefore, in the first round, player (1) and player (2) will select resource 6 and resource 1, respectively.

This model has at least the following difficulties:

- payoff values are not easy to be defined. This is one of the most important parts in the game model;
- service and network information is usually dynamic, which increases the difficulty to solve the game.

Besides, [CD08] also proposed to model the access admission control as a non-cooperative game. In this sense, networks play against each other so as to maximize their payoffs and admission control policy ensures maximum QoS for all service requests.

Table 2- 4 The first round of an example of a game between networks.

	Player 2						
	Strategy	1	2	3	4	5	6
	1		1,5	1,4	1,3	1,2	1,1
	2	2,6		2,4	2,3	2,2	2,1
Player 1	3	3,6	3,5		3,3	3,2	3,1
	4	4,6	4,5	4,4		4,2	4,1
	5	5,6	5,5	5,4	5,3		5,1
	6	6,6	6,5	6,4	6,3	6,2	

2.3.6. NP hard

2.3.6.1 Knapsack

Knapsack problems are a family of optimization problems that require a subset of some given items to be chosen so that the corresponding profit sum is maximized without exceeding the capacity of the knapsack(s).

A typical knapsack problem with a single knapsack and N items is formulated as follows:

$$\text{Maximize } Z = \sum_{i=1}^{N} p_i x_i \ , \ s.t. \sum_{i=1}^{N} w_i x_i \le C \ , \qquad (2\text{-}21)$$

where Z is the total profit, p_i is the profit of item i, w_i is the capacity cost of item i, $x_i \in [0,1]$ is the fraction of item i placed in the knapsack, and C is the capacity of the knapsack.

However, this single knapsack model does not fit for the network selection issue. A more general model that fits network selection issue is a combine of the 0 – 1 knapsack model and the multiple choice multiple dimension (MMKP) model [GV05]. This generalized model, multiple knapsacks are used and one item can only be put into one knapsack, as shown below:

$$\text{Maximize } Z = \sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij}x_{ij}\ ,\ s.t.\sum_{i=1}^{N} w_{ij}x_{ij} \le C_j\ , \qquad \text{(2-22)}$$

where Z is still the total profit, p_{ij} is the profit of item i placed in knapsack j, w_{ij} is the capacity cost of item i placed in knapsack j, $x_{ij} \in \{0,1\}$ is the placement of item i in the knapsack j when it equals '1', and C_j is the capacity of the knapsack.

To model the network selection issue,

- the traffic flows are mapped to items,
- networks are mapped to knapsacks,
- the user utility is mapped to the profit,
- the resource constraints of networks are mapped to capacities of knapsacks,
- QoS profiles of traffic flows in given network are mapped to profits of items in given knapsack, and
- costs of traffic flows for a given network's resource are mapped to costs of items for a given knapsack's capacity.

It is worth mentioning that

- this knapsack model is NP-complex;
- this model is trying to maximize one objective, e.g. the user utility, not multiple objectives. By contrast, network selection issue usually has multiple objectives;
- this model fits for the case when networks' capacities are limited or load balancing are strongly demanded, but the network selection issue is generally making a decision for a coming traffic flow among multiple available networks;
- the weights are quite difficult to be defined;
- when the capacity of networks is enough for a coming traffic flow, this model becomes a simple additive weighted model.

2.3.6.2 Bin packing

The classical bin packing problem is a well studied optimization problem: Given n objects with sizes $a_1, \ldots, a_n$ belongs to (0, 1], find a packing in unit-sized bins that minimizes the number of bins used. In the off-line version of this problem, it is possible to consider all the objects and choose the order of assignment. In the online version however, each object must be assigned in turn, without knowledge of the next objects. That is, given $n - 1$ already packed objects with sizes $a_1, \ldots, a_n$ belongs to (0, 1], the new object *n with size* a_n belongs to (0, 1] must be packed in such a manner that the number of used bins is minimized. Finding an optimal packing is known as an NP-hard problem.

Network selection can be formulated as the problem of finding the best way of allocating applications in networks in order to minimize the number of rejected applications, i.e. the blocking probability, and maximize whole system's capacity. In [MD06], the authors mapped the problem of network selection into the bin packing problem, where objects are applications arriving and bins are networks where these applications should be packed.

Five algorithms for the online bin packing problem were compared: FirstFit, BestFit, WorstFit, LessVoice and Random.

FirstFit: a network is randomly selected with equal probability among all the networks. The application is allocated to the selected network if there is enough space for this application. Otherwise, a next network will be randomly selected until a network with enough space is found. If no network has enough space, the application is rejected.

BestFit: a network is selected if there is enough space available for the application and if by allocating the applications, there will be less free space left in that network compared with others.

WorstFit: a network is selected if there is enough space available for the application and if by allocating the application, there will be more free space left in that network compared with others.

LessVoice: the network with optimal $r(s)$ will be selected, given as follows

$$r(s) = Size(a,s) / Size(voice,s)\,, \qquad (2\text{-}23)$$

where $Size(x, s)$ represents the capacity of application x in network s.

Random: a network is randomly selected with equal probability among all the networks. The application is allocated to this network if there is enough space, rejected otherwise. Compared with FirstFit, this algorithm will not select other networks when the first selected network has no space.

2.3.7. Power saving

Power saving is an important issue during network selection. Multiple interfaces are available on the terminal, but usually only one or several of them are used at any moment. Since some interfaces (e.g. WiFi) cost a lot of power, it is a waste of power to let the unused interfaces active. In the literature, there are many proposals on network selection with power saving feature.

[IJ07] proposed a network selection scheme considering power saving, which is achieved by predicting the expected lifetime of the mobile terminal regarding to the current battery level, traffic class and power consumption capabilities of various networks.

[NV07] and [NV08-2] proposed a terminal-controlled network selection scheme, which is user-centric, power-saving, cost-aware and performance-aware. Power-saving is achieved by a policy-based state manager of available interfaces, as shown in figure 2-28. The idea is to turn off the non-cellular interfaces when they are not used, because non-cellular interfaces usually consume a large portion of power.

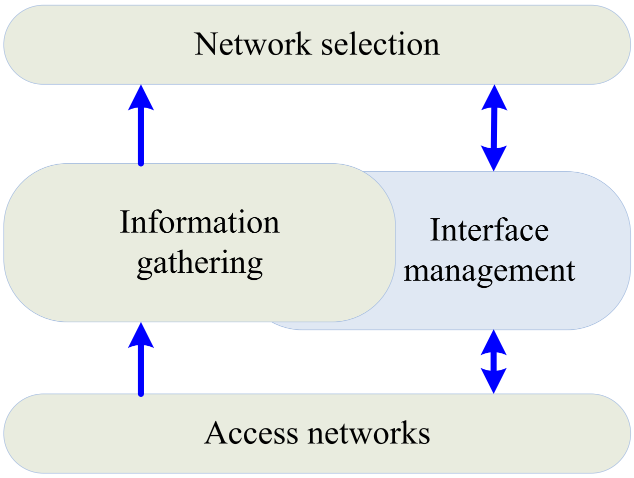


Figure 2- 28 Terminal-controlled network selection scheme.

Besides, [JH07] proposed a distributed uplink call admission control (CAC) and network selection scheme for hybrid CDMA/OFDMA networks. When the incoming call can be admitted by both of the two access networks, the one requiring less transmit power from the MT is selected as the target network for power saving purpose.

With power saving feature, it is quite possible that the interface of the newly selected best network should be activated after being selected. [WL07] suggested wait for a short period of time after the best network is decided, so that some necessary processes (e.g. channel searching and awakening) could be achieved.

2.3.8. User/operator combined strategy

Most of the schemes are user-centric, that is, they allow mobile users to decide when to handover and which network to attach based on users' preferences and perceived QoS. However, these schemes cannot completely guarantee successful handovers because network operators have the right to reject (or redirect from the preferred network to another one) any handover in order to maximize long-term revenue or save resources for the first-class users. Therefore, some proposals in the literature studied a tradeoff of benefits between operators and customers.

[AG05] considered a negotiation between network capabilities/availabilities and service requirements to decide whether a selected network can accept handover request.

In [JH06], a low-complexity, centralized network selection scheming was proposed, aiming to optimally distribute the end users to the networks of the heterogeneous wireless system, in the sense of maximizing the global spectrum efficiency. Two suboptimal algorithms, i.e. absolute bandwidth request (ABR) based algorithm and relative bandwidth request (RBR) based algorithm, were proposed.

[SS07] proposed a network selection process from the service provider's point of view. Utility functions for different services are defined, the competing service providers are defined as players in a game model, and the goal of the network selection is to maximize the completing service providers' revenue, i.e. achieving equilibrium.

In [SQ06] and [SQ08], an efficient QoS negotiation-based scheme was developed to balance between user satisfaction and network efficiency. The negotiation and decision step is formulated as a semi-Markov decision process. The considered information includes not only user preferences, network conditions, application requirements in terms of QoS parameters, but also network operator's interests, e.g. service policy for different classes of users and the plan for long-term revenue.

After all the candidate network list are sorted, the user will check with the network operators from the top of the list to see which network would like to accept the handover request and allocate the required resource to him. The operators accept handover requests selectively in order to maximize resource utilization and long-term revenue. Thus, users have the right to decide the rank of network, while operators have the right to accept, redirect or reject users' handover requests.

2.3.9. Integrated solution

Different mathematical models have different functionalities, so it is possible to combine multiple of above models as an integrated solution for the network selection issue. Context-aware network selection schemes described in subsection 2.4.4 and fuzzy MADM described in subsection 2.5.2 are examples of integrated solutions. It is common to combine context-aware information gathering with fuzzy logic and MADM as an integrated network selection solution.

After providing an overview of the most interesting and recent network selection schemes, classifying them into categories, and comparing their main features, [KM08] and [KM08-2] proposed a synthetic strategy that takes advantage of several interesting solutions.

The procedure of the strategy is given in figure 2-29. It contains mainly three parts: information gathering, handover decision and handover execution.

- Information gathering part collects (through monitoring and measurements) all the required context information for network selection.
- Handover decision part contains two steps. The first step is handover initiation performed by a Fuzzy Logic System, which decides if a handover is needed at any point of time or not. The second step is network selection, which decides the best target access network. This step includes criteria scoring, network scoring and decision making.
- After the best network is selected, Handover Execution is used to establish IP connectivity based on Mobile IP functionality.

It is worth mentioning that this strategy stores handover policies in a pot (called Handover Policies Repository), and these policies (e.g. user preferences and operator policies) express decision rules that govern choices of the whole process.

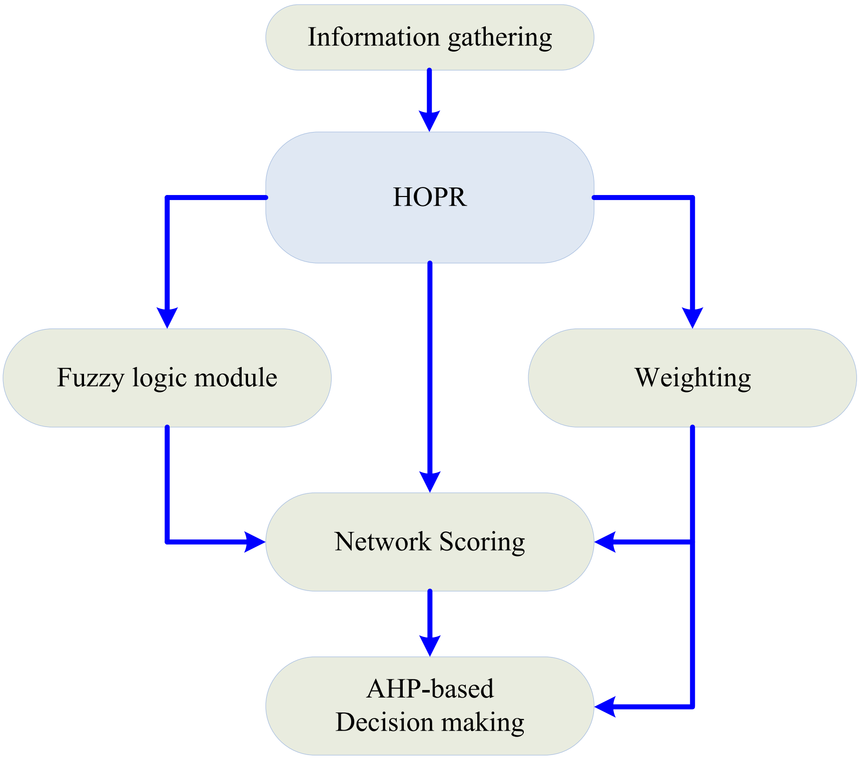


Figure 2- 29 A synthetic strategy of network selection.

3. MADM-BASED NETWORK SELECTION

Network selection has been studied a lot in recent years. Different mathematical tools have been used for this issue, as described in chapter 2, focusing on different aspects. In this dissertation, I choose MADM theory as the core of network selection scheme, for the following reasons:

- network selection issue is exactly to make a decision based on multiple attributes (as MADM);
- network selection scheme should make a decision in a few milliseconds, MADM fits this requirement. By contrast, those which model network selection into a NP-complex problem are definitely out of account;
- Game theory, knapsack model and bin packing model consider mainly the capacities of networks, so they fit well for the case when congestion is serious. My study is looking for a generic scheme that fits in all cases, so they are out of account;
- Markovian decision process is suitable for network selection issue, but this model also combines multiple attributes based on their weights, so it is just another form of MADM. Plus, Markov chain can be used to describe terminal mobility among HWNs, but it is not used for making a decision on which network is the best;
- Fuzzy logic and utility theory are usually combined with MADM in related works, and MADM was always the core theory in those combined schemes;
- Power saving is important, but not the only purpose of network selection. It could be considered together with other objectives in MADM. Both user's and operator's benefits should be considered and compromised while doing network selection, MADM fits for compromising multiple entities' benefits.

To sum up, MADM is selected as the core of the designed network selection scheme in this dissertation. For more information on the above theories and better understand my choice on MADM, please refer to section 2.3.2.

3.1. Configuration of simulator

To do extensive simulations, we established a network selection simulator which is configured as follows (for the details of this simulator, please refer to Appendix A):

Criteria: numerous network attributes are used as decision making criteria together, e.g. monetary cost (MC), bandwidth (BD), power consumption (PC), security level (SL), traffic load (TL), signal strength (SS), bit error rate (BER), jitter (JT), cell radius (CR), coverage percentage (CP), etc.

Requirements: two terminal properties (i.e. power condition and velocity) and four QoS levels (i.e. conversational, streaming, interactive and background) are considered. Besides, the customer prefers low monetary cost and good signal strength; while the operator wants load balancing to avoid congestion in the best network.

Networks: the HWNs is composed of WPAN, WLAN, WMAN and WWAN.

Weighting: weights of different criteria are calculated by the eigenvector method (explained in section 5.3) based on 9*9 pair-wise comparison matrices, the calculated weights are shown in table 3-1.

Adjusting: all the attributes' values are normalized and further adjusted through sigmoidal utility function.

Network ranking: five MADM algorithms, as explained above, are considered.

Table 3- 1 Weights of various scenarios (%).

		MC	BD	PC	SL	TL	SS	BER	JT	CR	CP
Low speed good power	Con	36	2	6	4	10	10	2	22	4	4
	Str	29	29	3	3	7	7	2	14	3	3
	Int	32	2	4	32	10	10	2	2	3	3
	Bac	36	3	4	11	11	11	11	3	5	5
High speed good power	Con	22	2	3	3	6	6	2	10	23	23
	Str	20	16	2	2	5	5	2	8	20	20
	Int	20	2	2	20	6	6	2	2	20	20
	Bac	23	2	3	6	6	6	6	2	23	23
Low speed poor power	Con	27	2	27	4	8	8	2	14	4	4
	Str	22	22	22	3	6	6	2	11	3	3
	Int	24	2	24	24	7	7	3	3	3	3
	Bac	28	2	28	8	8	8	8	2	4	4
High speed poor power	Con	19	2	19	2	5	5	2	8	19	19
	Str	17	14	17	2	4	4	1	7	17	17
	Int	17	1	18	18	5	5	1	1	17	17
	Bac	19	2	19	5	5	5	5	2	19	19

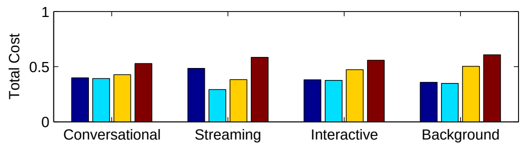
(Con: Conversational; Str: Streaming; Int: Interactive; Bac: Background)

3.2. Effects of terminal-side requirements

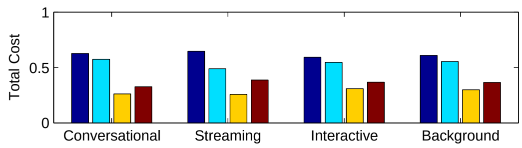
In MADM-based network selection schemes, terminal properties and application QoS requirements could affect the weights of network attributes, hence affect the selection result. However, we need to know how the selection result is affected and how much it is affected by the change of terminal properties or applications. In this sub-section, we study the impacts of terminal properties and application QoS requirements on network selection results by simulation. Weights affected by various requirements are listed in table 3-2, where left-hand requirements can result in high weights of right-hand criteria.

Table 3- 2 High weights corresponding to subjective requirements.

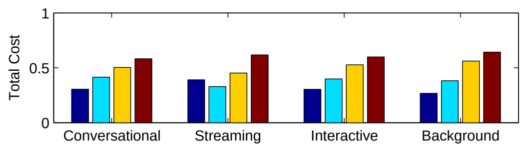
Subjective requirements		Criteria with high weights
low price	→	monetary cost
high speed	→	mobility-related criteria
poor power condition	→	power consumption
conversational applications	→	jitter
streaming applications	→	bandwidth and jitter
interactive applications	→	security level
background applications	→	BER and security level



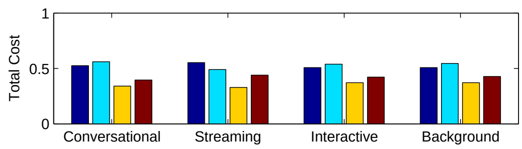
(a) Low-speed MT with good power condition



(b) High-speed MT with good power condition



(c) Low-speed MT with poor power condition



(d) High-speed MT with poor power condition

Figure 3- 1 Total costs for different terminal properties and applications (from left to right: WPAN, WLAN, WMAN and WWAN).

Total costs of four networks in these scenarios are shown in figure 3-1. When terminal properties and applications change, high weights are used for corresponding criteria, hence total costs of networks change and different networks are selected in different scenarios.

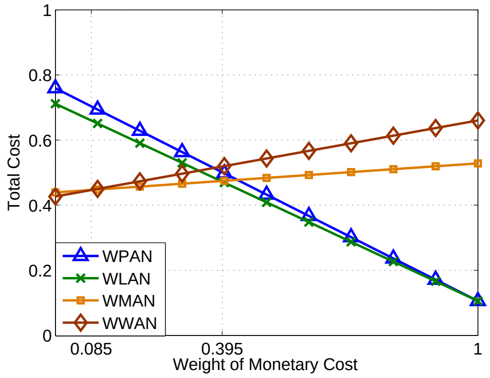
3.3. Coefficients of various MADM algorithms

MADM algorithms calculate coefficients of networks, so that the best network can be selected based on these coefficients. However, we wonder how these coefficients change and how their change affects the selection result. In this sub-section, we simulate four MADM algorithms (SAW, MEW, TOPSIS and GRA), and show their coefficients' changes with respect to certain criterion's weight, as shown in figure 3-2. ELECTRE is an algorithm that uses pair-wise comparisons between different networks, so it is not considered in this simulation.

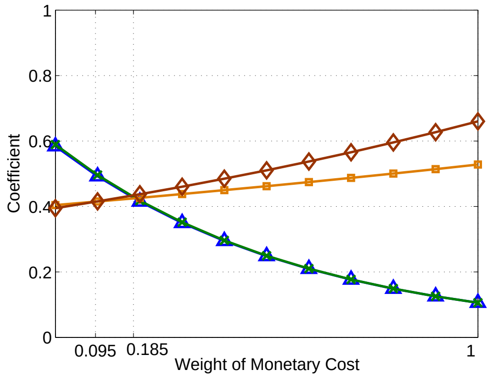
For SAW and MEW, the best network is the one with minimum coefficient; while for TOPSIS and GRA, it is the one with maximum coefficient, see section 2.3.2.

We can see from figure 3-2 that, some networks have similar performance, while some others are totally different. In most cases, several high-performed networks' coefficients are close, which means there is little difference by selecting any of them. This feature provides us the following information:

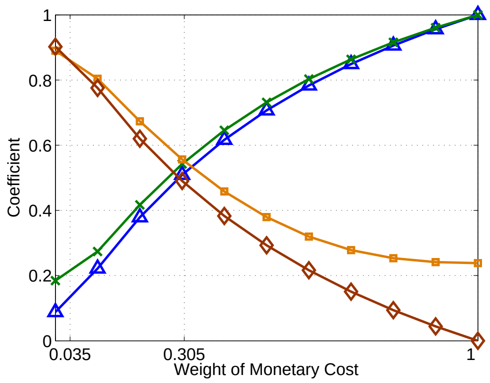
- VHO tradeoff is important, otherwise a customer might handover frequently between two networks with similar performance;
- load balancing is important, otherwise all the customers in an area might select the same network and ignore those networks with similar high performance. For example, WWAN is only a little bit better than WMAN in figure 3-2(a) when w>0.395, but all the customers will use WWAN (leading to congestion) and ignore WMAN if load balancing is not considered;
- due to normalization and sigmoidal utility function, some networks' coefficients increase, while some decrease. This feature fits for most of the criteria (e.g. mobility-related criteria in [WL09-3]), so it is easy to distinguish between good and poor networks and classify them into different groups.



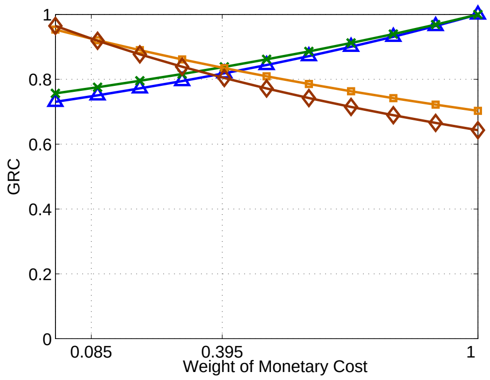
(a) SAW



(b) MEW



(c) TOPSIS



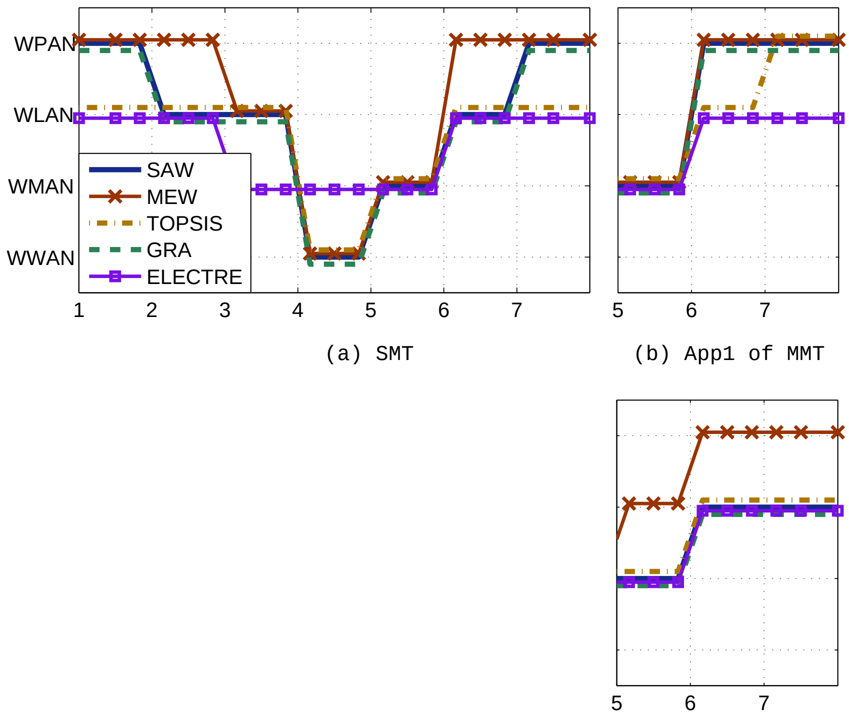
(d) GRA

Figure 3- 2 Coefficients' changes with respect to certain criterion's weight.

3.4. Selection results of various MADM algorithms

In this sub-section, we utilize the series of 7 scenarios described in table 2-1 of section 2.1.3 to evaluate different MADM algorithms' selection results. Two customers using separately single-homed MT (SMT) and multi-homed MT (MMT) move together within a heterogeneous environment consisting of four networks. SMT can only connect to the Internet through one interface at one time, so the selection of its best network should consider simultaneously all the applications together. By contrast, MMT is capable of connecting through multiple interfaces, so different applications might select different networks if necessary.

Figure 3-3 shows separately the two MTs selection results. For the first four scenarios, the single application uses the same network for both SMT and MMT; by contrast, for the last three scenarios, MMT could select different networks for different applications.



(c) App2 of MMT

Figure 3- 3 Selection results of MADM algorithms.

3.5. Traffic load assignment for MADM-based network selection

By considering traffic load as a dynamic attribute in the network ranking module, it is possible to affect the network selection result. Thus, networks with more resource will have more chance to be selected.

In this simulation, we assume that 1000 sessions of an MT with high speed and good power condition arrive one by one, and each of them occupies 0.1% of the selected network's resource. As shown in figure 3-4, when the weight of traffic load is small, WMAN is selected as the best network. Along with the increase of the weight, this networks' traffic load is considered in the network ranking procedure, so other networks are gradually selected. Finally, when the weight of traffic load is relatively large, each network takes approximately 1/4 of the whole traffic.

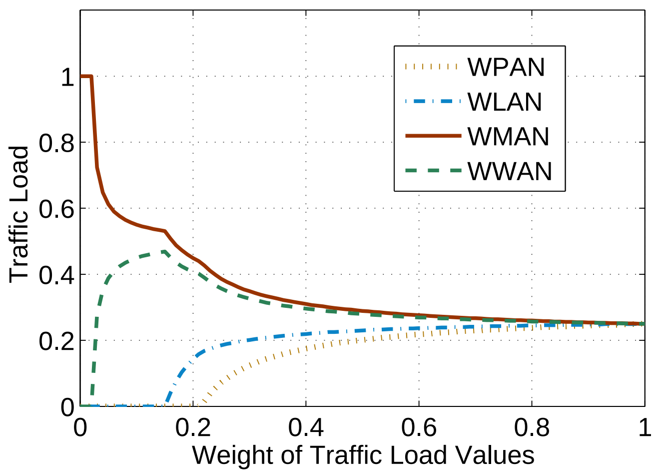


Figure 3- 4 Load balancing feature.

3.6. Important observations and issues

Based on the simulations above, we briefly summarize our most important observations as follows:

- it is feasible to use terminal-side and operator-side requirements to impact the weights of different criteria, but the pair-wise comparison matrix in AHP changes dynamically and frequently for different scenarios. In other words, for each scenario, a weighting matrix should be calculated by pair-wise comparison among all the network attributes. For different scenarios, the matrices are different;
- it is common to have several networks with performance close to the best one, so load balancing and VHO tradeoff are both important.

Moreover, it is possible to divide all the networks into groups, which will be further discussed in chapter 4;

- MADM algorithms may have different coefficients and selection results, but all of them can generally select reasonable networks in various scenarios;
- using traffic load as a criterion in the network ranking algorithm is a simple method of load balancing, and it works well among the networks with similar performance.

Then, we find several existing issues in the scope of MADM-based network selection:

Weighting method: it is inconvenient to manually evaluate weights for different scenarios based on pair-wise comparison matrices by AHP, so novel weighting method is required to efficiently and quickly evaluate weights for different scenarios (see chapter 5);

Mobility-related factors: VHO properties depend on the priorities of networks (i.e. permutation) and cannot be easily used as criteria for network ranking, so further study on how to combine these criteria with other criteria is required (see chapter 4);

VHO tradeoff: after a better network is found by the MADM-based network ranking algorithm, the MT might not want to handover to it in lots of scenarios. For example, the better network might be only a little bit better than the current network, the better network might disappear rapidly, a much better network might appear in a few time, etc. Therefore, a tradeoff is required before executing VHO (see section 7.2);

Load balancing: it is possible to balance the load by using traffic load as a criterion in the same way as others, but weight of this criterion compromises other criteria’s importance, even when load balancing is not required (see section 7.1).

4. MOBILITY-BASED NETWORK SELECTION

This chapter discusses one of our main contributions to the network selection issue: mobility-based network selection using the concept of permutation. Proposals are divided into two sections: one with two groups of networks and the other with *n* groups.

4.1. Introduction

A network selection scheme usually considers multiple groups of factors simultan-eously, including network attributes, operator policies, terminal properties, customer preferences, application QoS levels, VHO properties, etc., as described in section 2.2.1. Some factors are mobility-related, such as cell radius, coverage percentage, terminal velocity, HHO and VHO properties, etc. These factors, representing MTs' mobility features and networks' mobility support capabilities, can be gathered (e.g. by an MIH information server [802.21]) for network selection. For example, according to these factors, high speed MTs should not select a network with small cell radius; otherwise, live applications will be severely disturbed by frequent handovers.

Unfortunately, only a few proposals in the literature considered mobility-related factors. For example, authors of [MJ04] stated that some dynamic factors (e.g. terminal velocity, moving pattern, moving history and location information) should be considered by network selection schemes; a Markov decision process (MDP) model was proposed by [SN08] to take into account connection duration and VHO signaling load; the simulation in [CJ06] used diameter of access point (AP); the study in [WH99] considered cellular diameter and handover latency; the simulation in [OE08] showed different schemes' network re-selection frequencies; and the scheme proposed in [SQ05] assumed that the availability of a hotspot means that not only signal strength is strong enough for transmitting data, but also the MT would stay in its coverage for enough time to reduce the possibility of frequent handover. However, none of the above proposals studied whether these mobility-related factors can be used in the same way as other factors.

VHO properties include VHO signaling cost, latency, rate, etc., which are all important mobility-related factors for network selection. However, the usage of these factors in the above network selection framework is quite complicated. That is because VHO properties depend on not only features of the MT's mobility and different networks' coverage but also the permutation of networks. A *permutation* here is an ordering of all the networks which represents these networks' priorities without considering their availability. At anytime and anywhere, the first available network in the permutation should be selected.

For example, in an HWNs with 3 networks (e.g. UMTS, WiMAX and WLAN), network selection scheme selects the best network based on networks' total costs, including VHO cost (MT moving into or out of WLAN hotspots) and other costs (e.g. monetary costs). In order to explain the idea of permutation clearly, let's consider only VHO cost. VHO cost is not a cost of certain network but a cost of certain permutation. Using '>' to denote that the left-side network has higher priority than the right-side one, the left-side one should be selected when both of them are available. With 3 networks (UMTS, WiMAX and WLAN), there are 6 permutations, i.e.
UMTS>WiMAX>WLAN,
UMTS>WLAN>WiMAX,
WiMAX>UMTS>WLAN,
WiMAX>WLAN>UMTS,
WLAN>UMTS>WiMAX and
WLAN>WiMAX>UMTS.

Detailedly, permutation 'UMTS>WiMAX>WLAN' corresponds to 'no VHO cost' because UMTS is assumed always available due to its ubiquity. By contrast, permutation 'WLAN>UMTS>WiMAX' corresponds to 'large VHO cost'. Similarly, 6 permutations correspond to 6 different (or maybe the same) VHO costs, and the one with the minimum VHO cost is the best (that is 'UMTS>WiMAX>WLAN' or 'UMTS>WLAN>WiMAX'). The above decision is based on only VHO cost, and the best permutation may change if other costs are also considered (e.g. monetary cost). Moreover, the complexity of scheme becomes much higher if multiple costs are considered.

To sum up, different permutations lead to different VHO properties, hence different total costs. Thus, when we use a ranking algorithm to select an alternative based on different networks' total costs, this selected alternative is actually a permutation not a network. Thus, the selection of the best network should be the selection of the best permutation when VHO properties are used in network ranking algorithms.

However, when the heterogeneous environment consists of N networks, the number of permutations will be the factorial of N. Thus, coefficients of a large number of permutations should be calculated and compared to find the best one. Moreover, the evaluation of VHO properties of each permutation is also complicated due to irregular coverage of networks and various moving patterns of MTs. To solve the above problems, this chapter provides solutions by two steps to use mobility-related factors, especially VHO properties, in network ranking algorithms of the above framework.

4.2. Mobility-based scheme considering two groups of networks

4.2.1. Mobility modeling

As shown in section 3.1, a number of network attributes should be combined by the network ranking algorithm. Since these attributes usually have different units, they should be normalized before combining together. However, normalized values are not enough to represent these networks' capabilities. For example, considering some application with certain minimum bandwidth requirement, one network could provide exactly this required bandwidth and the other could provide 90% of it. For the latter, this application may not even work, so the utility is not 90% of the former but much lower than it. Therefore, to combine multiple attributes together, it is better to combine their utilities, not their normalized values.

Based on the study of utility functions in [NV08], sigmoidal form utility functions are suitable for adjusting values of various network attributes, given by:

$$u_2(x) = \frac{(x/x_m)^\eta}{1+(x/x_m)^\eta}, \ (x_m > 0, \eta \geq 2) \qquad (4\text{-}1)$$

where x_m corresponds to the threshold between satisfied and unsatisfied areas (i.e. the centre of the utility curve). η determines the steepness of the utility curve, which makes it possible to model the user sensitivity to variation of network attributes.

Considering an HWNs with UMTS, WiMAX, WLAN and Bluetooth, these networks can be classified into two groups: *ubiquitous networks* (i.e. UMTS and WiMAX) and *hotspot networks* (i.e. WLAN and Bluetooth). Networks in the same group usually have similar values on many criteria, such as monetary cost, power consumption, security level, mobility support capability, etc. Imagining that normalized values of certain attribute (e.g. monetary cost) are going to be adjusted by the sigmoidal utility function in figure 4-1, ubiquitous networks usually have large monetary costs (hence low utilities) and hotspot networks have high utilities. Linear utility function does not change the relative difference of these networks, but sigmoidal utility function could greatly increase the difference between the two groups of networks. This feature is also true for lots of network attributes that a network ranking algorithm considers. Therefore, an HWNs with two groups of networks is studies in this section.

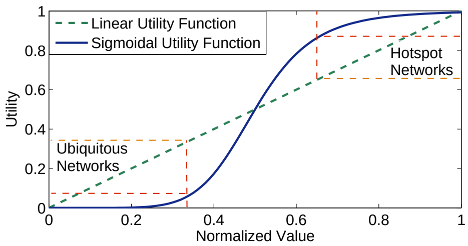


Figure 4- 1 Utilities of different networks through sigmoidal function.

We assume the hotspot network's deployment is based on customers' requirements. For example, personal areas could be covered by Bluetooth, coffee houses and offices could be covered by WLAN, etc. According to the randomicity of distributions of customers, coffee houses, offices, etc., we assume that hotspots are randomly distributed, and their deployment is independent of the ubiquitous network.

Figure 4-2 shows a square area $S = D \times D$ (can be imagined as small as a cell of the ubiquitous network, or as large as a city) with the hotspot network's K hotspots distributed by Poisson point process. When a random walking MT leaves one of these hotspots, the probability of transiting directly into another hotspot equals exactly the percentage of the former hotspot's border being covered by others. For example, the border of hotspot '*A*' in figure 4-2 is covered by two other hotspots, so the probability of directly transiting from this hotspot to another is large. By contrast, the MT has no possibility to transit directly to another hotspot when leaving hotspot '*B*'. Thus, considering the randomicity of hotspots' distribution, the transiting-out probability can be calculated as

$$P = 1 - Q_{K-1}, \tag{4-2}$$

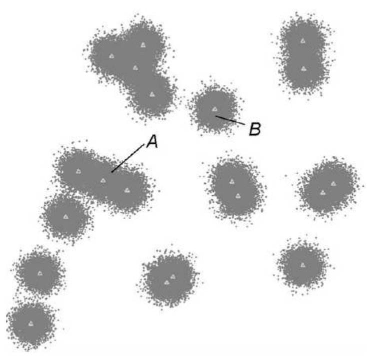


Figure 4- 2 Distribution of hotspots.

where P is the *transiting-out probability* which represents the probability of transiting out of the coverage of the hotspot network when an MT moves out of a hotspot; and Q_K represents the average coverage probability of K

random hotspots. Here, we have to assume that there is no 'border effect', which means only part of the hotspot area is within the whole square area when it is distributed near the border of the whole area.

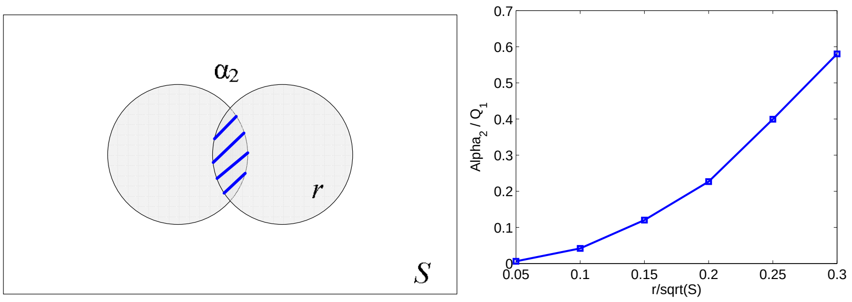
Based on the inclusion-exclusion principle [HJ01] (also known as the sieve principle), the average coverage probability of K random hotspots can be expressed as

$$Q_K = \sum_{i=1}^{K}\left[(-1)^{i-1}\binom{K}{i}\alpha_i\right], \tag{4-3}$$

where $\alpha_i\{2 \le i \le K\}$ represents the average probability covered by i hotspots, and $\alpha_1 = Q_1 = \pi \cdot r^2 / S$, where r is the radius of hotspot. It is quite difficult to calculate $\alpha_i\{2 \le i \le K\}$, so here an experimental result of α_2 is given below in figure 4-3. Then, Q_2 can be expressed as $Q_2 = 2\alpha_1 - \alpha_2$, as shown in figure 4-3(a).

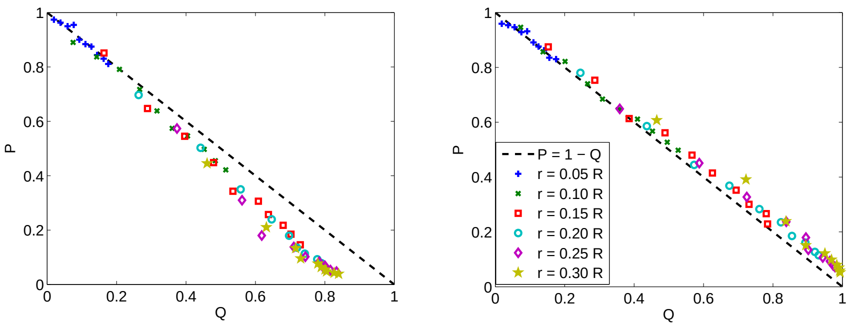
In the analysis below, a large area (S) covered with a large number of hotspots is considered, so there is approximately

$$P = 1 - Q_{K-1} \approx 1 - Q \ . \tag{4-4}$$



(a) (b)

Figure 4- 3 Experimental result of α_2.



(a) hotspots distributed in the central area (b) hotspots distributed in the whole area

Figure 4- 4 Monte Carlo simulation results.

For simplification, Q will be used from now on to represent the coverage of K hotspots instead of Q_K. For the formulation of Q by stochastic geometry for a network with dense hotspots, please refer to [NH07]. In this thesis, Monte Carlo simulation is used to get its value.

Figure 4-4 shows the verification of eq. (4-4) by Monte Carlo simulation. Given K uniformly distributed hotspots, Q is calculated by randomly distributing a large number of points. To calculate P, a number of points are distributed on the border of each hotspot with angles uniformly distributed within [0, 2π). Each figure in Figure 4-4 shows six groups of simulations with different r / R, where r and R are respectively the radius of hotspots and the equivalent radius of the whole square area in figure 4-2, which can be treated as the cell radius of certain ubiquitous network or a whole city. In each of the six groups, 10 simulations are performed (steadily increase the number of hotspots from 10 to 100). In each simulation, all the hotspots are distributed many times. We can see that the curves fit well to eq. (4-4) when r / R is small, and a little bit different when r / R increases. The Difference between figure 4-4 (a) and (b) is the area to distribute hotspots. Since we have to use a limited square area in our simulation instead of a really large area, so there is surely border effect in Monte Carlo simulation. Therefore, in the left figure, hotspots are distributed in the central part of the simulation area, i.e. (R – r)*(R – r), while in the right figure, hotspots are distributed in the whole simulation area, i.e. R*R. We can see that the simulation results are a little bit lower than the curve P = 1 – Q in the left figure, but a little bit higher in the right figure. This demonstrates that the tiny difference between the simulation results and P = 1 – Q is due to the border effect of the simulation, hence further verifies the correctness of eq. (4-4).

Figure 4-5 shows a Markov chain that denotes the MT's movement among this network's hotspots. 'a' and 'd' represent separately the states that the MT is covered by one hotspot or uncovered by any hotspot. U_a is the transiting rate out of one hotspot (i.e. $1/U_a$ is the mean residence time within a hotspot), and U_d is the transiting rate out of area 'd' (i.e. $1/U_d$ is the mean residence time in the uncovered area). P is transiting-out probability, so the probability of moving directly from one hotspot to others is $1-P$. Considering there are $K-1$ other hotspots, the probability of moving from the current hotspot i to another hotpost j is $p_{ij} = (1-P)/(K-1)$. The area of a hotspot is represented by its coverage probability as A/S, where A is the area of one hotspot.

When the Markov chain is stationary, the transiting probability from left to right equals that from right to left. This can be imagined as a huge number of uniformly distributed MTs randomly walking in the whole area, so these MTs will be always uniformly distributed in the future (assuming no 'border effect' as explained below eq. (4-2)). Thus, the following equation is obtained

$$U_d(1-Q) = U_a P \frac{A}{S} K \,. \tag{4-5}$$

When considering a large area covered by a number of hotspots which are not densely deployed, we have $A \cdot K / S \approx Q$. Thus, figure 4-5 can be simplified as figure 4-6 and eq. (4-5) turns into

$$U_d(1-Q) \approx U_a P Q \,. \tag{4-6}$$

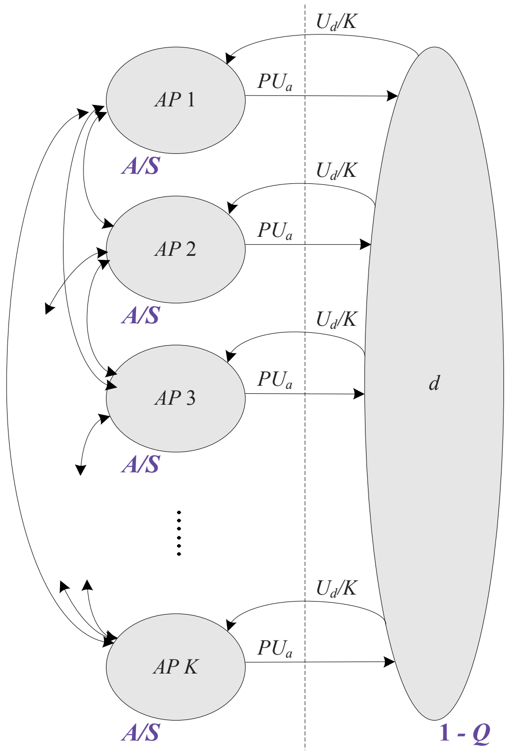


Figure 4- 5 Transition of an MT in an area with *K* hotspots.

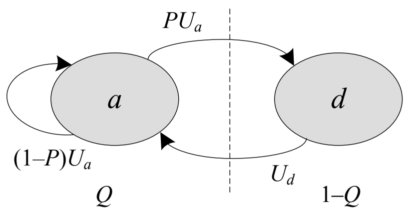


Figure 4- 6 Transition of an MT within 2 states.

Taking eq. (4-4) into eq. (4-6), we get

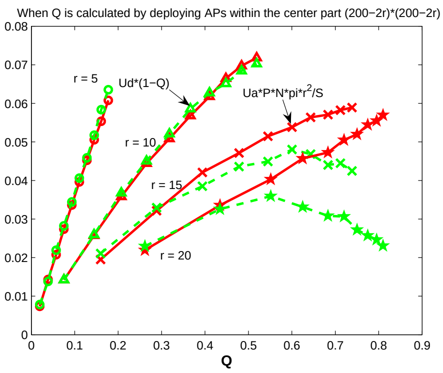
$$U_d = Q U_a \tag{4-7}$$

To support the above analysis and simplifications, Monte Carlo simulations are done as follows: In a simulation area *S* = 200 * 200, *K* hotspots are randomly deployed. Then, one MT starts its random walk from the centre of the simulation area. Q is calculated as explained below figure 4-4. Then, we record the time points when the MT moves into or out of a hotspot, so that the mean cell residence time T_a can be calculated, hence U_a. Similarly, U_d could also be calculated. Notice that, in order to decline 'border effect' when calculating U_a, hotspots should be not be deployed near the border of the simulation area. However, this will affect the calculation of U_d when Q is large, which will be shown below. Detailed simulation parameters are given in Table 4-1.

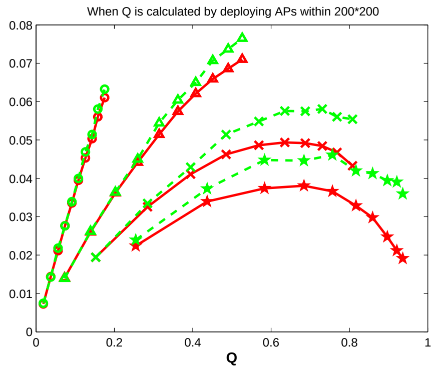
Figure 4-7 shows some simulation results. Seen from figure 4-7(a) and 4-7(b), the only difference is the consideration of 'border effect'. We can see that the green line is lower than the red line in the first figure but higher in the second figure, which demonstrates the seriousness of the 'border effect'. Figure 4-7(c) and 4-7(d) show the effect of the simplification in eq. (4-6). The left figure verifies the correctness of eq. (4-5). We can see that the two lines are quite close except when Q is large (i.e. border effect). The right figure shows the disparity between U_d and QU_a when more and more hotspots are deployed. Figure 4-7(e) is an enlargement of 4-7(d) for $Q < 0.2$, which demonstrates the accordance between U_d and QU_a when Q is relatively small.

Table 4- 1 Simulation parameters.

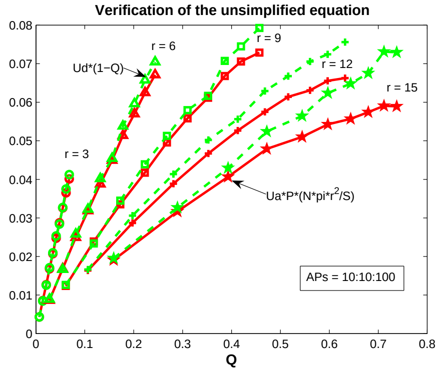
Simulation area (*S*)	200*200
Radius (*r*)	5 cases (from 3 to 15) or 4 cases (from 5 to 20)
Number of hotspots (*K*)	10 cases (from 10 to 100)
Number of rounds in each case	100
Random walking steps in each round	10000
Step length	3



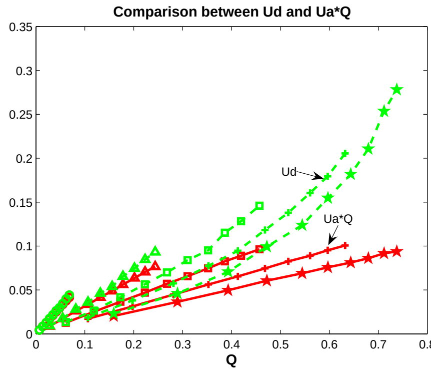
(a)



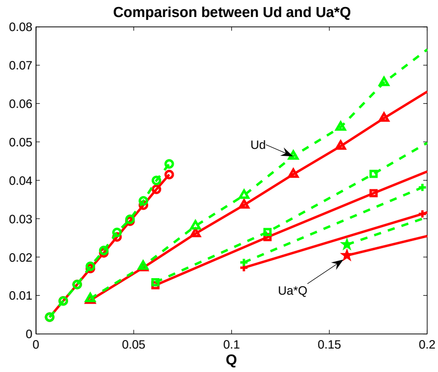
(b)



(c)



(d)



(e) Enlargement of (d) for Q < 0.2

Figure 4- 7 Monte Carlo simulation results.

According to fluid flow model [XH93] and by assuming all the hotspots have circular coverage area, we could further get

$$U_a = E[V]L/\pi S = 2E[V]/\pi r \,, \tag{4-8}$$

where *L* and *S* are perimeter and area of cell, respective. *V* is the velocity of MT. Taking eq. (4-8) into eq. (4-7), we get

$$U_d = 2QE[V]/\pi r \,. \tag{4-9}$$

4.2.2. Evaluation of HHO and VHO costs

Handover rates for both HHO and VHO are summarized in table 4-2. And, we simply define costs of the four handovers as *X*, *Y*, Z_1 and Z_2, including signaling costs, latency, etc., because the evaluation of these costs is out of the scope of this paper. Now, we can evaluate average handover costs of different permutations. Two permutations are considered: 'ubiquitous networks better than hotspot networks (UBN>HSN)' and 'hotspot networks better than ubiquitous networks (HSN>UBN)'.

Table 4- 2 Handover rates and costs.

HO	HO Rate	Cost per HO
UBN → UBN	$U_a r/R$	X
HSN → HSN	$U_a(1-P)Q$	Y
HSN → UBN	$U_a PQ$	Z_1
UBN → HSN	$U_d(1-Q)$	Z_2

For permutation 'UBN>HSN', ubiquitous networks will be always used. Thus, the average handover cost *HC* contains only HHO cost among the cells of ubiquitous networks, which is

$$HC_{UBN>HSN} = \frac{2E[V]}{\pi R} \cdot X \,. \tag{4-10}$$

By contrast, *HC* of the permutation 'HSN>UBN' contains the four parts as shown in table 4-2. To simplify the following derivation, we assume no hotspot spans two or more cells of a ubiquitous network [ZA07], so we get *HC* as

$$HC_{HSN>UBN} = \frac{2E[V]}{\pi R} \cdot X + \frac{2E[V](1-P)Q}{\pi r} \cdot Y + \frac{2E[V]PQ}{\pi r} \cdot Z_1 + \frac{2E[V]Q(1-Q)}{\pi r} \cdot Z_2 \,. \tag{4-11}$$

Generally speaking, a network selection scheme will consider multiple criteria besides the average handover cost, so we suppose that the combination of all the other criteria is Oth_{UBN} for ubiquitous networks and

Oth_{HSN} for hotspot networks with a combined weight of W_1, and HC has a weight of W_2 ($W_2 = 1 - W_1$). Therefore, taking SAW as an example, the total cost of 'UBN>HSN' can be expressed as

$$TC_{UBN>HSN} = \overline{HC}_{UBN>HSN} \cdot W_2 + \overline{Oth}_{UBN} \cdot W_1 , \tag{4-12}$$

where $\overline{HC}$ represents the normalized value of HC and $\overline{Oth}$ represents the combination of other criteria. Meanwhile, the total cost of 'HSN>UBN' is

$$TC_{HSN>UBN} = \overline{HC}_{HSN>UBN} \cdot W_2 + [\overline{Oth}_{HSN} \cdot Q + \overline{Oth}_{UBN} \cdot (1-Q)] \cdot W_1 . \tag{4-13}$$

Hotspot networks are preferred to ubiquitous networks if they have a smaller total cost, written as

$$TC_{HSN>UBN} < TC_{UBN>HSN} . \tag{4-14}$$

Taking eq. (4-12) and eq. (4-13) into eq. (4-14), we obtain the threshold as

$$W_2 < \frac{\overline{Oth}_{UBN-HSN}}{\overline{Oth}_{UBN-HSN} + \dfrac{\overline{HC}_{HSN>UBN} - \overline{HC}_{UBN>HSN}}{Q}} , \tag{4-15}$$

where $\overline{Oth}_{UBN-HSN}$ represents the difference between the other criteria's combination of ubiquitous networks ($\overline{Oth}_{UBN}$) and that of hotspot networks ($\overline{Oth}_{HSN}$). Taking eq. (4-4), eq. (4-10) and eq. (4-11) into eq. (4-15), this threshold can be further expressed as

$$W_2 < \frac{\overline{Oth}_{UBN-HSN}}{\overline{Oth}_{UBN-HSN} + \dfrac{QY + (1-Q)(Z_1 + Z_2)}{Norm}} , \tag{4-16}$$

where $Norm = \sqrt{\dfrac{r^2 X^2}{R^2} + \left\{ \dfrac{rX}{R} + Q[QY + (1-Q)(Z_1 + Z_2)] \right\}^2}$.

4.2.3. Performance evaluations

4.2.2.1 Configuration of network selection simulator

In this section, we evaluated the performance of the above mobility-based network selection scheme. Detailed configurations of our simulator are explained as follows:

- *Criteria*: besides the average handover cost, nine other criteria are considered, i.e. monetary cost, bandwidth, power consumption, security level, bit error rate, burst error rate, jitter, traffic load and signal strength.

- *Requirements*: the terminal velocity is relatively high and the power condition is good; the customer prefers low price; and application flow is conversational.
- *Networks*: the HWN environment is composed of four networks, i.e. Bluetooth, WLAN, WiMAX and UMTS.
- *Adjusting*: criteria are adjusted firstly by normalization, then through sigmoidal utility functions as shown in figure 4-1.
- *Weighting*: AHP is used to evaluate the weights.
- *Ranking*: four MCDM algorithms, i.e. SAW, MEW, TOPSIS and GRA, are used for network ranking.
- *Matrix*: an $m \times n$ value matrix is used to represent the values of different criteria of different networks, where *m* and *n* represent the number of networks and criteria, respectively. In our simulation, we suppose the values of the two dynamic criteria (i.e. traffic load and signal strength) of different networks are the same, in order that the results focus on the impact of the average handover cost. We also simply assume that costs per handover are ($X : Y : Z_1 : Z_2 = 2:3:4:4$) because performance evaluation of various handover strategies is out of the scope of our study.

4.2.2.2 Simulation results

Simulation results in figure 4-8 and figure 4-9 show change of the four networks' coefficients along with the increase of the weight of average handover cost (W_2) and the coverage percentage (*Q*), respectively. For SAW and MEW, the coefficient is the total cost, so it is the smaller the better. For TOPSIS and GRA, the coefficient is the preference value, so it is the larger the better. Based on these figures, we have the following important observations:

- networks in the same group have similar performance;
- as shown in figure 4-9, along with the increase of hotspot networks' coverage, the advantage of hotspot networks gradually increases. Meanwhile, the advantage of ubiquitous networks decreases due to the normalization process;
- when the average handover cost is not considered (i.e. the weight of average handover cost equals 0 in figure 4-8, hotspot networks are generally better than ubiquitous networks;
- when the weight of the average handover cost increases, ubiquitous networks gradually have more chance to be selected, and the threshold between selecting the two groups of networks is also shown in figure 4-8;

- moreover, we can see that different MCDM algorithms have different coefficients and different thresholds, but the trends in these figures are all the same. SAW, TOPSIS and GRA have quite similar thresholds.

4.2.4. Coverage of small-scale networks vs. network selection

Define the probability density function (PDF) of the MT speed as *f*(*V*), and the weight of handover cost for a MT with speed *V* as *W*(*V*). Thus, MTs whose speed is smaller than V_0 will prefer WLAN:

$$W(v_0) = \frac{Others_{UMTS-WLAN}}{\frac{QY + (1-Q)(Z_1 + Z_2)}{Norm} + Others_{UMTS-WLAN}}. \qquad (4\text{-}17)$$

Therefore, the number of consumers who prefer WLAN can be expressed as

$$n(Q) = Q \cdot N_0 \int_0^{V_0} f(V)dV \,, \qquad (4\text{-}18)$$

where N_0 is the total number of consumers, and

$$V_0 = W^{-1}\left(\frac{\overline{Oth}_{UBN-HSN}}{\overline{Oth}_{UBN-HSN} + \frac{QY + (1-Q)(Z_1 + Z_2)}{Norm}}\right). \qquad (4\text{-}19)$$

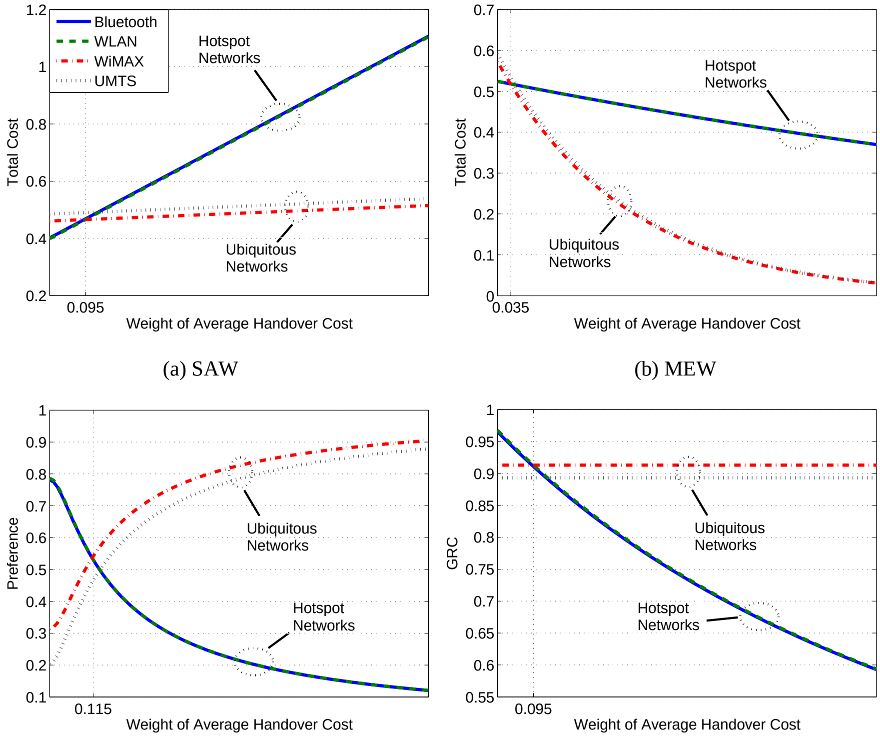
At the same time, we could get the WLAN's CIR as the derivative of eq. (4-18) with respect to *Q*. For example, if we assume the total number of consumers is 1,000,000, and *f*(*V*) and *W*(*V*) are

$$f(V) = e^{-V}, \text{ and } W(V) = 1 - e^{-V}, \; V \in [0, +\infty), \qquad (4\text{-}20)$$

respectively, as shown in figure 4-10. And, also taking eq. (4-4) into account, the CIR can be calculated as shown in figure 4-11. Along with the deployment of more hotspots are deployed, more and more customers will prefer WLAN and the rate (CIR) will not decrease a lot.

However, the increase of WLAN hotspots may not bring all the consumers to use this technology, as shown in figure 4-11(a). Take *Q* = 1 and *P* = 0 into eq. (4-19), we find that WLAN will never be preferred by those consumers whose speed is larger than certain limit as follows:

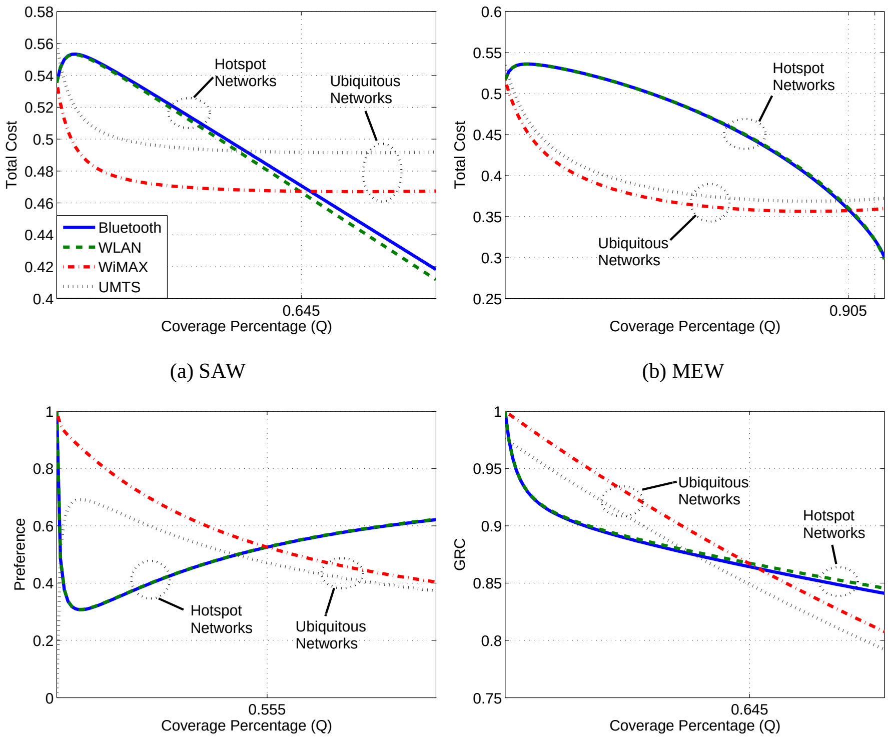
$$V > W^{-1}\left(\frac{1}{1 + \frac{1}{\overline{Oth}_{UBN-HSN}\sqrt{\frac{r^2X^2}{R^2Y^2} + \left(\frac{rX}{RY} + 1\right)^2}}}\right). \qquad (4\text{-}21)$$



(a) SAW (b) MEW

(c) TOPSIS (d) GRA

Figure 4- 8 Coefficients of various networks vs. Weights of average handover cost.



(a) SAW (b) MEW

(c) TOPSIS (d) GRA

Figure 4- 9 Coefficients of various networks vs. Coverage percentage.

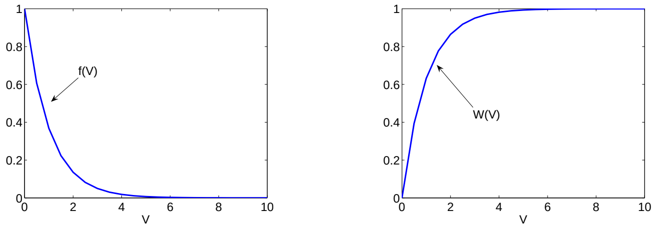


Figure 4- 10 Examples of *f*(*V*) and *W*(*V*).

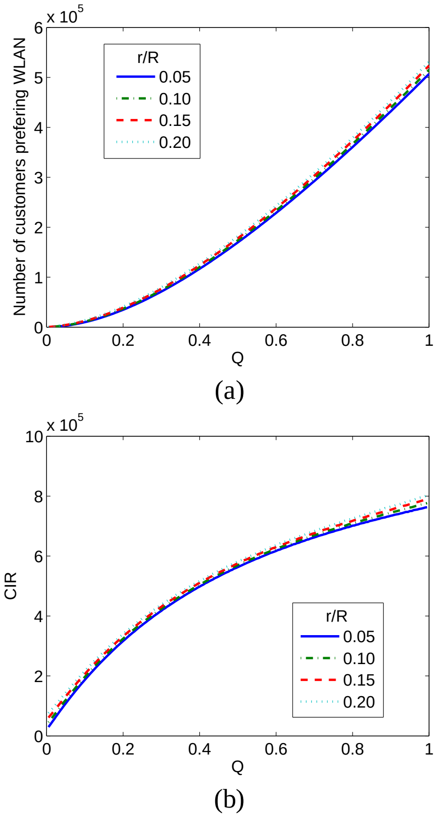


Figure 4- 11 Consumer increment rate (CIR) of WLAN.

4.3. Mobility-based scheme considering *N* groups of networks

We studied in the last section the scenario with two groups of network, but when the heterogeneous environment consists of *n* networks, the number of permutations will be the factorial of *n*. Thus, a large number of permutations' total costs should be calculated and compared to find the best one. Moreover, calculation of the average VHO cost of each permutation is also complicated due to irregular coverage of networks and various moving patterns of MTs.

To solve the time complexity problem of permutation-based schemes, our previous proposal [WL09-1] was to classify all the wireless networks into two groups, i.e. hotspot networks and ubiquitous networks. We modelled an

MT's mobility in this 2-network environment and calculated a threshold between the two groups. When sigmoidal utility function is used to adjust attributes' normalized values, it works well to classify all the networks into the above two groups, because this function enlarges the difference of networks on its two sides. However, sigmoidal utility function is not always used. For QoS-related attributes (e.g. bandwidth), it is necessary to have a value better than a threshold, hence sigmoidal utility function is suitable; while for non-QoS-related attributes (e.g. monetary cost), sigmoidal utility function has no obvious advantage compared with linear or other utility functions.

Mobility-related factors are non-QoS-related, so sigmoidal function might not be used for them. Hence, it is quite possible that networks are classified into more than two groups based on their mobility support capabilities. For example, a common classification of all the networks with four groups is WPAN, WLAN, WMAN and WWAN. Moreover, classifying networks into groups leads to 2nd round selection within the best group, which has an extra time cost. In a word, it is necessary to study the scenario when there are more than two groups of networks. Therefore, this section discusses permutation-based network selection schemes in a generic scenario with *N* groups of networks.

4.3.1. Evaluating total costs of permutations in a generic scenario

In an HWNs with *N* groups of networks, an MT can be covered by any subset of these *N* groups. By assuming deployments of all the groups of networks are independent of one another, we could easily obtain the MT's mobility properties within these HWNs.

Figure 4-6 shows an MT's transitions within a single network environment. '*a*' and '*d*' represent separately the states that the MT is covered and uncovered by this network. U_a is the transiting-out rate from a cell of this network (i.e. $1/U_a$ is the mean cell residence time), and U_d is the transiting-in rate from '*d*' to certain cell of the network (i.e. $1/U_d$ is the mean residence time in the uncovered area). *P* is transiting-out probability, so the probability of moving directly to another cell of this network (i.e. HHO within this network) is $(1-P)$. *Q* is the coverage percentage of this network.

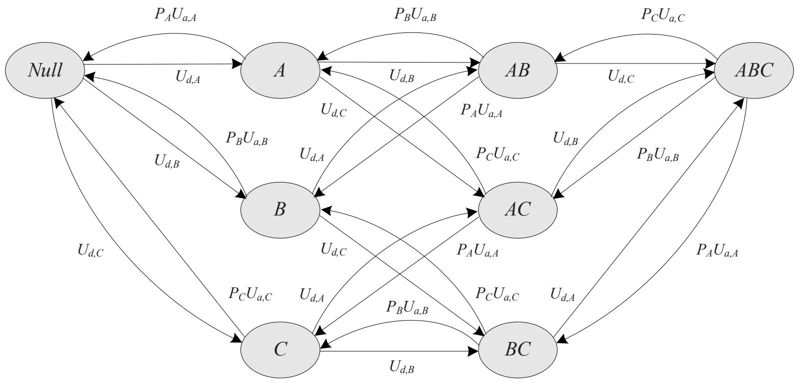
For a generic scenario, we define a *state* as an area covered by the same groups of networks, so there are totally 2^N states for *N* groups of networks. Figure 4-12 shows transitions among 2^3 states in the Markov chain of an HWNs with 3 groups of networks. For example, state '*AB*' means a MT being currently covered by group *A* and group *B*. Transition rates in this Markov chain are calculated based on the assumption that all networks' deployments are independent. Symbols used in this figure and later derivations of a generic scenario are summarized in table 4-3. For example, $U_{a,i}$ has a similar meaning of U_a in the 2-group scenario, but it represents specifically the transiting-out rate of certain cell of group *i* in a generic scenario with *N* groups.

To calculate the average VHO cost of certain permutation, we combine the states which have both the same number of groups and the same first available group as one *big state*. Taking the scenario shown in figure 4-12 as an example, state *AB* and state *AC* form a big state when permutation '*A*>*B*>*C*' is considered. Moreover, we assume a group of ubiquitous networks is always available, so the *Null* state is eliminated. Thus, the total number of states decreases to $N(N+1)/2$. As shown in the second Markov chain of figure 4-12, states in the same row have the same first available group. Thus, VHO is required when an MT transits into a state that contains a better first available group (i.e. from bottom-left to top-right) or a state that does not contain the former first available group (i.e. from top-right to bottom-left). Therefore, there is no need to consider all the transitions in the Markov chain, instead, the number of transitions leading to VHO can be decreased from $2\cdot\sum_{i=1}^{N}[C_N^{i-1}\cdot C_N^{i}]$ to $2\cdot\sum_{j=1}^{N-1}\sum_{i=2}^{N-j+1}[(i-1)\cdot C_{N-i}^{j-1}]$. In the proposed scheme, the average VHO cost of a permutation is calculated by considering only the transitions that lead to VHOs.

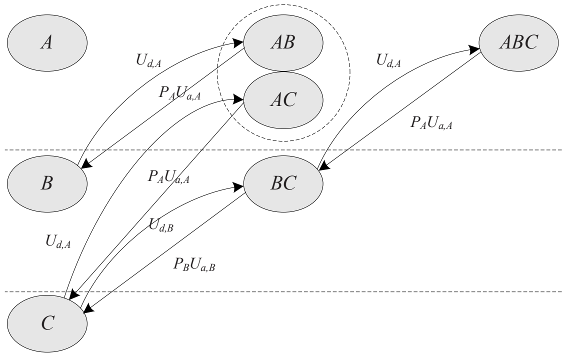
Based on figure 4-12, we now derive the formula of average handover cost. Handover cost is not only related with permutation, MT mobility and network coverage, but also related with mobility management strategy. However, performance evaluation of various mobility management strategies is out of the scope of this paper. Here, we only focus on transitions of an MT among HWNs, so that a generic scheme can be obtained. The derivation of the total cost of permutation '$g_1 > g_2 > ... > g_N$' is as follows:

There are two types of handovers, i.e. HHO and VHO, so the average handover cost contains both HHO and VHO costs. According to the Markov chain of figure 4-6, the rate of HHO to another hotspot for group *i* is $Q_i(1-P_i)U_{a,i}$, so the average HHO cost in unit time can be calculated as

$$HC_{HHO}=\sum_{i=1}^{N}\left\{[\prod_{j=0}^{i-1}(1-Q_j)]Q_i(1-P_i)U_{a,i}\cdot hho_i\right\}. \qquad (4\text{-}22)$$



(a) Complete transitions



(b) Simplified transitions

Figure 4- 12 Transitions in a 3-group scenario.

Table 4- 3 Symbols for a generic scenario.

n	number of networks
N	number of groups of networks
M	number of attributes except HC
L	number of groups better than the ubiquitous
g_i	the ith group in a permutation
$U_{a,i}$	transiting-out rate from a cell of group i
$U_{d,i}$	transiting-in rate to certain cell of group i
P_i	transiting-out probability of group i
Q_i	coverage percentage of group i
r_i	cell radius of group i
hho_i	cost of HHO within group i
$vho_{i,j}$	cost of VHO from group i to group j
$v_{i,j}$	value of the jth attribute of the ith group
ω_{HC}	weight of HC
ω_i	weight of attribute i (except HC)
λ_i	arriving rate of the scheme trigger event i
μ_i	total transiting rate of group i
V	MT velocity

For VHO, when the MT moves from a state uncovered by the first *i* groups (i.e. g_1, g_2, ... , g_N) to a state covered by group *i*, it should handover to group *i*. Considering this VHO could be from any group of (g_{i+1}, g_{i+2}, ... , g_N), the average cost of one VHO to group *i* can be expressed as

$$HC_{Single\_VHO}(i) = (1-Q_i)U_{d,i} \cdot [\prod_{j=0}^{i-1}(1-Q_j)] \cdot \left\{ vho_{i+1,i}Q_{i+1} + \sum_{k=i+2}^{N}[vho_{k,i}Q_k \prod_{j=i+1}^{k-1}(1-Q_j)] \right\} \quad (4\text{-}23)$$

Thus, the average VHO cost of moving into a better group can be calculated by summing up costs of VHOs to any group of (g_1, g_2, ... , g_{N-1}), which can be written as

$$HC_{VHO}^{+} = \sum_{i=1}^{N-1}\left\{ U_{d,i} \cdot \sum_{k=i+1}^{N}[vho_{k,i}Q_k \prod_{j=1}^{k-1}(1-Q_j)] \right\}. \quad (4\text{-}24)$$

Similarly, the average VHO cost of moving out of the former first available group is calculated as

$$HC_{VHO}^{-} = \sum_{i=1}^{N-1}\left\{ \frac{Q_i P_i U_{a,i}}{1-Q_i} \cdot \sum_{k=i+1}^{N}[vho_{i,k}Q_k \prod_{j=1}^{k-1}(1-Q_j)] \right\}. \quad (4\text{-}25)$$

By combining eq. (4-22), eq. (4-24) and eq. (4-25), the average handover cost is obtained as

$$HC = HC_{HHO} + HC_{VHO}^{+} + HC_{VHO}^{-}. \quad (4\text{-}26)$$

On the other hand, other attributes should also be evaluated. For permutation '$g_1 > g_2 > ... > g_N$', the other attributes are combined as

$$Oth = \sum_{i=1}^{N}[Oth_i \cdot Q_i \cdot \prod_{j=0}^{i-1}(1-Q_j)], \quad (4\text{-}27)$$

where $Oth_i = \sum_{j=1}^{M} v_{i,j} w_j$.

After *HC* and *Oth* are obtained as described above, the two parts of costs should be normalized and summed up. Therefore, the total cost of permutation '$g_1 > g_2 > ... > g_N$' is finally obtained as

$$TC = \overline{HC} \cdot w_{HC} + \overline{Oth} \cdot (1 - w_{HC}), \quad (4\text{-}28)$$

where $\overline{HC}$ and $\overline{Oth}$ represent the normalized value of certain attribute and the combination of a group of attributes, respectively.

4.3.2. Methods to get the best permutation

4.2.2.1 Simplified Besper

In classic best network schemes (*Besnets*), sorting algorithms, e.g. bubble sort, are used to get the rank of all the networks. In permutation-based scheme, it is not necessary to rank all the permutation because the best permutation is already the rank of all the groups of networks. Therefore, the basic best permutation scheme (*Basic Besper*) is to calculate total costs of all the permutations based on the results obtained in section above. Then, the permutation with the minimum total cost is selected as the best one. However, Basic Besper has an obvious problem, which is the time spent for all the calculations. More precisely, when there are *N* groups of networks, the number of permutations is the factorial of *N*. Hence, Basic Besper has to calculate *N*! permutations' total costs, and find the best one with $N!-1$ pair-wise comparisons between them. This is not efficient for real-time network selection during VHOs.

Due to this reason, we propose a modified method (*Simplified Besper*) to find the best permutation in a quicker way. Since we assume the group of ubiquitous networks always exists, the first step of Simplified Besper is to decide the position of this group in the best permutation. This is achieved by comparing each non-ubiquitous group with the ubiquitous one, which has been studied in our previous work of a 2-network scenario and a threshold between the two groups was obtained as eq. (4-19) in section 4.2.4. For the generic scenario discussed in this section, the threshold can be modified as

$$\theta_i = w_{HC}^{-1}\left(\frac{\overline{Oth}_{U-i}}{\overline{Oth}_{U-i} + \dfrac{Q_i hho_i + (1-Q_i)(vho_{i,U} + vho_{U,i})}{Norm}}\right), \tag{4-29}$$

where $\overline{Oth}_{U-i}$ represents the difference of the other attributes' combinations between the ubiquitous group ($\overline{Oth}_U$) and group i ($\overline{Oth}_i$), *Norm* is for normalization of *HC*s of different groups, and $w_{HC}(V)$ is a function of MT velocity. Assuming $w_{HC}(V)$ is monotonically increasing, group i is found better than the ubiquitous group if V is smaller than θ_i.

By doing the above comparisons, we get L ($0 \le L \le N-1$) groups better than the ubiquitous one, hence $N-L-1$ groups worse. Due to the ubiquity of ubiquitous group, the groups worse than it has almost no opportunity to be used. Hence, the ordering of these groups is not as important as the better ones.

In the end, total costs of L! permutations based on the L better groups are calculated, and the permutation with the minimum total cost will be selected.

Simplified Besper does not decrease the time complexity of Basic Besper (still O(N!)), but it really decreases the time cost for getting the best permutation in most scenarios. The number of permutations we need to consider decreases to $L!$, and the number of pair-wise comparisons decreases to $(L!-1)$. Since the number of groups of networks is generally not large (e.g. four groups as WPAN, WLAN, WMAN and WWAN), this simplification is actually quite helpful to decrease the time cost.

4.2.2.2 Enhanced Besper

Compared with Basic Besper, Simplified Besper obviously decreases the time cost, but there is the possibility that its time cost is still large in some scenarios. For example, when the ubiquitous group is found to be the worst after comparisons with other groups, we still need to consider $(N-1)$ permutations in later calculations, which has no obvious advantage compared with the calculation of N! permutations in Basic Besper. Therefore, we propose in this section a further modified method (*Enhanced Besper*) which contains two steps and is very practical for network selection with a time complexity of only O(N) for getting the best group before VHO.

Step 1: figure 4-13 shows the procedure of Enhanced Besper. The first step is to find the best group with permutation-based pair-wise comparisons, which is different from Besnets because we consider VHO properties in this step.

Since we assume the deployments of networks are independent, the selection of a high rank group has nothing to do with a low rank group's coverage. By assuming that g_1 and g_2 are the best two groups, permutations '$g_1 > g_2 > ... > g_N$' and '$g_2 > g_1 > ... \quad > g_N$' represent the two cases when g_1 and g_2 are the best, respectively. The ordering from g_3 to g_N is not important because they are low rank groups. Therefore, we obtain total costs of the two permutations as $TC\{g_1 > g_2 > ... > g_N\}$ and $TC\{g_2 > g_1 > ..., > g_N\}$ where TC of each permutation can be obtained based on the results in Section II.

By comparing the two total costs above, we find the better permutation g_x of g_1 and g_2. Then, we compare g_x with g_3 by assuming g_x and g_3 are the best two groups in a similar way. Thus, we can compare through all the groups by ($N-1$) comparisons and find the best one (g_x) at the end of Step 1.

The above procedure of Step 1 has a time complexity of O(N), which is as fast as using Besnets to find the best group. However, when several groups have similar performance, it is possible that the best group got by this procedure is different from Basic Besper. Therefore, a more robust option for Step 1 is to do $N \times N$ pair-wise comparisons among all the groups, which is similar to ELECTRE and has a time complexity of O(N^2).

Step 2: After Step 1, the MT performs VHO immediately. Then, Step 2 is performed to get the best permutation. We also provide two options for Step 2: a fast option is Simplified Besper; while a precise option is Basic Besper. Since the best group has been decided in Step 1, there is generally no need to consider it again in Step 2, which decreases the time cost of Step 2.

Since Enhanced Besper separates the task of getting the best permutation into two steps, it is quite suitable for practical usage. As shown in figure 4-13, the scheme is triggered at t_0 and Step 1 is firstly processed. When Step 1 completes at t_1, a best group is obtained based on permutation-based pair-wise comparisons, and VHO is performed to a certain network in the best group. Then, Step 2 is processed to get the best permutation, which takes a relatively long time and ends up at t_2, but VHO is not disturbed by it.

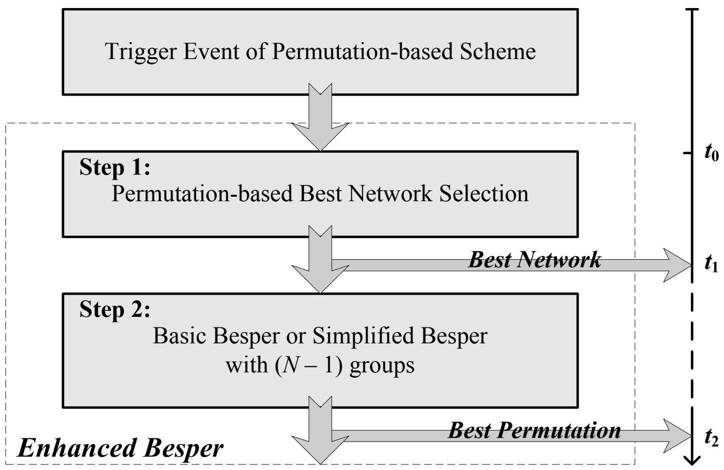


Figure 4- 13 Procedure of Enhanced Besper.

4.3.3. Performance evaluations

To evaluate the performance of Bespers, we establish in this section a network selection simulator, which is configured as follows:

Attributes: besides average handover cost, we also consider eight other attributes, i.e. monetary cost, bandwidth, power consumption capability, security level, bit error rate, jitter, traffic load and signal strength.

Weights: to calculate all the attributes' weights, we assume that the terminal velocity is relatively high, the power condition is good, the customer prefers low price and large bandwidth, the operator wants load balancing, application flows are streaming and conversational, etc. Based on these assumptions, AHP is used to calculate the weights.

Networks: for simulations of $N = 2$, the HWN environment is composed of WLAN and WWAN; for $N = 3$, of WPAN, WLAN and WWAN; for $N = 4$, of WPAN, WLAN, WMAN and WWAN.

Parameters: an $N \times (M + 1)$ value matrix is normalized and processed in our simulations. We assume firstly values of the two dynamic attributes (i.e. traffic load and signal strength) of various networks are the same, in order to focus on the effects of mobility-related factors. Secondly, $vho_{i,j}$ and hho_i are all assumed the same because the comparison of various mobility management strategies is out of the scope of this paper. Thirdly, by assuming the cells of all the networks are circular and according to fluid flow model [LX95], the transiting-out rate of a cell of group i is obtained as $U_{a,i} = 2E[V]/\pi r_i$. Further assuming that the Markov chain shown in figure 4-6 for each group is stationary, we get the transiting-in rate of group i as $U_{d,i} = 2Q_i E[V]/\pi r_i$. And, coverage percentages of the four groups are assumed as 20%, 40%, 90% and 100%, respectively.

VHO rate and Total cost: Bespers consider mobility-related factors, so a network is not preferred if it does not fit for the mobility requirement of MT, customer or applications. In numerous scenarios (e.g. an MT with high velocity), a network with small cells is not selected by Bespers, so the VHO rate is greatly decreased, as shown in figure 4-14. Due to the same reason, Bespers usually have lower total cost than Besnet, as shown in figure 4-15. Moreover, we can also see that the difference of total costs between Besper and Besnet becomes large when the weight of *HC* increases.

Scheme trigger rate: another important advantage of Bespers is the low trigger rate. As we know, any of the following events will trigger Besnet:

- Event-1: availability of one network;
- Event-2: parameter of terminal properties, applications, dynamic network properties or customer preferences.

Since Bespers provide a permutation of all the groups, the first available group will be used when the MT moves across various states, e.g. states in figure 4-12. Hence, the change of any network's availability will not trigger Bespers, which greatly decreases the trigger rate of the scheme (especially for high speed MTs).

Assuming the average arriving rates of Event-2 are $\lambda_1, \lambda_2, ...\lambda_K$ respectively, the total trigger rate by Event-2 is $\sum_{i=1}^{K} \lambda_i$. Total transmitting rate for the *i*th group includes both transiting-in and -out rates which can be written as

$\mu = Q_i P_i U_{a,i} + (1 - Q_i) U_{d,i}$ based on figure 4-6. Since all the groups are assumed independent, the total trigger rate by Event-1 is $\sum_{i=1}^{N} \mu_i$.

To sum up, the trigger rate of Besnet is $\sum_{i=1}^{K} \lambda_i + \sum_{i=1}^{N} \mu_i$, while that of Besper is only $\sum_{i=1}^{K} \lambda_i$. Figure 4-16 shows a comparison of trigger rates between Besnet and Besper. We can see that when the MT velocity or the number of networks is large, the trigger rate of Besper is much lower than Besnet.

Time complexity: this is a common problem of permutation-based schemes. As shown in figure 4-17, the increment rate of the time cost of Basic Besper is more than exponential growth with respect to *N*. Seen from figure 4-17 and figure 4-18, Simplified Besper and Enhanced Besper both have much lower time cost than Basic Besper, and the time cost for finding the best permutation decreases when the weight of *HC* is large. More important, time cost of Step 1 in Enhanced Besper increases linearly with respect to *N*, which is similar to that of Besnet. Therefore, the time complexity of permutation-based network selection schemes is not an issue any more.

Based on our simulation results, we obtain several important observations. For MTs who seldom move, Bespers and Besnets are similar; otherwise, Bespers are greatly better. For Basic Besper and Simplified Besper, networks should not be classified into too many groups; while for Enhanced Besper, the number of groups is not a problem because the time cost of its Step 1 is always small due to linear growth. Thus, instead of dividing networks into groups, we might use networks individually in Enhanced Besper.

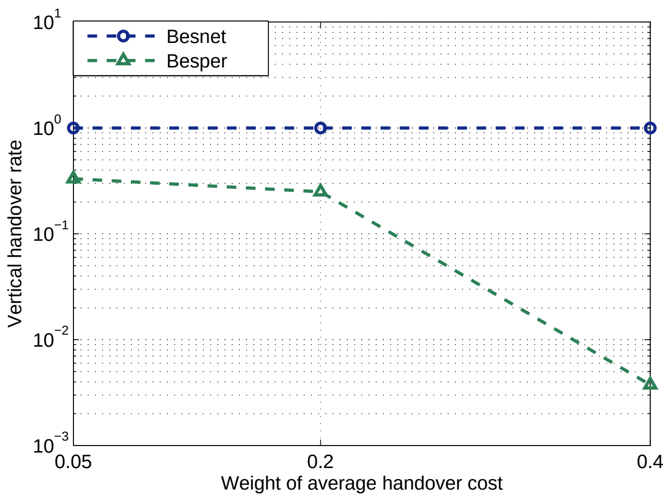


Figure 4- 14 Vertical handover rate comparison.

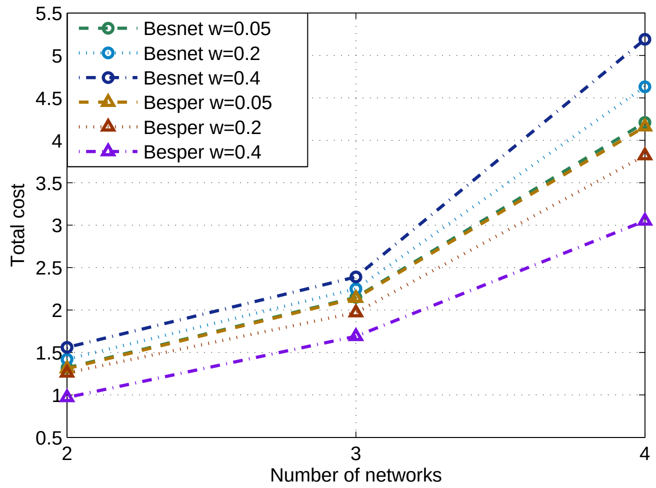


Figure 4- 15 Total cost comparison.

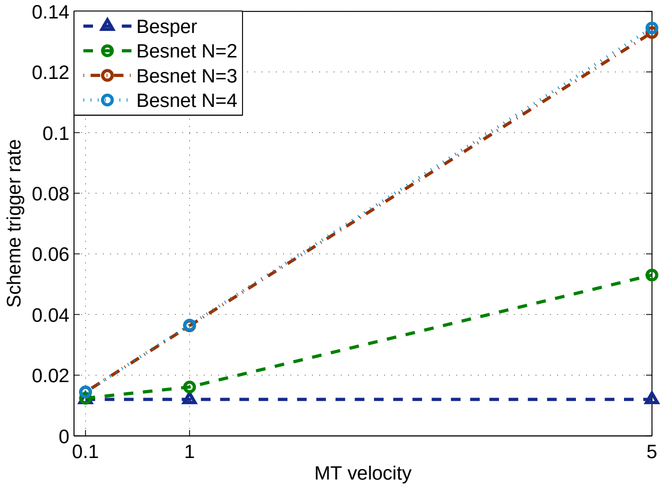


Figure 4- 16 Trigger rate comparison.

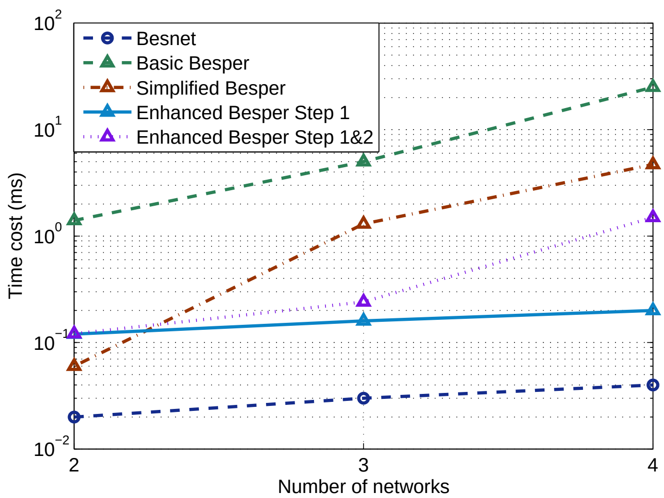


Figure 4- 17 Time cost vs. Number of groups of networks.

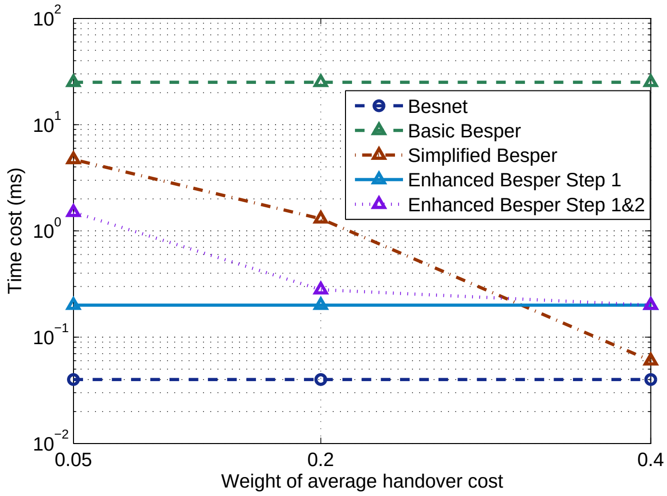


Figure 4- 18 Time cost vs. Weight of handover cost (with 4 groups of networks).

5. TRUST: A NOVEL WEIGHTING METHOD FOR NETWORK SELECTION

This chapter describes a new subjective weighting method specifically for the network selection issue. This method is called trigger-based because the weights are calculated based on the events triggering the current network selection. More important, as a weighting method for network selection, it should be fast and automatic.

5.1. Modeling the weighting issue for network selection

In the MADM-based network selection procedure, multiple attributes are usually adjusted by normalization, fuzzy logic, utility functions, etc. Then, they are combined based on their weights to obtain a total utility (or total cost) for each network. Finally, the network with the maximum utility (or minimum cost) will be chosen as the best network. Weights are calculated in two parts: network attributes are used for calculating objective weights, while subjective requirements (including terminal properties, customer preferences, application QoS requirements, operator policies, etc., see section 2.2.1) are used for subjective weights. Therefore, we model the weighting procedure as follows:

Suppose n attributes are used in a network selection scheme and the number of candidate networks is m, we have the following decision matrix:

$$\mathbf{NW} = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ . & . & . & . \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix}, \tag{5-1}$$

where a_{mn} represents the value of the nth attribute of the mth candidate network.

In order to combine these attributes together, it is necessary to know their relative importance in advance, so weights of these attributes should be calculated and employed in network selection schemes. There are two types of weights: objective and subjective. Objective weights are directly obtained based on the above decision matrix, which are represented as

$$\mathbf{W_O} = [wo_1 \quad wo_2 \quad ... \quad wo_n]. \tag{5-2}$$

Dissimilarly, subjective weights are usually obtained based on the subjective feelings of decision maker (DM). In the network selection issue, the DM is not exactly the customer. Subjective information should actually include customer preferences, terminal properties, application QoS requirements, operator policies, and so on. Based on these information, subjective weights can be obtained, which are represented as

$$\mathbf{W_S} = [ws_1 \quad ws_2 \quad \ldots \quad ws_n]. \tag{5-3}$$

The final weights are obtained by combining the objective and subjective weights, as follows:

$$\mathbf{W} = \frac{\mathbf{W_O} \cdot \mathbf{W_S^\Delta}}{\mathbf{W_O} \cdot \mathbf{W_S^T}} = [w_1 \quad w_2 \quad \ldots \quad w_n], \tag{5-4}$$

where w_j is the combined weight of attribute j, given by

$$w_j = \frac{ws_j \cdot wo_j}{\sum_{i=1}^{m}(ws_i \cdot wo_i)},$$

and $\mathbf{W_S^\Delta}$ is a diagonal matrix, given by

$$\mathbf{W_S^\Delta} = \begin{bmatrix} ws_1 & 0 & . & 0 \\ 0 & ws_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & ws_n \end{bmatrix}.$$

In the following two sections, we will discuss in detail on the methods to obtain the objective weights $\mathbf{W_O}$ and the subjective weights $\mathbf{W_S}$.

5.2. Objective weighting methods

To obtain objective weights, entropy is a widely used weighting method in various research and industrial fields. In information theory, entropy is a criterion for the amount of uncertainty represented by a discrete probability distribution which agrees that a broad distribution represents more uncertainty than does a sharply peaked one [BB07]. This measure of uncertainty is given by Shannon as

$$E = S[p_1 \quad p_2 \quad \ldots \quad p_m] = -k\sum_{i=1}^{m}[p_i \ln(p_i)], \tag{5-5}$$

where p_i are normalized values with $\sum_{i=1}^{m} p_i = 1$ and k is a positive constant.

When all p_i are equal to each other, the entropy E reaches its maximum value, which represents the information denoted by $[\,p_1 \quad p_2 \quad \ldots \quad p_m\,]$ is minimum.

For calculating the objective weights in the network selection procedure, the decision matrix **NW** is firstly normalized as follows

$$\mathbf{NW_{NORM}} = \begin{bmatrix} x_{11} & x_{12} & . & x_{1n} \\ x_{21} & x_{22} & . & x_{2n} \\ . & . & . & . \\ x_{m1} & x_{m2} & . & x_{mn} \end{bmatrix}, \tag{5-6}$$

where $x_{ij} = a_{ij} / \sum_{i=1}^{m} a_{ij}$.

The entropy of each attribute is calculated as

$$E_j = S[x_{1j} \quad x_{2j} \quad \ldots \quad x_{mj}] = -k\sum_{i=1}^{m}[x_{ij}\ln(x_{ij})], \tag{5-7}$$

where $k = 1/\ln(m)$ to guarantee that $0 \le E_j \le 1$.

The degree of diversification D_j of the information provided by the outcomes of attribute j can be defined as

$$D_j = 1 - E_j\,. \tag{5-8}$$

Finally, the objective weight of attribute j can be obtained by normalization as

$$wo_j = \frac{D_j}{\sum_{i=1}^{n} D_j}\,. \tag{5-9}$$

Another objective weighting method is the variance method which decides the weights of attributes according to their variations among all the candidate alternatives [JW08], given by

$$D_j = V[x_{1j} \quad x_{2j} \quad \ldots \quad x_{mj}] = \sqrt{\frac{\sum_{i=1}^{m}(x_{ij} - \overline{x}_j)^2}{m\overline{x}_j}}, \tag{5-10}$$

where $\overline{x}_j$ represents the mean value of the jth attribute. The objective weights of the jth attribute can be finally obtained by the same normalization as the entropy method.

5.3. Subjective weighting methods

To obtain subjective weights, pair-wise comparison between each pair of attributes is usually performed by the DM. By doing this, an *n*-by-*n* square matrix is obtained as

$$\mathbf{PW} = \begin{bmatrix} b_{11} & b_{12} & . & b_{1n} \\ b_{21} & b_{22} & . & b_{2n} \\ . & . & . & . \\ b_{n1} & b_{n2} & . & b_{nn} \end{bmatrix}, \tag{5-11}$$

where b_{ij} represents the comparison between attribute i and j, given as $b_{ij} = ws_i / ws_j$ for precise subjective estimation. Therefore, in the above pair-wise comparison matrix, we have $b_{ij} = 1/b_{ji}$ and $b_{ii} = 1$ for all $i, j \in \{1, 2, ..., n\}$.

Based on the above matrix, there are two commonly used methods for calculating the subjective weights: eigenvector method and weighted least square method.

In ideal case, the pair-wise comparison matrix of eq. (5-11) can be written as

$$\mathbf{PW} = \begin{bmatrix} \frac{ws_1}{ws_1} & \frac{ws_1}{ws_2} & . & \frac{ws_1}{ws_n} \\ \frac{ws_2}{ws_1} & \frac{ws_2}{ws_2} & . & \frac{ws_2}{ws_n} \\ . & . & . & . \\ \frac{ws_n}{ws_1} & \frac{ws_n}{ws_2} & . & \frac{ws_n}{ws_n} \end{bmatrix}. \tag{5-12}$$

Multiplying this matrix with subjective weights $\mathbf{W_S} = [ws_1 \quad ws_2 \quad ... \quad ws_n]$ yields

$$\mathbf{PW} \cdot \mathbf{W_S} = \mathbf{n} \cdot \mathbf{W_S}. \tag{5-13}$$

The above derivation is based on the assumption of $b_{ij} = ws_i / ws_j$ for all i, $j \in \{1, 2, ..., n\}$. Actually, the precise values of b_{ij} are unknown and must be estimated. In other words, the judgments by the DM are subjective and cannot be completely accurate to satisfy the above derivation. Since small perturbations in the coefficients imply small perturbations in the eigenvalues, we can define a matrix **PW** as the DM's estimate of matrix **PW** (surely with small perturbations). Then, a vector of subjective weights $\mathbf{W_S^*}$ can be calculated as the eigenvector of matrix **PW** corresponding to its largest eigenvalue λ_{MAX}. The eigenvector can be obtained by solving the following system of linear equations:

$$(\mathbf{PW}^* - \lambda_{MAX}\mathbf{I}) \cdot \mathbf{W}_\mathrm{S}^* = 0 \tag{5-14}$$

where **I** is an identity matrix.

Combining with the eigenvector method, AHP is usually employed, which is defined as a procedure to divide a complex problem into a number of deciding criteria and sub-criteria and integrate their relative importance to find the optimal solution. Based on AHP, attributes in the network selection issue are structured as a decision hierarchy. Then, eigenvector of each sub-layer is calculated to represent the weights of attributes in this sub-layer. In the end, weights of different sub-layers are synthesized as the final weights. An example of using AHP to calculate weights of QoS-related attributes can be found in [SQ05].

Another method to obtain the subjective weights is the weighted least square method, but this method has not been well used in the studies of the network selection issue. The concept of this method is to solve the following constrained optimization problem:

$$\min z = \sum_{i=1}^{n}\sum_{j=1}^{n}(b_{ij}^* ws_j - ws_i)^2 \tag{5-15}$$

$$s.t. \sum_{i=1}^{n} ws_i = 1,$$

where b_{ij}^* represents the value in the estimated matrix $\mathbf{PW}^*$.

The weights in the above model can be obtained by solving a system of linear equations as follows:

$$\sum_{i=1}^{n}(b_{ik}^* ws_k - ws_i)b_{ik}^* - \sum_{j=1}^{n}(b_{kj}^* ws_j - ws_k) + \eta = 0, \tag{5-16}$$

$$\text{and } \sum_{i=1}^{n} ws_i = 1,\ k = 1,2,\ldots,n,$$

where η represents the Lagrangian mutiplier.

5.4. Inappropriateness of the eigenvector method

For weighting the attributes in the network selection issue, various factors should be considered. Entropy method can be used to obtain the objective weights which denote the relative differences of candidate networks respecting to various attributes. However, it is not enough to use only objective weights for representing these attributes relative importance because most of the factors for deciding the weights in this issue are subjective information which requires subjective weighting methods.

The two subjective weighting methods described above both use pair-wise comparison matrices, but only eigenvector method with AHP is commonly used in the literature of network selection. We also have the same preference in our research because the calculation of the eigenvector is relatively simpler than solving an $n + 1$ system of linear equations in the weighted least square method, where n representing the number of attributes is usually large.

Unfortunately, the eigenvector method has an obvious problem while being used in the network selection procedure. As we know, pair-wise comparison matrices in this method are given by the DM based on his subjective feelings on the attributes, and the DM is usually human beings in most decision making processes. However, for the network selection issue, an automatic method is required because customers usually do not have the basic knowledge to construct the pair-wise comparison matrix. Moreover, the matrix changes in different situations (e.g. different applications and terminal properties), so customers do not want to be involved in the complicated pair-wise comparison process for each situation, even though they know how to do it. To sum up, when designing a network selection scheme, we should not suppose the customers to be the DMs who could provide pair-wise comparison matrix to calculate subjective weights. Furthermore, mobile terminals cannot be DMs either, because machines do not have subjective feelings on the attributes. In a word, evaluating the subjective weights in the network selection procedure is a tough work.

One possible approach to solve this problem is to do this work by the designer of the network selection scheme and to store the matrices for all scenarios in a ROM of the mobile terminal in advance. Imagining a mobile terminal which stores matrices of all the scenarios in its ROM, numerous factors should be considered, including application QoS levels, customer preferences, terminal properties, operator policies, dynamic network attributes, etc.

All the above factors contain two or more options (for example, terminal velocity could be divided into static, low speed, mid speed, high speed, etc.), so there are actually thousands of scenarios that a network selection scheme should consider. Furthermore, if the terminal is a multi-mode one which can only use one access at one time, the scheme should synthetically consider all the on-going applications to make the final decision, which leads to much more scenarios to consider. Therefore, we can see that it is not efficient to store in advance the pair-wise comparison matrices of all the scenarios into a ROM of the terminal by the scheme designer.

In order to evaluate the subjective weights in an efficient and automatic manner, we propose in the following sub-section a novel weighting method, named TRUST, which can be considered as an extended usage of the eigenvector (plus AHP) method, but only specific for the network selection issue.

5.5. TRUST

Once the network selection procedure is triggered, a subjective weighting method will be performed to calculate the subjective weights. The widely employed eigenvector method in research is actually not suitable for this task in practice, as we explained in the above sub-section. We emphasize here that the subjective weighting method for the network selection issue should satisfy at least the following conditions:

- subjective weights should be automatically obtained by the MT (or a network-side entity), not by customer’s pair-wise comparisons for each scenario;
- the procedure should be efficient and fast for obtaining appropriate weights in different scenarios.

As we know, the network selection procedure is usually triggered by the following events:

- new application comes or previous application ends;
- terminal property (e.g. velocity) obviously changes;
- customer or operator changes his preference;
- certain dynamic network-side attribute (e.g. traffic load) obviously changes, etc.

Now that the above events can trigger the network selection procedure, we wonder what kind of effect a given event actually brings into the network selection procedure. Table 5-1 shows the relationship between trigger events and changes of subjective weights. We can see that the content of this table is generally fixed, such as streaming applications require large bandwidth; conversational applications require low jitter; high-speed terminals require good handover capability. Therefore, in order to obtain the subjective weights $\mathbf{W}_S$, we just need to know the on-going events and their relative importance, then we will be able to evaluate $\mathbf{W}_S$ based on table 5-1.

Therefore, we define a k-by-n matrix **EA** as follows to represent the right part of table 5-1, where k is the total number of events and n is the number of attributes considered by the scheme:

$$\mathbf{EA} = \begin{bmatrix} c_{11} & c_{12} & . & c_{1n} \\ c_{21} & c_{22} & . & c_{2n} \\ . & . & . & . \\ c_{k1} & c_{k2} & . & c_{kn} \end{bmatrix}, \tag{5-17}$$

where c_{ij} represents the effect of the ith event on the jth attribute, and the value of c_{ij} is either TRUE (1) or FALSE (0).

The on-going events can be obtained by checking all the events one by one in the table. Since there are only a few dozens of events in the table, this process can actually be completed very quickly. By doing so, we obtain a diagonal matrix TF with k non-negative integer as follows:

$$\mathbf{TF} = \begin{bmatrix} tf_{11} & 0 & . & 0 \\ 0 & tf_{22} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & tf_{kk} \end{bmatrix}, \tag{5-18}$$

Table 5- 1 Relationship between trigger events and subjective weights.

Events & Weights				Attributes									
Layer-1		Layer-2		PR	BD	SC	PC	BER	JT	TL	HO	SS	...
Application QoS levels	0.37	Streaming	0.12		•								
		Conversational	0.16						•				
		Interactive	0.06			•		•					
		Background	0.03										
Terminal properties	0.21	Mid speed	0.03								•		
		High speed	0.12								•		
		Mid power	0.01				•						
		Low power	0.05				•						
Customer preferences	0.13	Low price	0.05	•									
		High security	0.05			•							
		Large bandwidth	0.03		•								
Operator preference	0.07	Load balancing	0.07							•			
Extreme dynamic attributes	0.22	Traffic load	0.07							•			
		Single strength	0.07									•	
		BER	0.03					•					
		Jitter	0.05						•				

where element tf_{ii} in this matrix represents whether the event i is currently true or not. For applications, this integer represents how many of this type of application are on-going at the moment of network selection.

The relative importance of these events is complicated to be obtained, but it is much easier than calculating the weights of attributes because weights of these events do not change frequently. For example, an operator of a telecommunication network providing some voice services feels that conversational application has higher importance than other applications; 'mid-speed' is less important than 'high-speed', although both of the two events require good handover capability; whether load balancing is more important than customer's preferences is decided by the operator's policy, etc. Therefore, the weights of these events will be calculated in advance and sent to the mobile terminal when the terminal is initiated by the customer. We use $\mathbf{W_E}$ to represent the weights of all the events in table 5-1, given by

$$\mathbf{W_E} = [we_1 \quad we_2 \quad \ldots \quad we_k]. \tag{5-19}$$

Eigenvector method (plus AHP) is suggested for calculating $\mathbf{W_E}$ by the scheme designer or the operator in advance. A hierarchy of several trigger events is formed in figure 5-1. There are two levels in the hierarchy: for the upper level, weights are obtained as

$$\mathbf{W_{E1}} = [we1_1 \quad we1_2 \quad \ldots \quad we1_{k1}], \tag{5-20}$$

and for the bottom level, weights of each group are obtained as

$$\mathbf{W_{E2i}} = [we2_{i,1} \quad we2_{i,2} \quad \ldots \quad we2_{i,k2}], \tag{5-21}$$

where i represents the ith group. The synthesized weights of the jth events in the ith group can be obtained as

$$we_{i,j} = we1_i \cdot we2_{i,j}\,. \tag{5-22}$$

Based on the preparation above, the procedure of TRUST is carried out as follows:

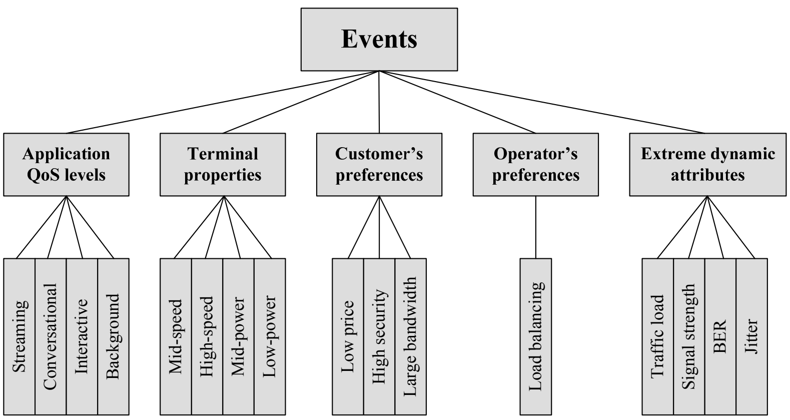


Figure 5- 1 Hierarchy of trigger events.

- In the mobile terminal, table 5-1 is stored. Suppose n network attributes and k events are considered for network selection, this table forms a k-by-n matrix **EA**;

- Since the relative importance of these events does not change frequently, their weights $\mathbf{W_E}$ can be calculated and transmitted to the terminal when it is initiated by the customer;
- When the network selection procedure is triggered, all the k events are checked to obtain a diagonal matrix **TF** with k non-negative integer as explained above;
- Finally, the subjective weights are obtained by

$$\mathbf{W_S} = \begin{bmatrix} ws_1 & ws_2 & \dots & ws_n \end{bmatrix} = \mathbf{W_E} \cdot \mathbf{TF} \cdot \mathbf{EA}, \qquad (5\text{-}23)$$

where the subjective weight of the jth attribute is

$$ws_j = \sum_{i=1}^{k} we_i \cdot tf_{ii} \cdot c_{ij}.$$

5.6. Performance comparisons

In this section, we compare the proposed method TRUST with the widely used eigenvector method in extensive scenarios. The weights of the eigenvector method are obtained by subjective pair-wise comparison matrices formed by the authors, while the weights of TRUST are calculated automatically by a simulation program using Matlab. 9 attributes are considered in the following comparisons, i.e. price (PR), bandwidth (BD), security (SC), power consumption (PC), bit error rate (BER), jitter (JT), traffic load (TL), handover properties (HO) and signal strength (SS). Three applications are used for comparison, i.e. streaming, conversational and interactive (the 3 columns in figure 5-2).

Table 5-1 is just a simple example of **EA** that is used in this paper to calculate the weights by TRUST. If this matrix is designed carefully, the weights obtained by TRUST can be much closer to the weights by the eigenvector method. Here we use a simple example of **EA** in order to show not only the consistency but also the difference of the two methods.

Table 5-2 shows weights of the 9 attributes obtained by the two methods in 24 scenarios. We can see that the main trends of these weights are the same for the two methods in various scenarios.

In order to further evaluate our proposed method TRUST, we calculate the *mean of difference* and the *correlation* of weights obtained by the two methods, as shown in figure 5-2. We can see that weights of some attributes in some scenarios are obviously different when calculated by the two methods, e.g. the weight of bandwidth for streaming applications, the weight of jitter for conversational applications and the weight of biter error rate for interactive applications. Meanwhile, we can see that the correlations of the first two attributes are low but the correlation of the third attribute is high. To further explain this problem, we draw specifically the three attributes' weights in three sub-figures (the last three sub-figures), and find that distribution of the third attribute's weights is linear, which is the reason that their correlation approaches 1. In the last three sub-figures, each figure (each application) contains 8 points representing 8 cases with different terminal properties, customer preferences and traffic states, as shown in table 5-2.

Actually, the above analysis finds out two unexpected phenomena of the TRUST method. One phenomenon is that the important attribute usually obtains a larger weight by the TRUST, which is because the unimportant attributes' weights are considered as 0s by this method. By contrast, weights of these attributes by the eigenvector method are usually tiny values. The adjustment is quite easy if the designer feels the unimportant attributes should be still somewhat considered. One possible solution is to insert a *virtual event* into table 5-1, which has a tiny weight (e.g. 0.02) but requires considering all the attributes.

Another phenomenon is that, sometimes, the difference of weights by the two methods is obvious but the correlation is good, such as the bit error rate of the interactive application in figure 5-2. This regular difference between the two methods is actually caused by the **EA** matrix. For example, interactive applications require large weights on both security and bit error rate and we suppose security is more important than bit error rate, but the **EA** matrix of TRUST in table 5-2 cannot take the two attributes' relative importance into account. One simple solution is to separate one event into two *semi-events* that correspond to the two important attributes.

By doing the above two adjustments, the obtained weights by TRUST can be quite close to the weights by the eigenvector method. Hence, TRUST is much more promising because it does not require complicated pair-wise comparisons.

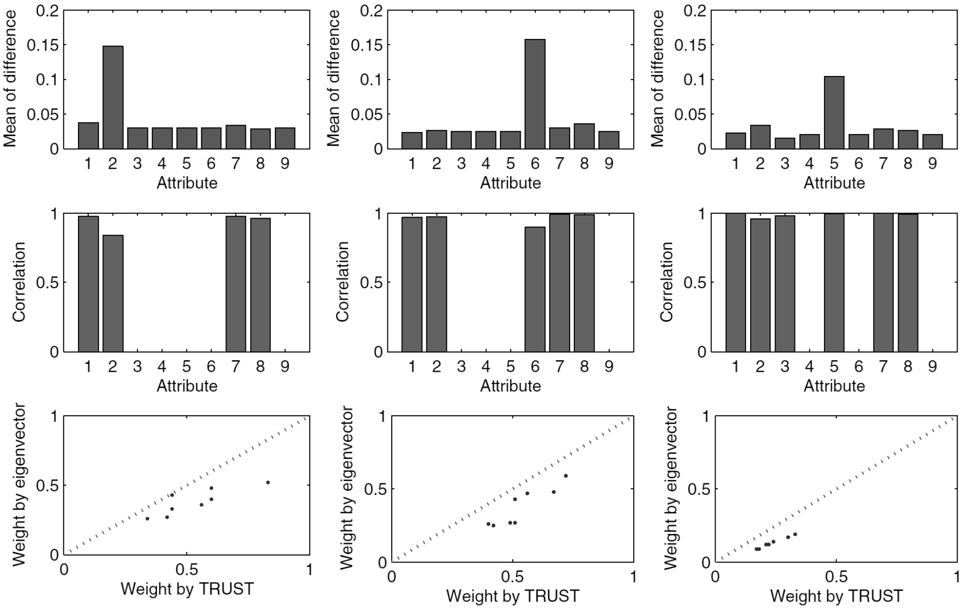


Figure 5- 2 Comparison results between TRUST and Eigenvector method.

Table 5- 2 Comparisons between TRUST and Eigenvector method.

QoS level	Terminal property	Customer preference	Dynamic attribute	Weighting method	PR	BD	SC	PC	BER	JT	TR	HO	SS
Streaming	Mid Speed	Low PR	Some networks	TRUST	0.19	0.44					0.26	0.11	
			with high traffic	Eigenvector	0.17	0.43	0.02	0.02	0.02	0.02	0.20	0.10	0.02
			All networks	TRUST	0.25	0.60						0.15	
			with low traffic	Eigenvector	0.20	0.48	0.03	0.03	0.03	0.03	0.03	0.14	0.03
		Large BD	Some networks	TRUST		0.60					0.28	0.12	
			with high traffic	Eigenvector	0.03	0.40	0.03	0.03	0.03	0.03	0.28	0.14	0.03
			All networks	TRUST		0.83						0.17	
			with high traffic	Eigenvector	0.04	0.52	0.04	0.04	0.04	0.04	0.04	0.20	0.04
	High speed	Low PR	Some networks	TRUST	0.14	0.34					0.18	0.34	
			with high traffic	Eigenvector	0.10	0.26	0.02	0.02	0.02	0.02	0.16	0.38	0.02
			All networks	TRUST	0.16	0.42						0.42	
			with low traffic	Eigenvector	0.11	0.27	0.03	0.03	0.03	0.03	0.03	0.44	0.03
		Large BD	Some networks	TRUST		0.44					0.21	0.35	
			with high traffic	Eigenvector	0.03	0.33	0.03	0.03	0.03	0.03	0.16	0.33	0.03
			All networks	TRUST		0.56						0.44	
			with low traffic	Eigenvector	0.04	0.36	0.04	0.04	0.04	0.04	0.04	0.36	0.04
Conversational	Mid Speed	Low PR	Some networks	TRUST	0.16					0.51	0.23	0.10	
			with high traffic	Eigenvector	0.17	0.02	0.02	0.02	0.02	0.43	0.20	0.10	0.02
			All networks	TRUST	0.21					0.67		0.12	
			with low traffic	Eigenvector	0.20	0.03	0.03	0.03	0.03	0.48	0.03	0.14	0.03
		Large BD	Some networks	TRUST		0.10				0.56	0.24	0.10	
			with high traffic	Eigenvector	0.02	0.11	0.02	0.02	0.02	0.47	0.21	0.11	0.02
			All networks	TRUST		0.14				0.72		0.14	
			with high traffic	Eigenvector	0.03	0.18	0.03	0.03	0.03	0.59	0.03	0.15	0.03
	High speed	Low PR	Some networks	TRUST	0.13					0.40	0.17	0.30	
			with high traffic	Eigenvector	0.10	0.02	0.02	0.02	0.02	0.26	0.16	0.38	0.02
			All networks	TRUST	0.15					0.49		0.36	
			with low traffic	Eigenvector	0.11	0.03	0.03	0.03	0.03	0.27	0.03	0.44	0.03
		Large BD	Some networks	TRUST		0.08				0.42	0.18	0.32	
			with high traffic	Eigenvector	0.02	0.13	0.02	0.02	0.02	0.25	0.13	0.39	0.02
			All networks	TRUST		0.10				0.51		0.39	
			with low traffic	Eigenvector	0.03	0.11	0.03	0.03	0.03	0.27	0.03	0.41	0.03
Interactive	Mid Speed	Low PR	Some networks	TRUST	0.19		0.22		0.22		0.26	0.11	
			with high traffic	Eigenvector	0.22	0.02	0.22	0.02	0.12	0.02	0.22	0.14	0.02
			All networks	TRUST	0.25		0.30		0.30			0.15	
			with low traffic	Eigenvector	0.28	0.02	0.28	0.02	0.17	0.02	0.02	0.17	0.02
		Large BD	Some networks	TRUST		0.12	0.24		0.24		0.28	0.12	
			with high traffic	Eigenvector	0.02	0.14	0.24	0.02	0.14	0.02	0.24	0.16	0.02
			All networks	TRUST		0.17	0.33		0.33			0.17	
			with high traffic	Eigenvector	0.02	0.19	0.31	0.02	0.19	0.02	0.02	0.21	0.02
	High speed	Low PR	Some networks	TRUST	0.14		0.17		0.17		0.19	0.33	
			with high traffic	Eigenvector	0.16	0.02	0.16	0.02	0.09	0.02	0.16	0.35	0.02
			All networks	TRUST	0.17		0.21		0.21			0.41	
			with low traffic	Eigenvector	0.19	0.02	0.19	0.02	0.12	0.02	0.02	0.40	0.02
		Large BD	Some networks	TRUST		0.09	0.18		0.18		0.20	0.35	
			with high traffic	Eigenvector	0.02	0.16	0.16	0.02	0.09	0.02	0.16	0.35	0.02
			All networks	TRUST		0.11	0.22		0.22			0.45	
			with low traffic	Eigenvector	0.02	0.19	0.19	0.02	0.12	0.02	0.02	0.40	0.02

6. MOBILITY SIGNALING COST EVALUATION AND MAP SELECTION

This chapter studies the signaling cost of terminal mobility which could be used as an important parameter of the previous discussed mobility-based network selection scheme. Moreover, a mobility anchor point (MAP) selection scheme is proposed for hierarchical MIPv6 networks.

6.1. Mobility signaling cost evaluation

6.1.1. Background on HMIPv6 signaling

In the mobility-based network selection schemes presented in chapter 4, cost of one handover (HHO or VHO) is assumed to be an available parameter. In order to obtain a generic expression, costs of HHO in network *i* and costs of VHO from network *i* to network *j* are represented by hho_i and $vho_{i,j}$ during the derivation. To use the proposed mobility-based network selection scheme, it is necessary to evaluate clearly the above HHO and VHO costs. However, these costs are related with many aspects:

- network structure, e.g. intra-MSC handover and inter-MSC handover.
- coupling manner, e.g. loose coupling and tight coupling.
- location management, e.g. MIPv6, route optimization, HMIPv6, PMIPv6, etc.
- handover scheme, e.g. soft handover, fast handover, etc.

Therefore, it is necessary to evaluate the exact costs based on the actual situations, so that the network selection scheme could select a network with really large benefit.

In this section, HMIPv6 is used as an example to show the evaluation of signaling cost of IP layer. This IP-layer signaling cost is named ‘mobility signaling cost’, which is composed of binding update messages sent to both HA and CNs. I would like to point out that this mobility signaling cost forms only a part of the total handover cost in mobility-based network selection schemes. hho_i and $vho_{i,j}$ should be more complicated, as explained above. In order to evaluate the mobility signaling cost of HMIPv6, let’s start from the signaling of MIPv6.

MIPv6 specification [JD04], allowing MNs to remain reachable while moving around in the IPv6 Internet, describes two modes depending on whether RO is used or not. When RO is not used, a MN registers its CoA to its HA with a BU message, so that its CNs could send packets to its HoA. All the packets are intercepted by the HA and delivered to the MN's CoA through a tunnel between the MN and its HA. In RO mode, the MN registers its CoA not only to the HA but also to all its CNs, so that packets between the MN and its CNs can be routed directly using its CoA. However, both of the two modes have obvious problems: non-RO mode has a triangular routing problem, and RO mode increases BU traffic all over the Internet. Besides, the BU latency to HA or CNs is usually large enough in either of the two modes to increase seriously handover latency.

Micro-mobility architectures [CA02] (e.g. HAWAII, Cellular IP, HMIPv6) could be used to decrease BU cost and handover latency. Among these architectures, people have paid a particular attention to HMIPv6 in recent years. In HMIPv6 [SH05], a MAP which serves as a local HA is used for micro-mobility. When a MN enters a Visited Network (VN), it first obtains a LCoA belonging to the prefix of the attached AR. Then, the MN registers this LCoA to a MAP and binds it to a RCoA belonging to the prefix of the MAP. After this local registration, only the RCoA is registered in the HA and CNs, while the LCoA is totally transparent to the outside of this MAP's coverage. When the MN performs a handover between ARs under the coverage of the same MAP, it only needs a local registration with that MAP, hence greatly decreasing BU cost and handover latency.

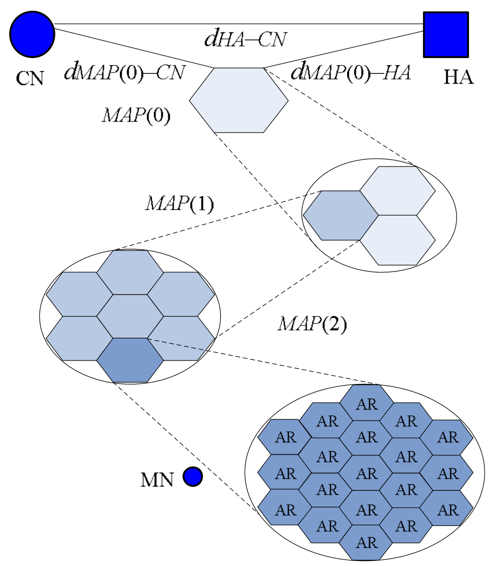


Figure 6- 1 A three-layer MAP model for HMIPv6 networks.

6.1.2. Defining a new parameter 'location rate'

MNs could have different mobility characteristics, so the capability of categorizing and distinguishing MNs' mobility patterns is essential for evaluating mobility signalling cost. In this section, the following types of MNs are considered:

- almost static,
- moving straight with certain velocity,
- randomly walking, and
- moving mainly within a relatively small region.

A new parameter is defined, called ‘location rate’, which could help to evaluate mobility signalling cost for MNs moving mainly within a relatively small region. Compared with terminal velocity, *location rate* is defined as the number of *different* ARs the MN has previously attached to in unit time. Therefore, an AR visited by the MN for several times appears in the AR list only once. Visited ARs’ addresses are recorded by the MN with an AR list which, by the way, requires a quite smaller storage than a group of MAP lists used in the scheme of [XY03]. The AR list is managed as follows:

- At the beginning of an observation period, the AR list is cleared, and the address of the currently attached AR is added into the list as the first entry;
- When the MN changes its attachment to a new AR, it checks if the new AR’s address is in the list. If not, a new entry is added;
- At the end of the observation period, the MN counts the number of entries in the list. The location rate is calculated as the number of entries in the list divided by the observation period.

6.1.3. Mobility signaling cost

The network model is shown in figure 6-1. Symbols are listed in table 6-1. Assumptions for the following derivations are given as follows:

- distances between the MN and the MAPs in the same layer are equal;
- distance between the MN and a MAP of ith layer is one hop larger than the distance between the MN and a MAP of (i –1)th layer: $d_{MN-MAP(i)} = d_{MN-MAP(i-1)} + 1$;
- all the packets between MN and HA are routed through the MAP, so $d_{MN-MAP(i)} + d_{MAP(i)-HA} = d_{MN-HA}$;
- the average number of ith layer MAPs the MN has visited, is the number of visited ARs divided by the square root of the number of ARs that an *i*th layer MAP averagely covers, shown in eq. (6-3).

Table 6- 1 Symbols in the proposal

$MAP(i)$	MAPs on the *i*th layer
d_{A-B}	Hop distance between A and B
T	Observation period
U_{AR}	Rate of the MN passes ARs
U_S	Session rate
P	Average number of packets per session
$U_{Location}$	Location rate
$U_{MAP(i)}$	Rate of the MN passes *i*th layer MAPs
L_{BU}, L_{BACK}	Length of BU and BACK message
L_{PT}	Length of tunneling header
$M(i)$	Number of ARs *i*th layer MAP averagely covers
$C(i)$	Mobility Signalling Cost of *i*th layer MAP
lowest	Level number of the lowest MAP

About the last assumption, the crossing rate of an AR could be expressed as [LX95]

$$U = vL / \pi S \qquad (6\text{-}1)$$

where v is the velocity, L is the boundary length, and S is the surface area of the cell. Therefore, the relationship between U_{AR} and $U_{MAP(i)}$ is $U_{MAP(i)} = U_{AR} / \sqrt{M(i)}$ [PS04] [BF94].

The above conclusion is based on the assumption that distributions of MNs' positions and moving directions are both uniform, as in RWMM. However, this is not precise enough to describe the real MNs' mobility characteristic. In reality, the probability of moving within a region (e.g. campus, sci&tech park, commercial district, residential area, etc.) is generally larger than the probability of moving out.

Considering the following three cases: first, when the MN is static or moving straight, it visits each AR only once, so we have $U_{Location} = U_{AR}$; second, when the MN walks randomly, the probability to re-visit a lot of former ARs is small, so we can assume $U_{Location} \approx U_{AR}$ in this case; third, when the MN is moving in a limited region, both the distribution density near the boundary and the probability to move against the centre are smaller than those of random walk, so $U_{Location}$ is more suitable than U_{AR} to be used to calculate $U_{MAP(i)}$. Based on the above analysis, the following relationship is used in our proposal:

$$U_{MAP(i)} = U_{Location} / \sqrt{M(i)} \,. \tag{6-2}$$

Table 6- 2 Costs of RO and Non-RO modes for one movement.

Non-RO mode	$C_{LBU}(i)$	$(L_{BU} + L_{BACK})d_{MN-MAP(i)}$
	$C_{GBU}(i)$	$(L_{BU} + L_{BACK})d_{MN-HA} + 2L_{PT}d_{MN-MAP(i)}$
	$C_{PT}(i)$	$U_S PL_{PT}(2d_{MN-MAP(i)} + d_{MAP(i)-HA})$
RO mode	$C_{LBU}(i)$	$(L_{BU} + L_{BACK})d_{MN-MAP(i)}$
	$C_{GBU}(i)$	$[(L_{BU} + L_{BACK})\overline{d_{MN-CN\&HA}} + 2L_{PT}d_{MN-MAP(i)}](1+N)$
	$C_{PT}(i)$	$U_S PL_{PT} \cdot 2d_{MN-MAP(i)}$

We summarize mobility signalling costs of non-RO and RO modes per movement in table 6-2 ($\overline{d_{MN-CN\&HA}}$ represents the average distance from a MN to its HA and all CNs, and N represents the number of CNs communicating with this MN). The mobility signalling cost in unit time is composed of LBU cost and GBU cost, written as:

$$C(i) = U_{AR} \cdot C_{LBU}(i) + U_{MAP(i)} \cdot C_{GBU}(i) \,. \tag{6-3}$$

6.2. MAP selection in HMIPv6 networks

6.2.1. Related work

However, HMIPv6 brings new cost. That is the increment of Packet Tunneling (PT) cost, as a tunnel between the MN and its MAP must be established to deliver packets. This PT cost becomes vital when the session rate becomes large, and could even results in increment of the total cost (consisting of BU and PT costs). A threshold was found to notify whether the usage of HMIPv6 could result in lower total cost or not [PS07].

An important issue of HMIPv6 is MAP selection [SH05]. Since the mobility signaling cost of HMIPv6 has been well evaluated in the last section, it is sure that this evaluated cost could be used for MAP selection. Thus, a MAP selection scheme leading to minimum mobility signaling cost can be found. And, compared with those MAP selection schemes that do not correctly evaluate (or even do not consider) this cost, it is not a surprise that the proposed scheme could have some obvious benefits. Therefore, a location history based MAP selection scheme is proposed in this section, and the performance is evaluated and compared with other schemes to show its advantages. Let's first review the existing schemes of MAP selection.

In HMIPv6 networks, a MAP is a local HA which processes not only BU registration but also encapsulation and decapsulation of tunneled packets. If only a single MAP is deployed in a HMIPv6 network, it could become a traffic bottleneck when the number of MNs becomes huge. Therefore, multiple MAPs are deployed, and they are on different layers on top of the ARs. The total cost [PS03] [PS04] and handover delay [LY07], when different MAPs are selected, could be considerably different. For example, the local BU cost and PT cost are both much smaller if a nearer MAP is selected rather than a further one, while the global BU cost could be huge if this MN moves fast. In a word, it's quite necessary to find out a good scheme to select preferred MAPs for different MNs.

When a MN enters a HMIPv6 network, it firstly selects the furthest MAP as default in order to avoid frequent re-registration. Then, it starts an observation period to collect the information for further MAP selection, and calculates the parameters of its preference at the end of the observation period in order that certain selection scheme could be used to select a better MAP.

In the HMIPv6 specification, two MAP selection schemes are recommended [SH05]. One is distance-based scheme, which always selects the furthest available MAP in order to avoid frequent global BU registration. However, this scheme is only suitable for MNs that continuously move across a large region. Besides, if all MNs use the furthest MAP, this MAP could become a bottleneck as we explained above. Another kind of schemes based on MN's preference is also recommended. This kind of schemes achieves several benefits, such as low handover delay, low traffic cost, load balancing, etc. However, it's not easy to select the optimal MAP, because it's difficult to find out the variable that could well describe the MN's preference. In previous schemes, the MN's velocity, mobility rate, session rate and MAPs' traffic load information are generally used. Those schemes are briefly described in the following paragraphs.

To begin with, some schemes take networks' current traffic loads into consideration in order to have load balancing benefit [HX05] [TT05] [CY06], but there are others which do not. In our opinion, as long as MNs have enough probability to select lower layer MAPs based on their preferences, there would be load balancing benefit even if traffic load is not considered. In mobility-based schemes [NE05] [KK04], velocity can be estimated by MN itself in different ways. When a MN enters a new MAP domain, its residence time in the previous MAP domain is the new BU time minus the old BU time, and the velocity is calculated as distance divided by the residence time. However, distance is quite difficult to estimate [CY06] [KK04], so some other methods are used to get the velocity information. In session-to-mobility ratio (SMR) based scheme [PS04], velocity (named mobility rate in this scheme) is calculated as the number of ARs visited by the MN in unit time. SMR is used for MAP selection because both the mobility rate and the session rate are important. If mobility rate is relatively large, a higher level MAP should be selected to decrease the global BU cost. On the other hand, if session rate is relatively large, a lower level MAP should be selected to decrease the local BU and PT cost. Actually, this SMR-based scheme has already achieved an optimal selection to minimize the total cost with Random Walk Mobility Model (RWMM).

Mobility rate is calculated as number of ARs the MN passes in unit time, but this is not enough to describe the MN's mobility feature, because the scope information, i.e. the span of the region where the MN is moving, is not contained. In the literature, some schemes were proposed to use scope information. In [XY03], all the MAPs are organized as a tree before selection. With this MAP tree, a list containing the MAPs ranging from the tree root to each AR is broadcasted to every MN which passes this AR. After passing several ARs in observation period, the MN records a bunch of MAP lists. Then, an intersection operation is used to find MAPs that exist in all received MAP lists, and the nearest of all these MAPs is finally selected. In [KT04], each MN holds a mobility history list, storing IP addresses of previously visited ARs. After a MN attaches to a new AR, it submits this mobility history list, so that the AR could select an appropriate MAP for it based on this information. The rule is to select the lowest MAP which could cover more than *X*-percent number of ARs in the mobility history list.

The above two schemes take the mobility scope information into account, but they have obvious drawbacks. Firstly, their requirements are rigorous: the first scheme requires every MN to have a large storage for a bunch of MAP lists; the second scheme requires ARs to know all the coverage information of their upper layer MAPs. Secondly, they don't simultaneously consider any other information, e.g. mobility rate or session rate. In order to find a better scheme, we use location history (to avoid confusion with mobility-based and SMR-based schemes, we don't use the term mobility history in this paper) information together with mobility rate and session rate to find the MAP with minimum total cost. A compound variable and a group of thresholds are derived for simplifying the selection procedure. Simulations based on a modified RWMM show that our scheme achieves good and steady performance in different situations, while other schemes are only suitable for certain cases.

6.2.2. Location history based MAP selection scheme

As we explained section 6.2.1, the usage of the HA brings in mainly two additional costs: BU cost and PT cost. MAPs in HMIPv6 decreases BU cost by changing Global BU (GBU) into Local BU (LBU), while PT cost increases, since a tunnel has to be established between the MN and its MAP. In table 6-2, PT cost is also formulated. Hence, the total cost, including both mobility signaling cost and PT cost can be written as

$$C(i) = U_{AR} \cdot C_{LBU}(i) + U_{MAP(i)} \cdot C_{GBU}(i) + C_{PT}(i)\,. \tag{6-4}$$

Lemma: *For certain HMIPv6 network structure, costs of MAPs on different layers obey the following two statements (a sufficient while not necessary condition is briefly deduced later):*

1) *if* $C(i-1) \geq C(i)$, *then* $C(i-x) \geq C(i)$ *for all* $x \in [1, i]$;

2) *if* $C(i+1) \geq C(i)$, *then* $C(i+x) \geq C(i)$ *for all* $x \in [1, lowest-i]$.

Prove of the Lemma: The two statements in the lemma are similar to each other, so we only derive the first one here. Combining eq. (6-2), eq. (6-4) and table 6-2, mobility signaling cost of $MAP(i-x)$ can be expressed as

$$C(i-x) = U_{AR}(C_{LBU(i)} + xL_{BU} + xL_{BACK}) + \frac{U_{Location}}{\sqrt{M(i-x)}}(C_{GBU}(i) + 2xL_{PT}) + (C_{PT}(i) + xU_S PL_{PT})\,. \tag{6-5}$$

Therefore, we have the following analysis:

$$C(i-1) \geq C(i) \Leftrightarrow U_{AR}(L_{BU} + L_{BACK}) + U_S PL_{PT} + \frac{U_{Location}}{\sqrt{M(i-1)}}(C_{GBU}(i) + 2L_{PT}) - \frac{U_{Location}}{\sqrt{M(i)}}C_{GBU}(i) \geq 0\,, \tag{6-6}$$

$$C(i-x) \geq C(i) \Leftrightarrow U_{AR}(L_{BU} + L_{BACK}) + U_S PL_{PT} + \frac{U_{Location}}{x\sqrt{M(i-x)}}(C_{GBU}(i) + 2xL_{PT}) - \frac{U_{Location}}{x\sqrt{M(i)}}C_{GBU}(i) \geq 0\,. \tag{6-7}$$

Considering the two inequations above, we can find a sufficient but not necessary condition as follows

Statement (1)

(6-8)

$$\Leftarrow \frac{U_{Location}}{x\sqrt{M(i-x)}}(C_{GBU}(i)+2xL_{PT})-\frac{U_{Location}}{x\sqrt{M(i)}}C_{GBU}(i)$$

$$\geq \frac{U_{Location}}{\sqrt{M(i-1)}}(C_{GBU}(i)+2L_{PT})-\frac{U_{Location}}{\sqrt{M(i)}}C_{GBU}(i)$$

$$\Leftrightarrow \left[\frac{1}{x\sqrt{M(i-x)}}-\frac{1}{\sqrt{M(i-1)}}+\frac{x-1}{x\sqrt{M(i)}}\right]\frac{C_{GBU}(i)}{2L_{PT}} \geq \frac{1}{\sqrt{M(i-1)}}-\frac{1}{\sqrt{M(i-x)}}.$$

Let $C_{GBU}(i)/2L_{PT}=nx$, where n is a generally much larger than 1. Take the network structure in figure 6-1 as an example, n could be around 10.

Let $M(i)=A_0$, $M(i-1)=A_0A_1$ and $M(i-x)=A_0A_1\cdots A_x$, where A_0 is the average number of ARs in an ith layer MAP's coverage, and $A_j\ \{1\leq j\leq x\}$ is the average number of (i–j+1)th layer MAPs in a (i–j)th layer MAP's coverage.

Taking n and $A_j\ \{1\leq j\leq x\}$ into the above inequation, we finally obtain

$$Statement\ (1) \Leftarrow \frac{1}{\sqrt{A_2A_3\cdots A_x}}(n+1)+\sqrt{A_1}(nx-n)\geq nx+1 \cdot \tag{6-9}$$

Moreover, statement (2) can be derived with a similar procedure, and the result is obtained as

$$Statement\ (2) \Leftarrow \sqrt{B_2B_3\cdots B_x}(n-1)+\frac{1}{\sqrt{B_1}}(nx-n)\geq nx-1, \tag{6-10}$$

where $B_j\ \{1\leq j\leq x\}$ represents the average number of (i+j)th layer MAPs in a (i+j–1)th layer MAP's coverage. Therefore, we finally obtain the two inequations above as a sufficient but not necessary condition for the lemma. ■

According to the lemma, we don't need to compare C(i) with all the other MAP's total costs. Instead, only C(i–1) and C(i+1) need to be compared with C(i) to find out the condition under which MAP(i) achieves minimum total cost. Let's take non-RO mode as an example.

According to eq. (6-4) and assumptions made in part. B, if MAP(i) is selected, while not MAP(i–1), their costs must satisfy

$$C(i-1)\geq C(i) \Leftrightarrow \tag{6-11}$$

$$\frac{U_{AR}(L_{BU}+L_{BACK})+U_S PL_{PT}}{U_{Location}} \geq \left(\frac{1}{\sqrt{M(i)}}-\frac{1}{\sqrt{M(i-1)}}\right)\times$$

$$\left[(L_{BU}+L_{BACK})d_{MN-HA}+2L_{PT}d_{MN-MAP(i)}\right]-\frac{2L_{PT}}{\sqrt{M(i-1)}}.$$

And, if MAP(i) is selected, while not MAP(i+1), we obtain

$$C(i+1) \geq C(i) \Leftrightarrow$$

(6-12)

$$\frac{U_{AR}(L_{BU}+L_{BACK})+U_S PL_{PT}}{U_{Location}} \leq \left(\frac{1}{\sqrt{M(i+1)}}-\frac{1}{\sqrt{M(i)}}\right)\times$$

$$\left[(L_{BU}+L_{BACK})d_{MN-HA}+2L_{PT}d_{MN-MAP(i+1)}\right]-\frac{2L_{PT}}{\sqrt{M(i)}}.$$

To sum up with eq. (6-5) and eq. (6-6), the compound variable is

$$Z=\frac{U_{AR}(L_{BU}+L_{BACK})+U_S PL_{PT}}{U_{Location}}, \quad (6\text{-}13)$$

and thresholds for a group of MAPs are written as eq. (6-14) below.

When a MN enters a HMIPv6 network, it selects the default MAP and starts an observation period. Thresholds could be calculated during the observation period, while the variable *Z* is calculated when the observation period is over. Then, the variable *Z* is compared with the thresholds to find the optimal MAP. The rule of our MAP selection is "*The LARGER the variable Z, the LOWER the MAP-layer*".

In RO mode, the form of variable *Z* is the same as eq. (6-13), but thresholds are a little more complex, because GBU should also be sent to a group of CNs, as show in eq. (6-15).

$$Thds=\left\{\left(\frac{1}{\sqrt{M(i)}}-\frac{1}{\sqrt{M(i-1)}}\right)\left[(L_{BU}+L_{BACK})d_{MN-HA}+2L_{PT}d_{MN-MAP(i)}\right]-\frac{L_{PT}}{\sqrt{M(i-1)}}\middle| 1\leq i\leq lowest\right\}$$

(6-14)

$$Thds\_RO= \quad (6\text{-}15)$$

$$\left\{\left(\frac{1}{\sqrt{M(i)}}-\frac{1}{\sqrt{M(i-1)}}\right)\left[(L_{BU}+L_{BACK})(d_{MN-HA}+N\overline{d_{MN-CN\&HA}})+2L_{PT}(N+1)d_{MN-MAP(i)}\right]\right.$$

$$\left.-\frac{L_{PT}(N+1)}{\sqrt{M(i-1)}}\middle| 1\leq i\leq lowest\right\}$$

6.2.3. Performance evaluation

In order to evaluate the performance of the proposed scheme and compare with previous schemes, we would like to modify the RWMM to fit for more types of mobility.

In the literature, RWMM is used to calculate the crossing rate of MAPs. The probabilities of moving toward and against the centre of a MAP region for one movement are given in [AI02]. However, the RWMM is not precise enough to describe the mobility characteristics of MNs, because it assumes that MNs are randomly moving. In reality, MNs would have more probability to stay within than to move out of a region (e.g. customers in campus, sci&tech park, commercial district, residential area, etc.), so this model is not suitable for MNs in reality.

Our modification to the RWMM is to multiply the probability of moving against the centre by a residence factor $\theta \in [l, 1]$, so the modified RWMM's probabilities become:

$$a_{r,r+1} = \begin{cases} 1 & \text{if } (r = 0) \\ \theta \cdot [1/3 + 1/(6r)] & \text{if } (1 \le r \le R), \end{cases} \tag{6-16}$$

$$b_{r,r-1} = 2/3 - \theta \cdot [1/3 + 1/(6r)] \quad \text{for all } (1 \le r \le R)\,, \tag{6-17}$$

where $a_{r,r+1}$ and $b_{r,r-1}$ are the probabilities of moving against and toward the centre respectively, and *R* represents the outmost circle of this MAP region. For a parallel study of this modification, please refer to [MI08] and [CV09]. The two publications provided quite similar but much more detailed studies, which could be used to further evaluate the novel MAP selection scheme's performance.

With the above equations, we can see that the probability of moving toward the MAP centre becomes larger, hence decreasing the probability of moving across MAPs. This fits for the reality that MNs prefer to move within a limited region, compared with the RWMM. The residence factor θ is not the same for MNs with different mobility preferences. Its floor limit *l* is an experimental parameter related with MNs' preferences and the network structure. In the following simulation, we set *l*=0.2, which is small enough to describe the MN who seldom moves out.

With this modified RWMM, crossing rate of MAP(2) is calculated as

$$U_{MAP(2)} = a_{2,3} Q_{2,2} U_{AR} = \frac{\frac{5}{24}\theta^2}{\frac{5}{24}\theta^2 - \frac{19}{36}\theta + \frac{10}{9}} U_{AR}\,, \tag{6-18}$$

where $Q_{x,y}$ is the probability of residence in circle-*x* of an area with *y* total circles. Crossing rates of *MAP*(1) and *MAP*(0) are calculated in the same way as in RWMM, respectively given by

$$U_{MAP(1)} = a_{1,2} Q_{1,1} U_{MAP(2)}\,, \tag{6-19}$$

$$U_{MAP(0)} = a_{0,1} Q_{0,0} U_{MAP(1)} \cdot 2/3\,. \tag{6-20}$$

The network structure we used for simulation is the same as figure 6-1. Parameters are shown in table 6-3, where t represents unit time. In our simulation, the novel scheme is compared with four other schemes: "furthest", "nearest", mobility-based and SMR-based. The "furthest" and the "nearest" schemes are selecting respectively the furthest and the nearest layer of MAPs, such as *MAP*(0) and *MAP*(2) in figure 6-1. The mobility-based scheme used for comparison in our simulation is the same as the

HMIP-UP algorithm [NE05], in which the velocity thresholds are uniformly distributed. Further adjustment of thresholds for mobility-based schemes is based on different factors [KK04], but the amelioration is limited according to our experience. As SMR-based scheme achieves optimal selection using RWMM, we calculate costs of different MAP layers based on this mobility model and find the one with minimum cost. Then, we calculate this MAP's actual mobility signalling cost in the modified RWMM for comparison.

In our scheme, addresses of ARs are recorded in an AR list, which makes our scheme not memoryless. To avoid complex network simulation, we still use mobility model to evaluate the performance, but assumptions should be made: intuitively, location rate is distributed in the interval $(0, U_{AR}]$, and its probability density function is related with θ. When θ is relatively small, the MN has larger probability to move toward the centre than to move against, so the probability of returning to a former AR is relatively large, in other words, the location rate is relatively small, vice versa. However, we can't know exactly the distribution of the location rate for certain HMIPv6 network. In our simulation, we simply assume that $U_{Location}$ obeys a uniform distribution within the interval of (θ –0.2, θ). To our experience, the simulation curves are not exactly the same if the distribution is changed, but they could generally show the same conclusions as discussed below.

The simulation results of mobility signalling costs obtained with different schemes are shown in figure 6-2. When the residence factor equals 1, SMR-based scheme is better than the others, because SMR-based scheme achieves an optimal selection for RWMM [PS03]. However, when the residence factor is small, SMR-based scheme's performance becomes poor. Compared with others, our scheme is good in different situations. When the residence factor is very small, which means MNs have more probability to move among a few ARs, our scheme's performance is the same as the nearest scheme and much better than the others. When the residence factor is very close to 1, which means MNs are almost randomly walking, the only scheme that is a little better than ours is the SMR-based scheme. Generally, residence factor should be neither very small nor very close to 1 for most MNs, so our scheme is the best choice for selecting the minimum-cost MAP, as illustrated by the middle segment of residence factor in figure 6-2(b).

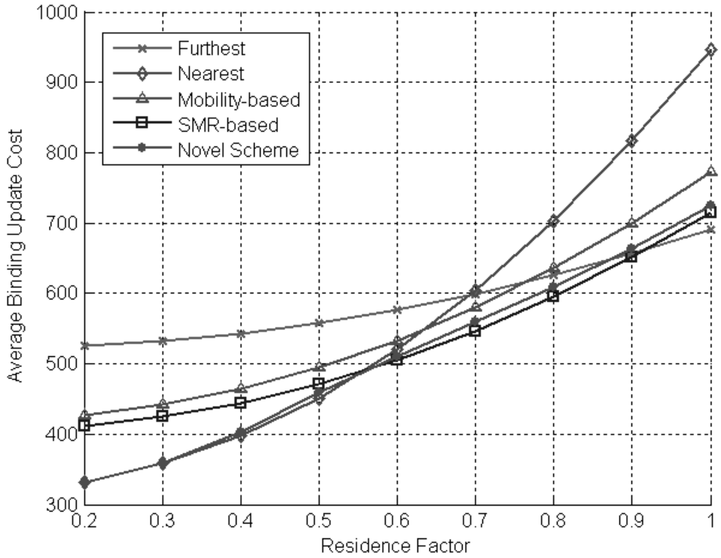
Besides mobility signalling cost, we also evaluate handover delay of the novel scheme. Average handover delay contains global and local BU delays. The two delays are calculated based on inter- and intra- handover procedures of HMIPv6, analyzed as in [LY07]. Their probabilities should be calculated based on eq. (6-18) eq. (6-19) and eq. (6-20) due to our modified RWMM. Layer 2 handover delay, wireless link transmission delay, one-hop transmission delay in wired link and router advertisement interval are 0.2, 0.5, 0.2 and 1 of unit time, respectively. Figure 6-3(a) shows comparison of average handover delays for different schemes. As we know, global registration incurs a large delay, so some mechanisms [MW04] have been proposed to change the global registration into a local registration to previous FA (PFA) or previous AR (PAR). In this paper, we also evaluate a scenario with RCoA registered to previous MAP (PMAP) to decrease handover delay. We can see from figure 6-3(b) that the handover delay in most schemes is decreased compared with figure 6-3(a), except for the "furthest" scheme as it doesn't bring any global registration.

Then, we evaluate these schemes with different SMRs. As shown in figure 6-4 and figure 6-5, we can see that our scheme is better than SMR-based scheme when the residence factor is small, no matter how much SMR is. When the SMR is large, the two schemes are both have more and more preference to select the nearest MAP.

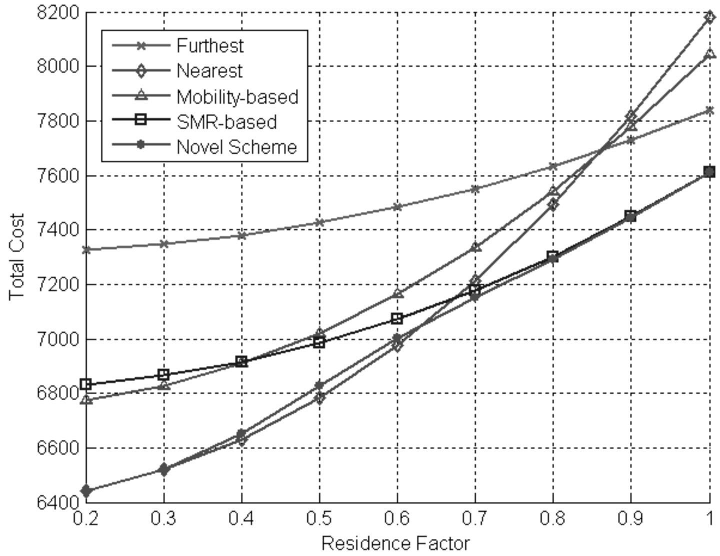
For further study on the MAP selection issue, we suggest: optimal length of observation period; strategy of initiating new selection procedures; storage cost (AR list vs. MAP list); computation cost (compound variable Z & thds vs. SMR & its thresholds). In our near future work, we are also going to improve our scheme with different optimizing targets, and extend this location history based scheme for agent selection of different multiple-agent scenarios (e.g. Proxy-MIP and Global-HAHA).

Table 6- 3 Main simulation parameters.

$d_{MAP(0)-CN}$	20 hops	L_{PT}	40 bytes
$d_{MAP(0)-HA}$	20 hops	U_{AR}	[0, 5] handoffs/t
d_{HA-CN}	20 hops	U_S	[0, 1] sessions/t
L_{BU}, L_{BACK}	52 bytes	P	10 packets/session

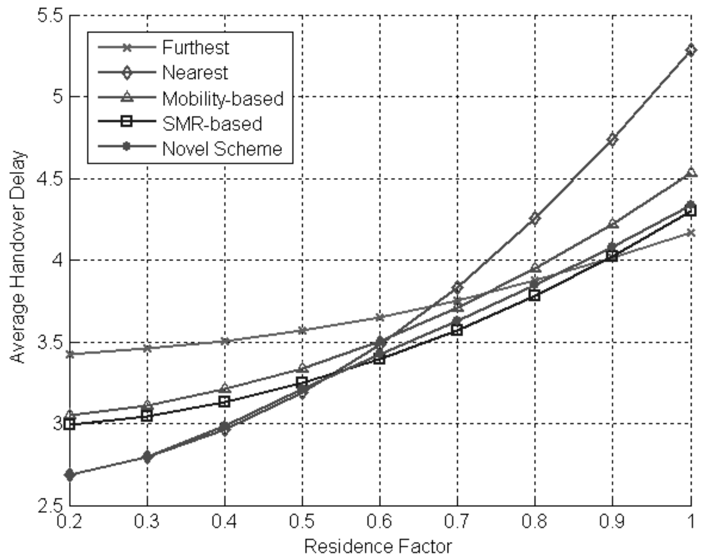


(a) in unit movement

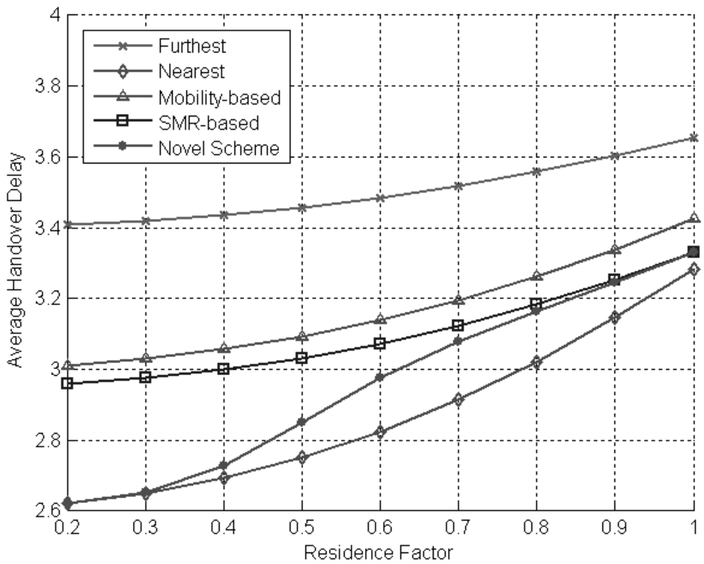


(b) in unit time

Figure 6- 2 Cost comparison with SMR = 0.2.

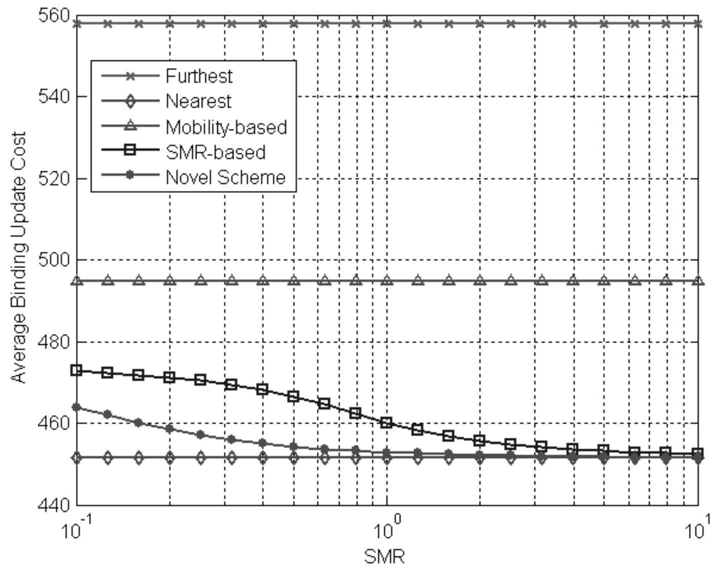


(a) without RCoA BU to PMAP

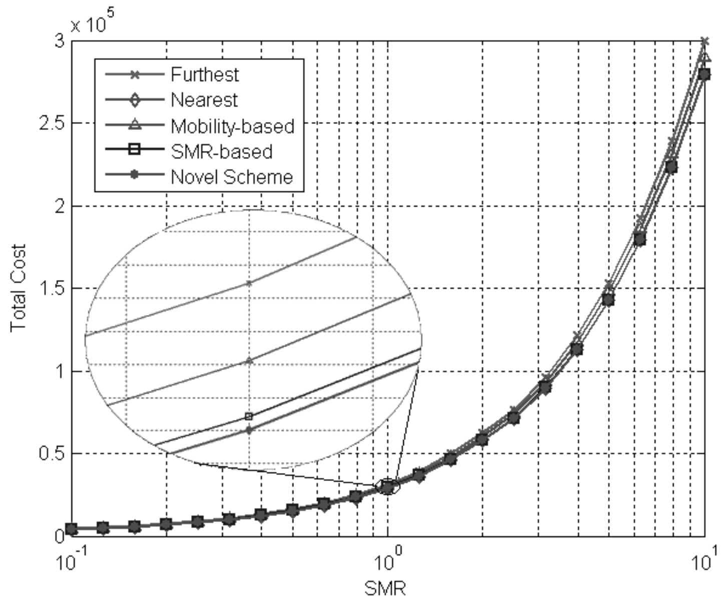


(b) with RCoA BU to PMAP

Figure 6- 3 Handover delay comparison in unit movement with SMR = 0.2.

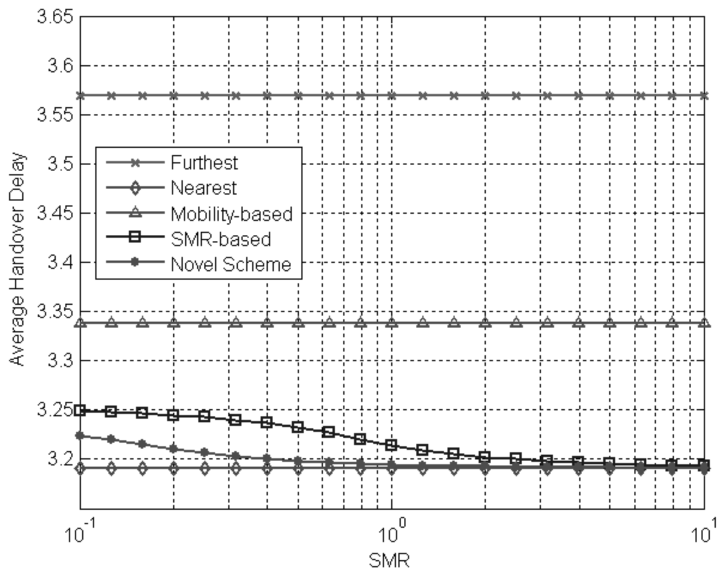


(a) in unit movement

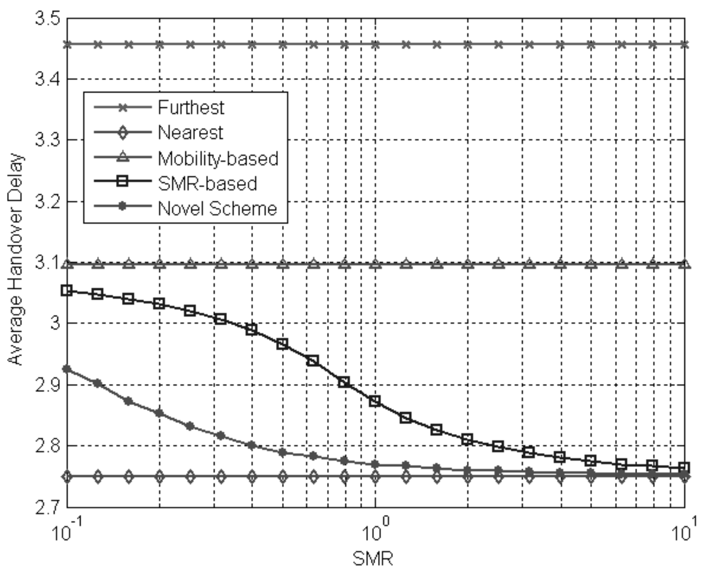


(b) in unit time

Figure 6- 4 Cost comparison with Residence Factor = 0.5.



(a) without RCoA BU to PMAP



(b) with RCoA BU to PMAP

Figure 6- 5 Handover delay comparison with Residence Factor = 0.5.

7. OTHER ISSUES ON NETWORK SELECTION

In this chapter, some other issues about network selection are discussed, including the usage of traffic load as a parameter of network selection, vertical handover decision, the way to integrated various proposals and network selection for mobile networks (NEMO).

7.1. Traffic load assignment during network selection

When selecting a network for a given mobile terminal, it is necessary to know the current traffic condition of all the available networks. Otherwise, a network without enough resource might be selected, which breaks the on-going applications and affects the usage of new applications. However, it is not easy to combine this information with others, because a network's traffic condition changes from time to time. Therefore, we will study how to usage the networks' traffic conditions to affect network selection result in this section.

According to the summary of WG IEEE P1900.4, there are the following forms of resource assigning and sharing:

- single operator: when the resource of one access network is mostly occupied, new users might not be able to get access through this network (RAN1). If another network (RAN2) has enough resource, it can temporarily borrow to RAN1. This new assignment of resource can be operated by a 'network reconfiguration management (NRM)' module of this operator.
- multiple operator with independent NRMs: when the two access networks belong to two operators with independent NRMs, the two NRMs should cooperate with each other to assign resource of the two networks.
- multiple operators with a centralized NRM: it is possible to use a centralized NRM between the two operators to manage their resource.
- terminal-side solution: in all the above solutions, resource of a network can be borrowed to another network, so that the latter could support more users during its peak period. Another obvious way to do this is to select the network with enough resource during the

network selection procedure, but this method requires the network selection scheme consider networks' traffic condition.

In this thesis, the idea to consider networks' current traffic load during the network selection procedure is called 'traffic load assignment', in order to distinguish with traffic load balancing. This is because the load is usually not balanced among all the networks, due to the fact that traffic load is only considered as one of multiple factors (usually not a decisive factor, except the preferred network has not enough resource). There are many methods to do traffic load assignment:

Using traffic load as a criterion: Seen from our simulation results in section 3.5, it is feasible to use traffic load values as a criterion in the ranking algorithm for traffic load assignment, and it works well for networks with similar performance. However, for networks with quite different performance, this method has an obvious immoderation problem. Considering two networks with both low but totally different traffic loads, normalization process will ignore the two networks' actual low traffic loads but retain only the relative large difference, which leads to immoderate traffic load assignment between the two networks and compromises the importance of other criteria. For example, we suppose there are two networks with traffic load of 0.1% and 10%, respectively. Intuitively, both 0.1% and 10% are small percentages (representing that a large amount of resource is still available), so it is not necessary to consider this factor during network selection. However, after the two values of the traffic load attribute pass through the normalization module, they becomes approximate 0% and 100%, respectively. This big difference of their normalized values will obviously lead to the usage of the first network, so flows will select it even though the latter has only 10% of its capacity occupied. This is called 'immoderate traffic load assignment' issue.

To solve this problem, I suggest not adjust traffic load values in the same way as other criteria. For example, we could use a special sigmoidal function $U(x) = x^{\eta}/(1+x^{\eta})$, $(\eta \geq 2)$ to calculate the utility, where x is the traffic load value and η is an experiential constant. Compared with eq. (4-1), this sigmoidal function has a centre of $x_m = 1$. Figure 7-1 shows the above function with different η. We could use a large value for η to avoid the immoderate traffic load assignment.

Let's check the example of two networks with 0.1% and 10% traffics again. Based on the above sigmoidal function with η = 10, we get the adjusted values as 0.0% and 0.0%, respectively. Therefore, this factor will not affect the selection of the best network. By contrast, if the traffics of the two networks are 10% and 90% respectively, the adjusted values will be 0.0% and 51.7%, which could affect network selection.

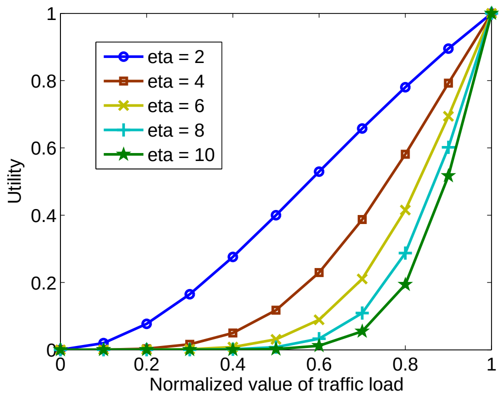


Figure 7- 1 Utility function for traffic load.

To sum up, using traffic load as a criterion of network selection is a feasible method, but the adjustment of this attribute should use a special sigmoidal function. This method overcomes the immoderate traffic load assignment issue.

Using knapsack or game model: this is another way to combine the traffic load information with other criteria. Load balancing can be achieved by the procedure of the model itself, while other criteria are used as the profit or payoff defined in the models. For more details, see section 2.3.5 and 2.3.6.

Independent operation: the above two methods use traffic loads by combining with other criteria, but it is also possible to use a separated operation. For example, we could check the traffic load of all networks before the network selection procedure and remove the ones without resource from the available network list. Or, we could check traffic load of networks in the available list from top to bottom after network ranking, so the first network with enough resource will be selected.

7.2. Vertical handover decision scheme

According to our study on mobility-based network selection, a network with small cell radius could lead to plenty of handovers (both horizontal and vertical). Therefore, an average handover cost was calculated by considering the network's coverage and the mobile terminal's mobility style based on assumptions of random walking terminal and randomly distributed cells. Then, the average handover cost was used as an attribute to represent the network's mobility support capability in chapter 4.

However, the above scheme does not consider prediction of the future events that the mobile terminal might encounter. Detailed speaking, when we find a better network with good mobility support capability, there is still the possibility that the terminal is about to leave the coverage of this network. If we could predict that the better network's availability lasts only a little time, we should not handover to this network. VHO tradeoff discussed in this section is exactly for this purpose.

Since a mobile terminal's movement is not regular, it is quite difficult to predict its movement during a long period. In other words, the most believable prediction results are those in a short time. Therefore, only a limited time of prediction for VHO tradeoff should be considered.

Supposing a terminal is using network C and network B is found better than network C at time 0, many predicted events might affect the VHO tradeoff, discussed one by one as follows

1) the better network is only a little bit better

Supposing the terminal finds a better network (B), but it is only a little bit better than the current network (C). Then, VHO is not suggested if

$$[U(B)-U(C)]\cdot T < VHO_{Cost}(C,B)\,, \qquad (7\text{-}1)$$

where T is the predictable time period, $U(i)$ represents the utility of network i and $VHO_{Cost}(i,j)$ represents the cost of one handover from network i to network j.

2) the better network might disappear soon

If network B is obviously better than network C, but it is predicted that network B will be available for only time *t*. In this case, if the terminal handover to network B, it will normally have to handover back to network C after time *t*, leading to two handovers in time *t*. Therefore, VHO is not suggested if

$$[U(B)-U(C)]\cdot t < VHO_{Cost}(C,B)+VHO_{Cost}(B,C)\,. \qquad (7\text{-}2)$$

3) a much better network might be available soon

If network B is obviously better than network C, and it could last for a long time, but we predict that a much better network (A) will be available in time *tt*. This means we will handover to network A in time *tt*, no matter we choose network B or not. Then, VHO is not suggested if

$$[U(B)-U(C)]\cdot tt < VHO_{Cost}(C,B)\,. \qquad (7\text{-}3)$$

The above three scenarios are relatively simple. There are much more complicated scenarios as follows:

4) a much better network (A) is predicted to be available in time tt, but it is not sure whether it is worth to handover to it.

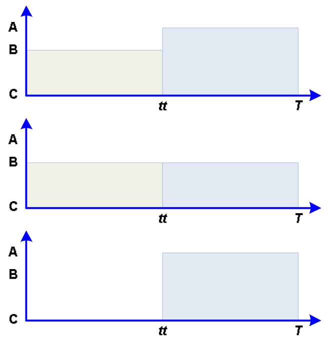


Figure 7- 2 VHO tradeoff scenario 4.

In this scenario, we suggest use the above figure to help analyzing the relationship between costs and utilities, where the colored area represents utilities. There are three possibilities:

a) the terminal handover to network B at time 0, then handover to network A at time *tt*;

b) the terminal handover to network B but NOT handover to network A at *tt*;

c) the terminal does NOT handover to network B, but handover to network A at *tt.*

The comparison between (a) and (b) is easy, which requires only to check whether it is worth to handover to network A at *tt*. Based on eq. (7-1), it is written by

$$[U(A) - U(B)] \cdot (T - tt) < VHO_{Cost}(B, A) . \qquad (7\text{-}4)$$

The comparison between (a) and (c) is also easy, which requires only to check whether it is worth to use network B for only *tt* time. Based on eq. (7-3), it is written by

$$[U(B) - U(C)] \cdot tt < VHO_{Cost}(C, B) . \qquad (7\text{-}5)$$

After the above two comparisons, if (a) is found better than either (b) or (c) (or both of them), the problem is solved. Otherwise, (a) is the worst case and we have to compare between (b) and (c). (b) and (c) both require only one handover, so the comparison is actually about the utility (area in the two figures), given by

$[U(B) - U(C)] \cdot tt < [U(A) - U(B)] \cdot (T - tt)$, for (c) better than (b). (7-6)

5) a much better network (A) is predicted to be available at time tt, but it is not sure whether it is worth to handover to it. Moreover, network B is going to be unavailable at time t (tt<t<T).

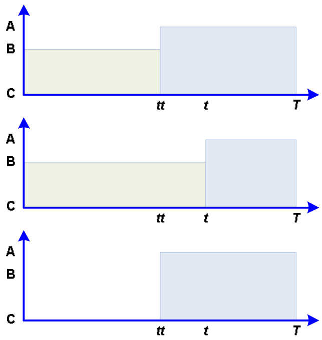


Figure 7- 3 VHO tradeoff scenario 5.

According to the figures above, it is obvious that (a) is better than (b), so we only compare between (a) and (c), given by

$$[U(B)-U(C)]\cdot tt < VHO_{Cost}(C,B). \tag{7-7}$$

This scenario seems simpler than the fourth one. That is because network B is predicted to be unavailable before time *T*.

6) a much better network (A) is predicted to be available at time tt, and is going to be unavailable at time t (tt<t<T), so it is not sure whether it is worth to handover to it.

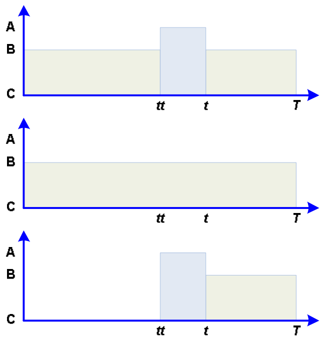


Figure 7- 4 VHO tradeoff scenario 6.

This scenario is a little bit complicated than those before, but the conditions can be easily obtained with similar analysis as above.

It is necessary to point out that the above scenarios cover only part of the possibilities. It is also possible that network B and A both become unavailable before *T*. Or, this tradeoff could involve more than three networks.

To sum up, there are quite a few scenarios that we should consider, and the number of scenarios increases dramatically when there are more networks.

7.3. An integrated strategy of network selection

In this sub-section, we propose an integrated strategy for MADM-based network selection, based on our study of the above issues, as shown in figure 7-5. The strategy contains four steps: the first step is to monitor the triggers and to gather the required information; the second step is the preparation before combining all the criteria, including weighting and adjusting of attributes; the third step is to combine multiple criteria based on certain MADM algorithm; and the last step is a VHO tradeoff algorithm. Further explanation on some key designs in our strategy is as follows:

Efficient subjective weighting: the proposed weighting method TRUST will be used for subjective weighting, while entropy method will be used for objective weighting. Then, the two groups of weights will be combined as explained in chapter 5.

Mobility-based network selection: as explained in chapter 4, mobility-based factors should be carefully considered in a network selection scheme. In our integrated strategy, both Besnet and Besper are used.

In Besnet, we get the best network by permutation-based pair-wise comparisons among all the networks, and VHO is performed immediately. Meanwhile, in Besper, we find the best permutation as the one with minimum total cost. The result of Besnet will be used for urgent selection, while the result of Besper for precise selection.

Moreover, when there are many access networks, we classify all the networks into several groups at the end of the adjusting module, in order to further decrease the time cost. This grouping operation is based on adjusted values of several most important criteria, e.g. cell radius, bandwidth, monetary cost, etc., which is reasonable according to our simulations in chapter 3.

VHO tradeoff scheme: as explained in section 7.2, a VHO tradeoff is required at the end of the network selection scheme. Networks in the available list will be checked one by one from the top to the bottom, the first network that passes the tradeoff will be selected. However, if no network passes the tradeoff, the current network continues to be used, which means no network is really worth to be handed-over to.

Traffic load assignment: traffic load information is used as a criterion of the MADM-based network ranking. For more information, see section 7.1.

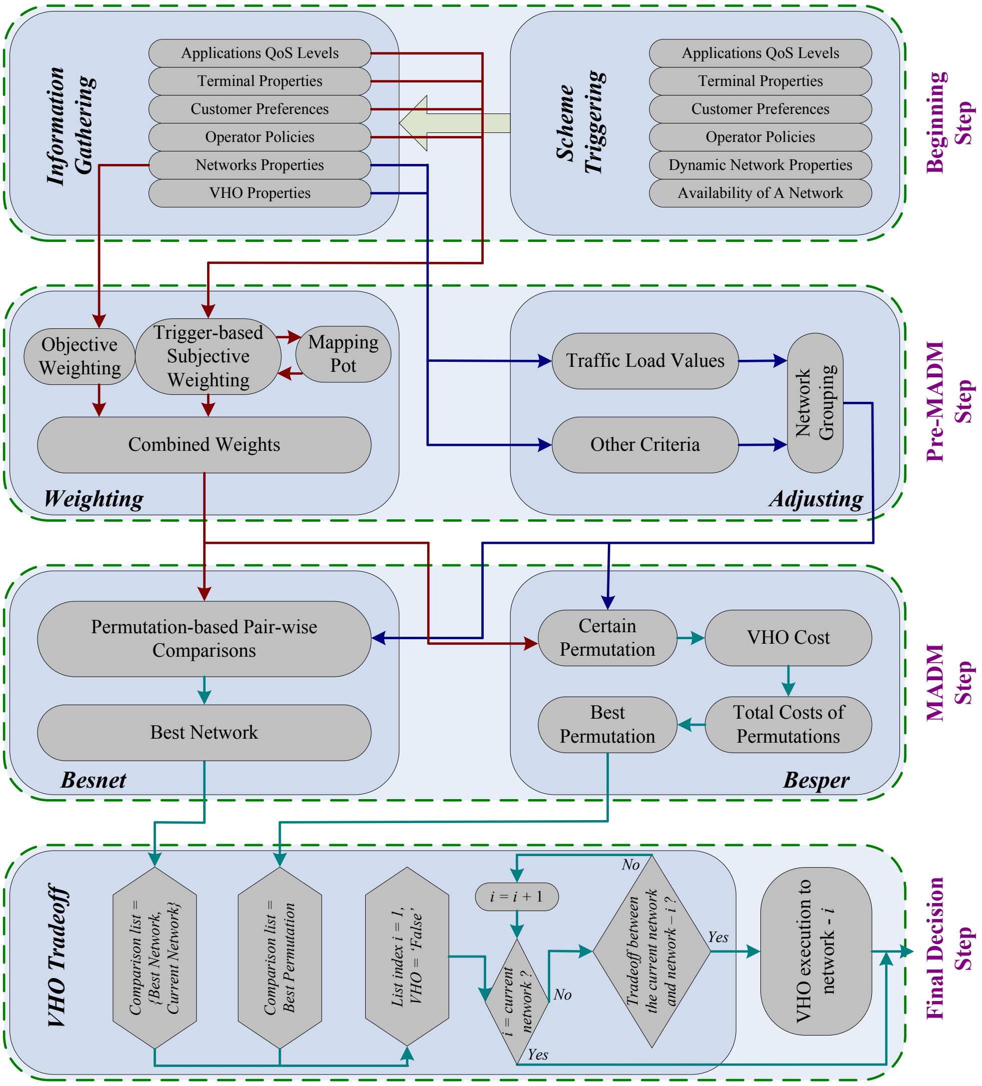


Figure 7- 5 An integrated strategy of network selection.

7.4. An analysis of network selection for NEMO

Mobility and multihoming of a NEMO can be totally transparent to the mobile network nodes in the NEMO. As shown in figure 7-6, the mobile router (MR) is in charge of the NEMO's mobility and some of multihoming issues (at least regarding access systems the MR is connected to). Besides, it connects to one or multiple access networks, and delivers packets between mobile network nodes (MNNs) and these access networks.

However, RFC3963 does not define any functionality of network selection on MR. Moreover, an MNN which has functionality of network selection can not decide the delivery of its packets through the MR based on current NEMO protocols. Therefore, it requires further study on how to do network selection for a NEMO. Here, we suggest a MR-side network selection scheme which can be simply deployed based on RFC3963.

Information gathering: similar to network selection for a terminal, information should be gathered for network selection. Here, all the information is gathered to the MR, including network properties, operator policies, NEMO mobility properties, application QoS requirements, etc. Among all the above groups of information, network-side ones can be delivered through access networks and NEMO mobility properties (e.g. velocity) can be detected by the MR itself with the help of some functions (e.g. GPS). By contrast, application QoS requirements are known by MNNs, and should be delivered to the MR within the NEMO. However, NEMO protocols have not defined this functionality, and we do not want to change too much of these protocols.

One possible method is to change the routing header by inserting just a few bits to indicate the packet's QoS requirement, called QoS indicator. This is to say that the MR has its network selection strategy for different QoS levels, and the only thing it needs to do is to check the packet's QoS indicator. In other words, the routing policy of the MR is different from a common router which delivers packets based on the destination address. By contrary, after the MR receives packets from a MNN, it delivers packets to different access networks based on the packet's source address and the QoS indicator. The source address is used to identify the MNN which might has specific user preferences. And, the QoS indicator is used to indicate the QoS requirement.

Advantages: this method is relatively simple, which does not need to increase signaling between the MR and its MNNs. The packet delivery process is fast, because the check of source address and QoS indicator in the routing header requires almost the same time as the check of destination address.

Disadvantages: the routing function of MR changes to source routing, which is a relatively big change.

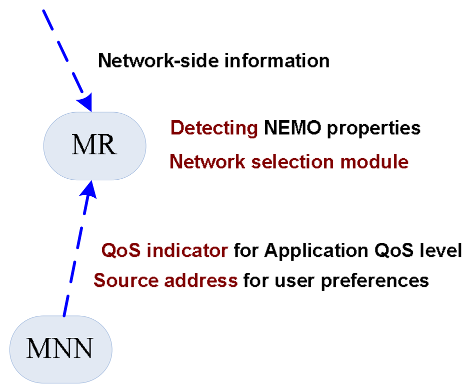


Figure 7- 6 MR-side network selection.

8. CONCLUSIONS

In this dissertation, the challenge of selecting the best network for a terminal or its applications was addressed. Current and future wireless access technologies bring us a heterogeneous wireless network environment, which contains multiple networks with different features. Thus, ABC becomes an important issue, in which network selection is one of its main components.

In our study, we firstly surveyed the existing network selection schemes. And, we established a simulator for MADM-based network selection and simulated vast scenarios to demonstrate the feasibility of using the MADM mathematical model for the network selection issue. Moreover, we found out several existing issues based on our simulations, including usage of mobility-related factors, requirement of efficient subjective weighting method, immoderate load balancing compromising other criteria during network selection, and VHO tradeoff for handing-over to the new best network after network ranking.

Secondly, we pointed out that the selection of the best network becomes the selection of the best permutation when VHO properties are taking into account. Then, we studied specifically on mobility-based network selection for scenarios with two and *n* groups of networks, respectively. For the scenario with two groups of networks, we studied the usage of mobility-related factors based on sigmoidal utility function, calculated horizontal and vertical handover costs, and derived a threshold between selections of the two groups. Simulations showed that it was reasonable to categorize all the networks into two groups when sigmoidal utility function was used. They also showed that our proposal could easily select the best group of networks based on mobility-related factors. Then, for the scenario with *n* groups of networks, we formulized the total handover cost, and proposed methods to obtain the best permutation of networks rapidly. We compared by simulations best network schemes (Besnets) and our proposed best permutation schemes (Bespers), and found that Bespers out-performed Besnets in many aspects, including total cost, VHO rate, trigger rate of the selection procedure, time cost for finding the best network, etc.

Thirdly, we explained that due to AHP method’s slow and complicated pair-wise comparison of attributes, it does not fit for calculating attributes’ subjective weights in the network selection issue. Hence, we proposed a trigger-based automatic subjective weighting method, called TRUST, which considers the relationship between trigger events and their effects on subjective weights. Compared with AHP subjective weighting, TRUST is a quick and efficient method to obtain similar subjective weights. Finally, we suggest combine the subjective weights obtained by TRUST and the objective weights obtained by Entropy method as the weights used in the network selection procedure.

Fourthly, we analyzed the signaling cost for mobility within HMIPv6 networks. This signaling cost can be used as an important mobility-related attribute in network selection schemes. A location history based MAP selection scheme was also proposed.

Finally, we studied and provided suggestions on several other issues of network selection, including traffic load assignment during network selection, VHO tradeoff for handing-over to the new best network. Based on all the above studies, an integrated network selection structure was proposed. Besides, we analyzed the possibility to use our strategy for NEMO, and mentioned a MR-side information gathering and network selection strategy.

To sum up, we have done a throughout study on the network selection issue. Several research articles are published in different international conferences in related domain. An initial study on mobility-based network selection with two groups of networks in section 4.2 was published in [WL09-3]. The best permutation for a generic scenario in section 4.3 was published in [WL09-1]. The novel subjective weighting method in chapter 5 was published in [WL09-4]. The simulative study in chapter 3 and the final integrated strategy was published in [WL09-2]. The location history based MAP selection scheme in chapter 6 was published in [WL08-3].

However, I would like to point out that many topics still require further study, including

- *appropriate utility function for traffic load values*: according to the analysis in section 7-1, traffic load should be adjusted by a specific utility function in the proposed integrated structure of network selection. Figure 7-1 gives some examples of utility functions for traffic load values. However, it is difficult to evaluate different utility functions and which function is the best one for traffic load is still unknown.
- *appropriate tradeoff function for VHO tradeoff*: in section 7-2, different scenarios are analyzed for VHO tradeoff. However, in a network selection scheme, it is inefficient to consider separately different scenarios. Hence, a generic VHO tradeoff decision function is required.
- *evaluation of mobility signaling cost and average handover cost*: it is important to know these costs for network selection, but it is quite complicated to evaluate them precisely. Chapter 6 shows a way to evaluate the mobility signaling cost in HMIPv6 networks, but more studies are required for more generic scenarios.

- *network selection for NEMO*: this topic is not widely studied, but there are really something interesting. First, in a NEMO, different information exists at different places (e.g. the customer's MT and the MR), so this gives new challenge to information gathering and network selection. Second, a NEMO could result in group handover, and network selection during group handover could cause some issues.

Moreover, it would be interesting to implement these network selection schemes on smartphones or laptops to do some tests in the current UMTS/WLAN environment.

APPENDIX

A. Network selection simulator

For our study on network selection, we need to establish a simulator to evaluate the performance of various schemes. This simulator should be able to simulate integratively consider multiple factors to make the decision on network selection. Moreover, it should be able to show the difference of selection between different specific cases.

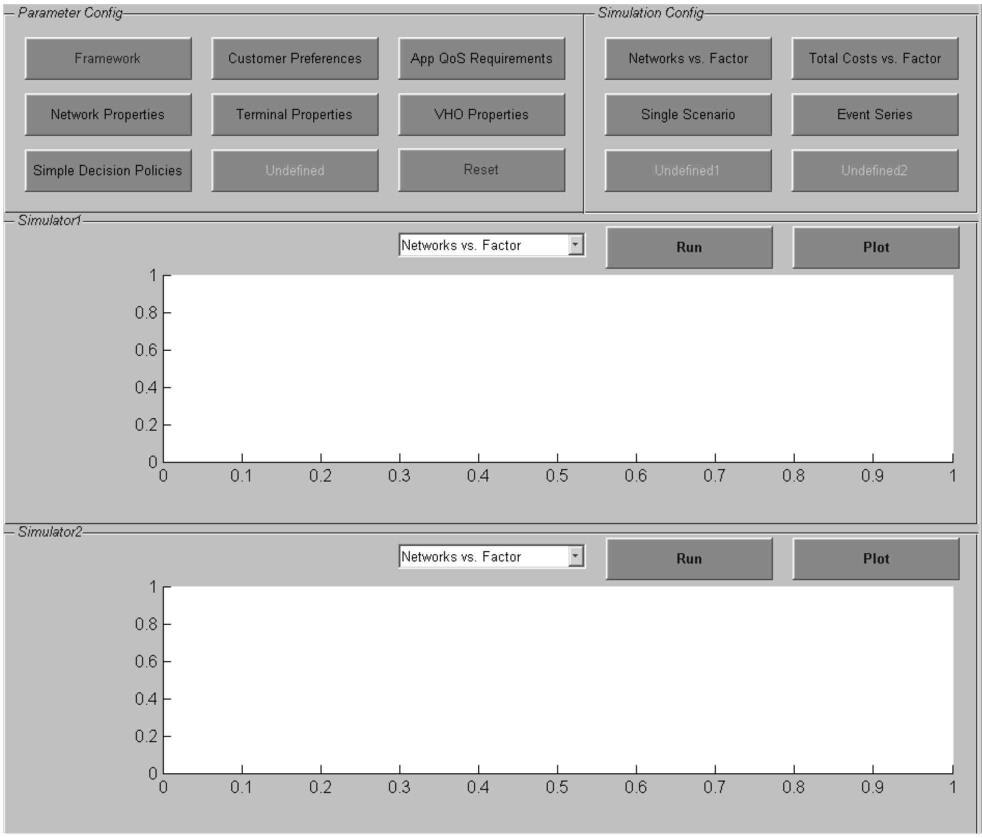


Figure A- 1 Main-board of our simulator.

Figure A-1 shows the main-board of our simulator by Matlab, which includes three parts: parameter config, simulation config and simulation results.

- The parameter config field is used to configure all the basic information that is required for making the decision, such as network-side information, terminal-side information, simulated network selection scheme, etc.

- The simulation config is to configure some information of the simulation, that is, what kind of simulation is required and what simulation result needs to be shown below.
- Simulation results part shows the simulation curves. There are two windows, so that you could easily compare different ideas. This part is only for a first view of the results, because most simulations are too complicated that we have to collect the results and draw figures separately. Therefore, all the simulation results will be stored in a file called *num_rslt.max* in numerical format, so that the user could draw different types of figures as required.

Before a simulation is run, there are many things you should configure. In this appendix, all the configurations are explained one by one as follows:

1) framework: the first button on the main-board opens the framework window, which shows the main procedure of the framework and some key options, as shown in figure A-2. These options decide what kind of scheme is simulated, including MADM algorithm, weighting, normalization, fuzzification, VHO tradeoff, etc.

- For MADM algorithms, we have realized SAW, MEW, GRA, TOPSIS, ELECTRE, etc.
- For weighting methods, we have Equality, Entropy, AHP, TRUST, etc.

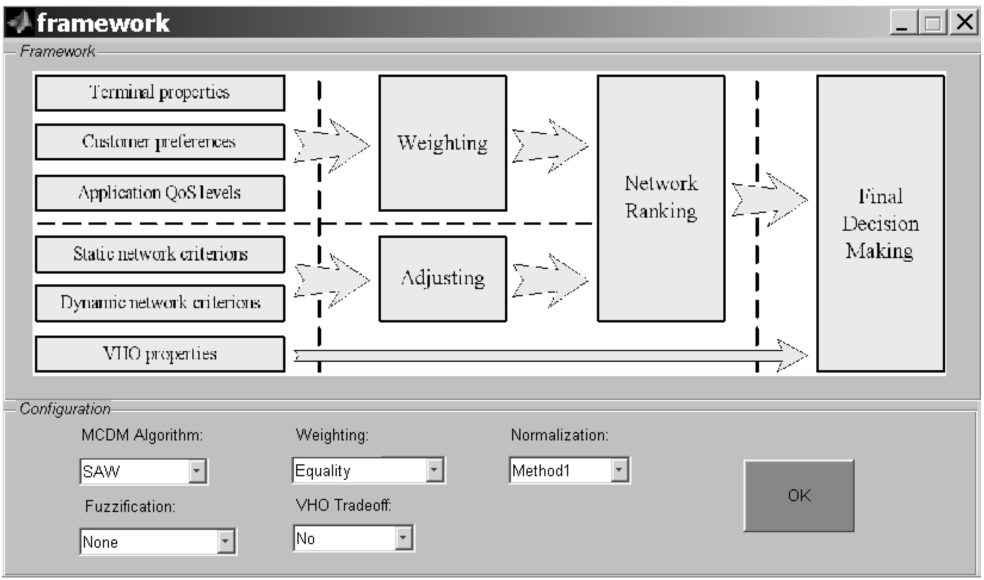


Figure A- 2 Framework window.

- For normalization, we have the methods as explained in chapter 2.
- For fuzzification and VHO tradeoff, you could select either use it or not.

It is worth mentioning that the simulation result will be the best network by network ranking algorithm if you do not use VHO tradeoff. Otherwise, the simulation result will tell you whether it is worth to handover to another network or not. During the study of our thesis, we have not complete the VHO tradeoff function, so simulations in this thesis only use the no VHO tradeoff option.

2) Network properties: this window is used to configure the available access networks and the considered criteria for network selection.

The button '*Factor value database*' brings you to the .m file containing the matrix of all the networks' attributes. This can be treated as an advanced configuration when the user wants to change the default values of networks' attributes.

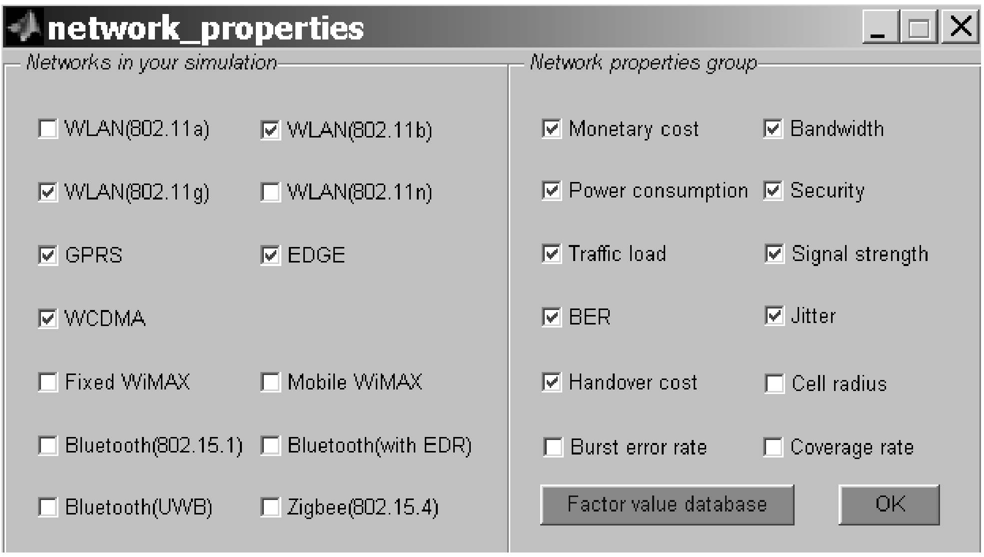


Figure A- 3 Network properties window.

3) Customer preferences: different customers might have different preferences, so this window provides you the chance to select the preferences of customers in your simulation.

It is worth mentioning that you should do corresponding configurations in other windows too. For example, if you are simulating a user who prefers *Low monetary cost*, you need make sure at least '*monetary cost*' is chosen as a considered attribute in the *Network properties window*. Otherwise, there is no difference between selection this option or not.

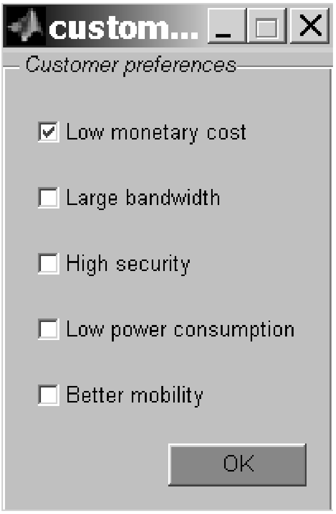


Figure A- 4 Customer preferences window.

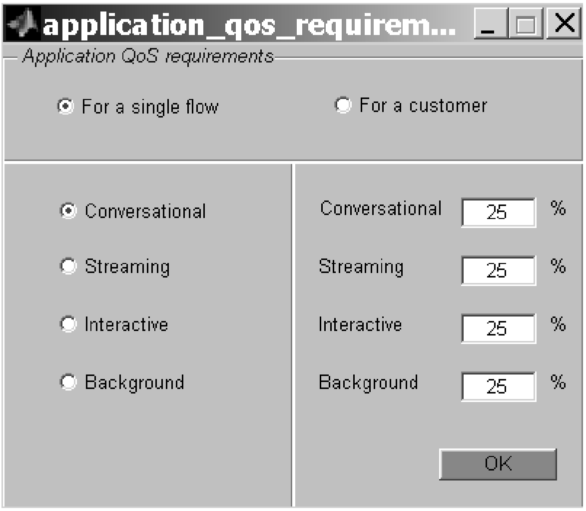


Figure A- 5 Application QoS requirements window.

4) Application QoS requirements: we divide all the applications into four levels (i.e. conversational, streaming, interactive and background). The simulation could be the selection of a network for a single flow or for a customer with multiple flows. For the latter case, you could freely decide the percentages of these applications, as shown in figure A-5.

5) Terminal properties: this window is used to configure the properties of the terminal, e.g. percentage of battery, velocity, etc., as shown in figure A-6. It is possible to add more options, such as the motion style of the terminal, but we did not have time to consider other motion style, except random walking, in our simulations during this thesis.

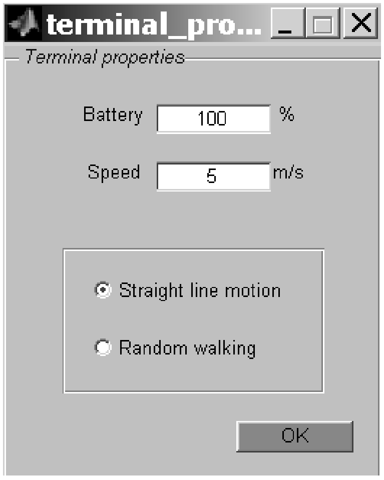


Figure A- 6 Terminal properties window.

6) Simulation_NF_gui (network vs. factor): if you want to know the change of the best network with regard to certain attribute, this simulation should be performed.

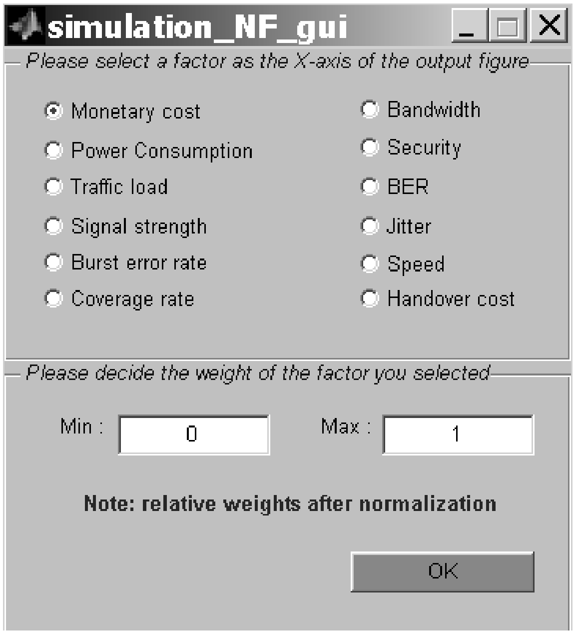


Figure A- 7 Network vs. factor simulation.

Configured by the window shown in figure A-7, this simulation could show us the change of the best network with respect to the value of certain network attribute. A simple example is given by the figure below, which shows the change of the best network from 'WiMax' to '802.11g' when the weight of the criterion 'monetary cost' becomes larger than 15.5%.

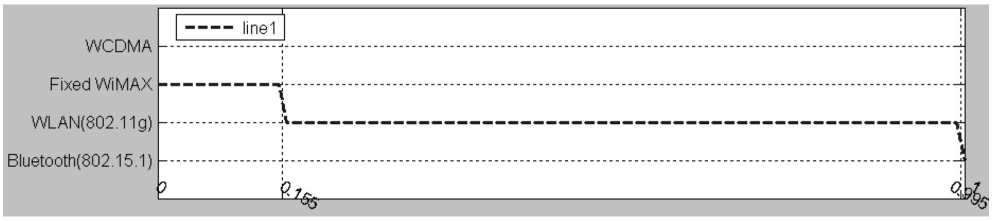


Figure A- 8 Simulation results of *Simulation_NF_gui*.

If you simulate and compare the results of different network selection schemes, there will be multiple curves drawn in figure A-8.

7) Simulation_TF_gui (coefficient vs. factor): if you want to know the change of total costs of different networks with regard to certain attribute, this simulation should be performed.

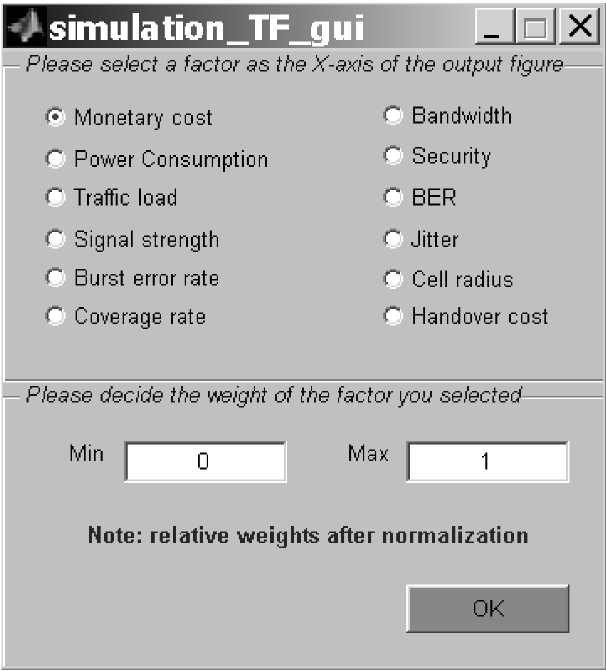


Figure A- 9 Coefficient vs. factor simulation.

With a similar configuration window to figure A-7, this simulation configuration window is shown in figure A-9. An example is shown in figure A-10, which draws the total costs of all the available networks and the switch point of the best network (15.5%).

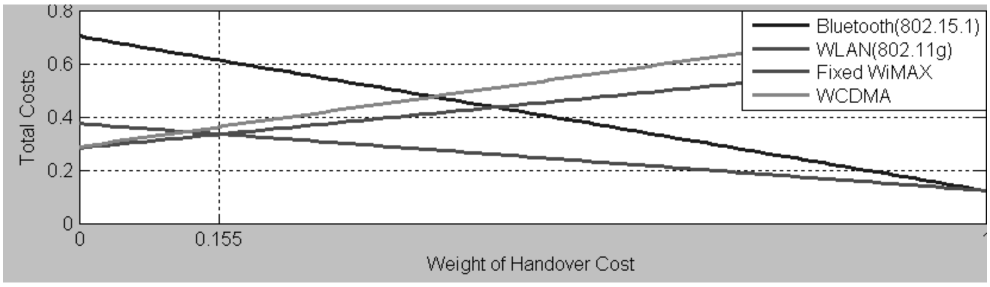


Figure A- 10 Simulation results of *Simulation_TF_gui*.

8) Simulation_SS_gui (single scenario): this simulation is used to show only the best network to you. It is used for some complicated scenarios when you want to know the best network.

9) Simulation_ES_gui (event series): this simulation is the key option provided by our simulator. The user could configure any series of events and see the change of the best network triggered by these events. The default events include change of signal strength, traffic load of certain network, application, battery, terminal velocity, customer preference, etc. You could easily configure the event time and the related value of an event. A usage of this simulation option is shown in section 3.4.

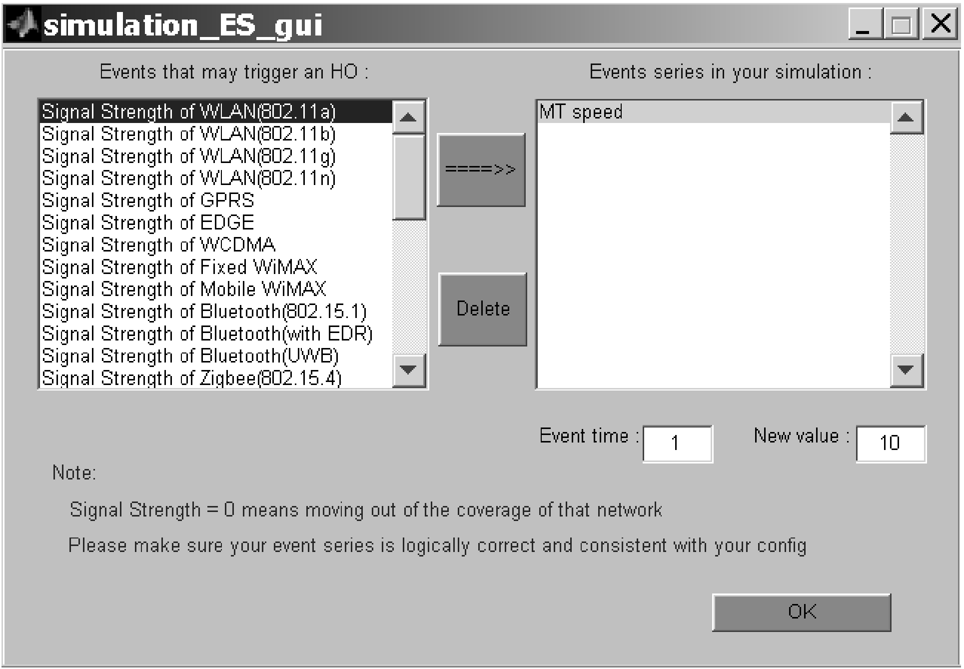


Figure A- 11 Event series simulation.

10) Advanced simulation options

Besides the above simulations, our simulator could do plenty of other simulations after only a few changes, such as the simulation of ‘best network vs. certain network’s coverage’, ‘best network vs. the terminal’s velocity’, ‘best network vs. traffic load’, etc. Examples of these simulations could be found in performance evaluations of our proposals in chapter 3, 4 and 5.

REFERENCES

[1900.4] IEEE P1900.4 standard working group

[3gpp] 3GPP, "Service and service capabilities (Release 8)," 3GPP TS 22.105 V8.2.0, Dec. 2006.

[802.21] IEEE 802.21 standard working group

[AG05] T. Al-Gizawi, K. Peppas, D. Axiotis, E. Protonotarios and F. Lazarakis, "Interoperability criteria, mechanisms, and evaluation of system performance for transparently interoperating WLAN and UMTS-HSDPA networks," *IEEE Network*, pp. 66 - 72, July-Aug. 2005.

[AI02] I. F. Akyildiz and W. Wang, "A Dynamic Location Management Scheme for Next-generation Multitier PCS Systems," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 178 - 189, Jan. 2002.

[AI04] I. F. Akyildiz, J. Xie and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," no. 4, vol. 11, pp. 16 – 28, Aug. 2004.

[AM05] M. Atiquzzaman and A. S. Reaz, "Survey and classification of transport layer mobility management schemes," in *proc. IEEE PIMRC 2005*, pp. 2109 – 2115, Sept. 2005.

[AM07] M. M. Alkhwlani and A. Ayesh, "Access network selection using combined fuzzy control and MCDM in heterogeneous networks," *International Conference on Computer Engineering & Systems 2007*, pp. 108 - 113.

[AT06] T. Ahmed, K. Kyamakya, M. Ludwig, "A context-aware vertical handover decision algorithm for multimode mobile terminals and its performance," *IEEE/ACM Euro American Conference on Telematics and Information Systems 2006*, pp. 19 - 28.

[BB07] B. Bakmaz, Z. Bojkovic and M. Bakmaz, "Network selection algorithm for heterogeneous wireless environment," *IEEE PIMRC 2007*, pp. 1 - 4.

[BC09] C. Bernardos, T. Melia and P. Seite, "Multihoming extensions for proxy mobile IPv6," draft-bernardos-mif-pmip-00, July 2009.

[BF07] F. Bari and V. C. M. Leung, "Automated network selection in a heterogeneous wireless network environment," *IEEE Network*, vol. 21, no. 1, pp. 34 - 40, Jan. - Feb. 2007.

[BF07-2] F. Bari and V. C. M. Leung, "Multi-Attribute Network Selection by Iterative TOPSIS for Heterogeneous Wireless Access," *IEEE CCNC 2007*, pp. 808 - 812.

[BF07-3] F. Bari and V. C. M. Leung, "Application of ELECTRE to Network Selection in A Hetereogeneous Wireless Network Environment," *IEEE WCNC 2007*, pp. 3810 - 3815.

[BF07-4] F. Bari and V. C. M. Leung, "Use of non-monotonic utility in multi-attribute network selection," *WTS 2007*, pp. 1 - 8.

[BF07-5] F. Bari and V. C. M. Leung, "Network selection with imprecise information in heterogeneous all-IP wireless systems," *International conference on Wireless Internet 2007*.

[BF94] F. V. Bauman and I. G. Niemegeers, "An evaluation of location management procedures," 3rd ICUPC, pp. 359 - 364, 1994.

[BS04] S. Balasubramaniam and J. Indulska, "Vertical handover supporting pervasive computing in future wireless networks," *Computer Communications*, 27 (8), 2004, 708 - 719.

[CA02] A. T. Campbell, J. Gomez, S. Kim, C. Y. Wan, Z. R. Turanyi and A. G. Valko, "Comparison of IP Micromobility Protocols," *IEEE Wireless Commun.*, vol. 9, pp. 72 - 82, Feb. 2002.

[CA02] A. T. Campbell, J. Gomez, S. Kim, C. Y. Wan, Z. R. Turanyi and A. G. Valko, "Comparison of IP micromobility protocols," IEEE Wireless Communications, no. 1, vol. 9, pp. 72 – 82, Feb. 2002.

[CD08] D. Charilas, O. Markaki and E. Tragos, "A theoretical scheme for applying game theory and network selection mechanisms in access admission control," *ISWPC 2008*.

[CD08] D. Charilas, O. Markaki, D. Nikitopoulos and M. Theologou, "Packet-switched network selection with the highest QoS in 4G networks," *Computer Networks*, vol. 52, no. 1, pp. 248 - 258, Jan. 2008.

[CHH92] S. J. Chen, C. L. Hwang and F. P. Hwang, "Fuzzy multiple attribute decision making: methods and applications," *Springer-Verlag 1992*.

[CJ06] J. Chen, K. Yu, Y. Ji, and P. Zhang, "Non-cooperative distributed network resource allocation in heterogeneous wireless data networks," *IST Mobile Summit 2006*.

[CP01] P. Chan, R. Sheriff, Y. Hu, P. Conforto and C. Tocci, "Mobility management incorporating fuzzy logic for a heterogeneous IP environment," *IEEE Communications Magazine*, 39 (12), 2001, 42 - 51.

[CP02] P. Chan, Y. Hu and R. Sheriff, "Implementation of fuzzy multiple objective decision making algorithm in a heterogeneous mobile environment," *IEEE WCNC 2002*, pp. 332 - 336.

[CT02] T. Camp, J. Boleng and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communication & Mobile Computing*, vol. 2, no. 5, pp. 483 - 502, Sept. 2002.

[CV09] V. Casares-Giner and P. Escalle-Garcia, "On movement-based location update. A lookahead strategy," *5th NGI*, July 2009.

[CX07] X. Cai, L. Chen, R. Sofia and Y. Wu, "Dynamic and User-Centric Network Selection in Heterogeneous Networks," *IEEE IPCCC 2007*, pp. 538 - 544.

[CY06] Y. W. Chen and M. J. Huang, "A Novel MAP Selection Scheme by Using Abstraction Node in Hierarchical MIPv6," *IEEE ICC 2006*, pp. 5408 - 5413.

[CY07] Y. Chen and Y. Yang, "A new 4G architecture providing multimode terminals always best connected services," *IEEE Wireless Commun.*, vol. 14, no. 2, pp. 36 - 41, April 2007.

[DV05] V. Devarapalli, R. Wakikawa, A. Petrescu and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol," *RFC 3963*, Jan. 2005.

[DV06] V. Devarapalli, R. Wakikawa and P. Thubert, "Local HA to HA protocol," *draft-devarapalli-mip6-nemo-local-haha-01.txt*, March 2006.

[Erxn] Technical overview and performance of HSPA and Mobile WiMAX, *Ericsson whitepaper*, Jan. 2009.

[FD09] D. Farinacci, V. Fuller, D. Meyer and D. Lewis, "Locator/ID separation protocol (LISP)," draft-farinacci-lisp-12.txt, March 2009.

[GE03] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Commun.*, 10(1):49 - 55, Feb. 2003.

[GE03] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Commun.*, vol. 10, no. 1, pp. 49 - 55, Feb. 2003.

[GE03] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Commun.*, vol. 10, no. 1, pp. 49 – 55, Feb. 2003.

[GH05] G. Huston, "Architectural approaches to multihoming for IPv6," draft-ietf-multi6-architecture-04.txt, August 2005.

[GQ05] Q. Guo, J. Zhu and X. Xu, “An adaptive multi-criteira vertical handoff decision algorithm for radio heterogeneous network,” *IEEE ICC 2005*, pp. 2769 - 2773.

[GS08] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury and B. Patil, “Proxy Mobile IPv6,” RFC 5213, Aug. 2008.

[GV05] V. Gazis, N. Alonistioti and L. Merakos, “Toward a generic ‘always best connected’ capability in integrated WLAN/UMTS cellular mobile networks (and beyond),” *IEEE Wireless Commun.*, vol. 12, no. 3, pp. 20 - 29, June 2005.

[GV05] V. Gazis, N. Alonistioti and L. Merakos, “Toward a generic ‘always best connected’ capability in integrated WLAN/UMTS cellular mobile networks (and beyond),” *IEEE Wireless Commun.*, vol. 12, no. 3, pp. 20 – 29, June 2005.

[HA08] A. Hasib and A. O. Fapojuwo, “A mobility model for heterogeneous wireless networks,” *IEEE RWS 2008*, pp. 815 - 818.

[HJ01] J. Hendricus van Lint and R. M. Wilson, “A course in combinatorics: Chapter 10,” Cambridge University Press, 2001.

[HJ06] J. Hou and D. C. O’Brien, “Vertical handover decision making algorithm using fuzzy logic for the integrated Radio-and-OW system,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 176 - 185, Jan. 2006.

[HP06] P. Houze, S. Ben Jemaa and P. Cordier, “Common Pilot Channel for network selection,” *IEEE VTC 2006 Spring*, pp. 67 - 71.

[HPUX] HP-UX Mobile IPv6 A.01.00 Administrator’s Guide, http://www.docs.hp.com.

[HT05] T. Hain, “A pragmatic report on IPv4 address space consumption,” *The Internet Protocol Journal CISCO*, vol. 8, no. 3, Sept. 2005.

[HX05] X. Hu, J. Song and M. Song, “An Adaptive Mobility Anchor Point Selection Algorithm for Hierarchical Mobile IPv6,” *IEEE ISCIT 2005*, pp. 1148 - 1151.

[HY81] C. L. Hwang and K. Yoon, “Multiple attribute decision making methods and applications,” *Springer-Verlag*, 1981.

[IJ07] I. Joe, W. T. Kim and S. Hong, “A Network Selection Algorithm considering Power Consumption in Hybrid Wireless Networks,” *IEEE ICCCN 2007*, pp. 1240 - 1243.

[JA07] J. Antoniou and A. Pitsillides, “4G Converged Environment: Modeling Network Selection as a Game,” *16th IST Mobile and Wireless Communications Summit 2007*, pp. 1 - 5.

[JD04] D. Johnson, C. Perkins and J. Arkko, “Mobility Support in IPv6,” *IETF RFC 3775*, June 2004.

[JD04] D. Johnson, C. Perkins and J. Arkko, “Mobility Support in IPv6,” *RFC 3775*, June 2004.

[JH06] H. Jia, Z. Zhang, P. Cheng, H. H. Chen and Shiju Li, “Study on Network Selection for Next-Generation Heterogeneous Wireless Networks,” *IEEE PIMRC 2006*, pp. 1 - 5.

[JH07] H. Jia, Z. Zhang, G. Yu and S. Li, “A Distributed Call Admission Control and Network Selection Scheme for Hybrid CDMA-OFDMA Networks,” *IEEE WCNC 2007*, pp. 3774 - 3779.

[JW08] W. Jiang, H. Cui, Q. Yan, X. Wang and J. Chen, “A novel application-oriented dynamic network selection in an integrated UMTS and WiMAX environment,” *ChinaCom 2008*, pp. 158 - 161.

[KH06] H. J. Kim, Y. G. Hong and S. H. Kim, “IPv6 multihoming for NGN,” ICACT 2006, pp. 2144 – 2147.

[KK04] K. Kawano, K. Kinoshita and K. Murakami, “Multilevel Hierarchical Mobility Management Scheme in Complicated Structured Networks,” *IEEE LCN 2004*, pp. 34 - 41.

[KM08] M. Kassar, B. Kervella and G. Pujolle, “An overview of vertical handover decision strategies in heterogeneous wireless networks,” *Comput. Commun.*, vol. 31, no. 10, pp. 2607 - 2620, June 2008.

[KM08-2] M. Kassar, B. Kervella and G. Pujolle, “An Intelligent Handover Management System for Future Generation Wireless Networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, no. 6, pp. 1 - 12, Jan. 2008.

[KR05] R. Koodli, “Fast handovers for mobile IPv6,” *RFC 4068*, July 2005.

[KS05] S. Kher, A. K. Somani and R. Gupta, “Network selection using fuzzy logic,” *International Conference on Broadband Networks 2005*, 3-7 Oct. pp. 876 - 885.

[KT04] T. Kumagai, T. Asaka and T. Takahashi, “Location Management Using Mobile History for Hierarchical Mobile IPv6 Networks,” *IEICE Trans. Commun.*, vol. E87-B, no. 9, pp. 2567 - 2575, Sept. 2004.

[LA07] H. Labiad, H. Afifi and C. De Santis, "Wi-Fi, Bluetooth, Zigbee and WiMAX," *Springer*, 2007.

[LH07] H. Lin and H. Labiod, "Hybrid handover optimization for multiple mobile routers-based multihomed NEMO networks," in *proc. IEEE International Conf. Pervasive Services 2007*, pp. 136 – 144, July 2007.

[LI08] I. Lassoued J. M. Bonnin and A. Belghith., "Towards an architecture for mobility management and resource control," *IEEE WCNC 2008*, pp. 2846 - 2851.

[LP08] P. Lescuyer and T. Lucidarme, "Evolved packet systems (EPS)," *John Wiley & Sons*, 2008.

[LX95] X. Lagrange and P. Godlewski, "Teletraffic analysis of a hierarchical cellular network," *IEEE VTC 1995*, pp. 882 - 886.

[LY07] Y. X. Lei and G. S. Kuo, "Impact of MAP Selection on Handover Performance for Multimedia Services in Multi-level HMIPv6 Networks," *IEEE WCNC 2007*, pp. 3901 - 3906.

[MD06] D. Mariz, I. Cananea, D. Sadok and G. Fodor, "Simulative Analysis of Access Selection Algorithms for Multi-Access Networks," *International Symposium on on World of Wireless, Mobile and Multimedia Networks 2006*, pp. 219 - 227.

[MG02] S. G. Matthew, "802.11 wireless networks : the definitive guide," O'Reilly, 2002.

[MH03] H. Matsuoka, T. Yoshimura and T. Ohya, "End-to-end robust IP soft handover," in *proc. ICC 2003*, pp. 532 – 536, May 2003.

[MI08] Ignacio Martinez-Arrue, Pablo Garcia-Escalle and Vicente Casares-Giner, "Location management based on the mobility patterns of mobile users," *Wireless Systems and Mobility in Next Generation Internet*, LNCS, *(also see 4th International Workshop of the EuroNGI/EuroNF Network of Excellence)*, vol. 5122/2008, pp. 185 - 200, DOI. 10.1007/978-3-540-89183-3, Oct. 2008.

[MJ04] J. Mcnair and F. Zhu, "Vertical handoffs in fourth-generation multinetwork environments," *IEEE Wireless Communications*, vol. 11, no. 3, pp. 8 - 15, June 2004.

[MO07] O. Markaki, D. Charilas and D. Nikitopoulos, "Enhancing quality of experience in next generation networks through network selection mechanisms," *IEEE PIMRC 2007*.

[MR06] R. Moskowitz and P. Nikander, "Host identity protocol (HIP) architecture," RFC 4423, May 2006.

[MR07] R. Moskowitz, P. Nikander, P. Jokela and T. Henderson, “Host Identity Protocol,” *draft-ietf-hip-base-07.txt*, Feb. 2007.

[MW04] W. Ma and Y. Fang, “Dynamic Hierarchical Mobility Management Strategy for Mobile IP Networks,” *IEEE J. Select. Areas Commun.*, vol. 22, pp. 664 - 676, May 2004.

[NE05] E. Natalizio, A. Scicchitano and S. Marano, “Mobility Anchor Point Selection Based on User Mobility in HMIPv6 Intergrated with Fast Handover Mechanism,” *IEEE WCNC 2005*, pp. 1434 - 1439.

[NE07] E. Nordmark and M. Bagnulo, “Shim6: Level 3 Multihoming Shim Protocol for IPv6,” *draft-ietf-shim6-proto-08.txt*, May 2007.

[NH07] H. Q. Nguyen, F. Baccelli, D. Kofman, “A stochastic geometry analysis of dense IEEE 802.11 networks,” *IEEE INFOCOM 2007*, pp. 1199 - 1207.

[NV07] Q. T. Nguyen-Vuong, N. Agoulmine and Y. Ghamri-Doudane, “Terminal-controlled mobility management in heterogeneous wireless networks,” *IEEE Communications Magazine 2007*, vol. 45, no. 4, pp. 122 - 129, April 2007.

[NV08] Q. T. Nguyen-Vuong, Y Ghamri-Doudane, N Agoulmine, “On utility models for access network selection in wireless heterogeneous networks,” *IEEE NOMS 2008*, pp. 144 - 151.

[NV08-2] Q. T. Nguyen-Vuong, N. Agoulmine and Y. Ghamri-Doudane, “A user-centric and context-aware solution to interface management and access network selection in heterogeneous wireless networks,” *Computer Networks*, vol. 52, no. 18, pp. 3358 - 3372, Dec. 2008.

[OE08] E. hwee Ong and J. Y. Khan, “Dynamic access network selection with QoS parameters estimation: a step closer to ABC,” *IEEE VTC 2008 Spring*, pp. 2671 - 2676.

[OM07] O. Markaki, D. Charilas and D. Nikitopoulos, “Enhancing Quality of Experience in Next Generation Networks Through Network Selection Mechanisms,” *IEEE PIMRC 2007*, pp. 1 - 5.

[OO05] O. Ormond, P. Perry and J. Murphy, “Network selection decision in wireless heterogeneous networks,” *IEEE PIMRC 2005*, pp. 2680 - 2684.

[OO06] O. Ormond, J. Murphy and G. M. Muntean, “Utility-based Intelligent Network Selection in Beyond 3G Systems,” *IEEE ICC 2006*, pp.1831 - 1836.

[OO06-2] O. Ormond, G. M. Muntean and J. Murphy, “Economic Model for Cost Effective Network Selection Strategy in

Service Oriented Heterogeneous Wireless Network Environment," *IEEE NOMS 2006*, pp. 1 - 4.

[PD06] D. Pareek, "The business of WiMAX," *Wiley*, 2006.

[PG96] G. Pollini, "Trends in handover design," *IEEE Communications Magazine*, 34(3) 1996, 82 - 90.

[PI03] I. Popescu, "Supporting multimedia session mobility using SIP," *CNSR 2003.*

[PK00] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. Makela, R. Pichna, J. Vallstron, "handoff in hybrid mobile data networks," *IEEE Personal Communications*, 7(2) 2000, 34 - 47.

[PS03] S. Pack and Y. Choi, "Performance Analysis of Hierarchical Mobile IPv6 in IP-based Cellular Networks," *IEEE PIMRC 2003*, pp. 2818 - 2822.

[PS04] S. Pack, M. Nam and Y. Choi, "A Study on Optimal Hierarchy in Multi-Level Hierarchical Mobile IPv6 Networks," *IEEE Globecom 2004*, pp. 1290 - 1294.

[PS07] S. Pack, X. Shen, J. W. Mark and J. Pan, "Adaptive Route Optimization in Hierarchical Mobile IPv6 Networks," *IEEE Trans. Mobile Computing*, vol. 6, pp. 903 - 914, Aug. 2007.

[RE06] E. P. Ribeiro and V. C. M. Leung, "Minimum delay path selection in multi-homed systems with path asymmetry," *IEEE Communications Letters*, vol. 10, no. 3, pp. 135 - 137, March 2006.

[SC08] C. Sun, E. Stevens-Navarro and V. W. S. Wong, "A Constrained MDP-based Vertical Handoff Decision Algorithm for 4G Wireless Networks," *IEEE ICC 2008*, pp. 2169 - 2174.

[SD04] D. Saha, A. Mukherjee, I. S. Misra, M. Chakraborty and N. Subhash, "Mobility support in IP: a survey of related protocols," IEEE Network, no. 6, vol. 18, pp. 34 – 40, Nov. – Dec. 2004.

[SH05] H. Soliman, C. Castelluccia, K. El Malki and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," *IETF RFC 4140*, Aug. 2005.

[SH05] H. Soliman, C. Castelluccia, K. El Malki and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," *RFC 4140*, Aug. 2005.

[SH07] S. Horrich, S. B. Jamaa and P. Godlewski, "Adaptive Vertical Mobility Decision in Heterogeneous Networks," *IWCMC 2007*, pp. 1 - 6.

[SN06] E. Stevens-Navarro and V. W. S. Wong, “Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks,” *IEEE VTC 2006 Spring*.

[SN08] E. Stevens-Navarro, Y. Lin and V. S. W. Wong, “An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1243 - 1254, March 2008.

[SQ05] Q. Song and A. Jamalipour, “Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques,” *IEEE Wireless Communications*, vol. 12, no. 3, pp. 42 - 48, June 2005.

[SQ05-2] Q. Song and A. Jamalipour, “A network selection mechanism for next generation networks Communications,” *IEEE ICC 2005*, pp. 1418 - 1422.

[SQ06] Q. Song and A. Jamalipour, “A Negotiation-Based Network Selection Scheme for Next-Generation Mobile Systems,” *IEEE Globecom 2006*, pp. 1 - 5.

[SQ08] Q. Song and A. Jamalipour, “A quality of service negotiation-based vertical handoff decision scheme in heterogeneous wireless systems,” *European Journal of Operational Research*, 191(3): 1059-1074 (2008).

[SS07] S. Sengupta, M. Chatterjee and K. Kwiat, “Pricing-based service and network selection in overlaid access networks,” *International Conference on Information, Communications & Signal Processing 2007*, pp. 1 - 5.

[SW07] W. Shen and Q. A. Zeng, “Cost-Function-Based Network Selection Strategy in Integrated Wireless and Mobile Networks,” *AINAW workshop 2007*, pp. 314 - 319.

[SW08] W. Shen and Q. A. Zeng, “Cost-Function-Based Network Selection Strategy in Integrated Wireless and Mobile Networks,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3778 - 3788, Nov. 2008.

[TP06] P. Thubert R. Wakikawa and V. Devarapalli, “Global HA to HA Protocol,” *draft-thubert-nemo-global-haha-02.txt*, Sept. 2006.

[TT05] T. Taleb, T. Suzuki and N. Kato, “A Dynamic and Efficient MAP Selection Scheme for Mobile IPv6 Networks,” *IEEE Globecom 2005*, pp. 2891 - 1895.

[v4ad] IPv4 address report, http://www.potaroo.net/tools/ipv4/.

[VB07] E. Van den Berg, P. Gopalakrishnan, B. Kim, B. Lyles, W. I. Kim, Y. S. Shin and Y. J. Kim, “Dynamic Network Selection using Kernels,” *IEEE ICC 2007*, pp. 6049 - 6054.

[VE03] E. Vanem, S. Svaet and F. Paint, “Effects of multiple access alternatives in heterogeneous wireless networks,” *IEEE WCNC 2003*, pp. 1696 - 1700.

[WH99] H. J. Wang, R. H. Katz, and J. Giese, “Policy-enabled handoffs across heterogeneous wireless networks,” *2nd IEEE workshop on Mobile Computing Systems and Application 1999*, pp. 51 - 60.

[WL07] L. C. Wang, A. Chen, and H. H. Chen, “Network Selection with Joint Vertical and Horizontal Handoff in Heterogeneous WLAN and Mobile WiMax Systems,” *IEEE VTC 2007 Spring*, pp. 794 - 798.

[WL08-1] L. Wang and D. Binet, “A Markov chain mobility model for heterogeneous wireless networks,” *ICLAN 2008*, pp. 1 - 6.

[WL08-2] L. Wang, D. Kofman and D. Binet, “Towards a framework of network selection in heterogeneous wireless networks: a survey,” *5th internal EuroNF workshop of wireless and mobility*, Oct. 2008, pp. 1 - 6.

[WL08-3] L. Wang, B. Gaabab, D. Binet and D. Kofman, “Novel MAP selection scheme using location history in hierarchical MIPv6 networks,” *IEEE WCNC 2008*, pp. 1 - 6.

[WL09-1] L. Wang and D. Binet, “Best permutation: a novel network selection scheme in heterogeneous wireless networks,” *ACM IWCMC 2009*, pp. 1 - 6.

[WL09-2] L. Wang and D. Binet, “A simulative study of MADM-based network selection in heterogeneous wireless networks,” *CTIF Wireless Vitae 2009*, pp. 1 - 6.

[WL09-3] L. Wang and D. Binet, “Mobility-based network selection in heterogeneous wireless networks,” *IEEE VTC 2009 Spring*, pp. 1 - 5.

[WL09-4] L. Wang and D. Binet, “TRUST: a trigger-based automatic subjective weighting method for network selection”, *IARIA AICT 2009*, pp. 1 - 7.

[WL10-1] L. Wang, “Network selection in heterogeneous wireless networks: a survey,” submitting to IEEE Communications Surveys & Tutorials, pp. 1 - 15, work in progress.

[WL10-2] L. Wang, “Best permutation scheme for network selection in heterogeneous wireless networks,” submitting to *IEEE Transactions on Vehicular Technology*, pp. 1 - 15, work in progress.

[WQ06] Q. Wei, K. Farkas, C. Prehofer, P. Mendes and B. Plattner, “Context-aware handover using active network technology,” *Computer Networks*, 50 (15), 2006, 2855 - 2872.

[WR04] R. Wakikawa, V. Devarapalli, and P. Thubert, “Inter Home Agents Protocol (HAHA),” *draft-wakikawa-mip6-nemo-haha-01*, February 2004.

[WR07] R. Wakikawa, T. Ernst, K. Nagami and V. Devarapalli, “Multiple Care-of Addresses Registration,” *draft-ietf-monami6-multiplecoa-02.txt*, March 2007.

[WT08] T. Wu, H. Jing, X. Yu, X. Qu and Y. Wang, “Cost-Aware Handover Decision Algorithm for Cooperative Cellular Relaying Networks,” *IEEE VTC 2008 Spring*, pp. 2446 - 2450.

[WY08] Y. Wang, J. Yuan, Y. Zhou, G. Li and P. Zhang, “Vertical handover decision in an enhanced media independent handover framework,” *IEEE WCNC 2008*, 2693 - 2698.

[WY08-2] Y. Wang, J. Yuan, Y. Zhou, G. Li and P. Zhang, “An Enhanced Media Independent Handover Framework for Heterogeneous Networks,” *IEEE VTC 2008 Spring*, 2306 - 2310.

[XH93] H. Xie and D. J. Goodman, “Mobility models and biased sampling problem,” *2nd International Conf. Universal Personal Coomunications*, pp. 803 - 807.

[XY03] Y. Xu, C. J. Lee and L. L. Thing, “A local Mobility Agent Selection Algorithm for Mobile Networks,” *IEEE ICC 2003*, pp. 1074 - 1079.

[YJ07] Y. Ji, K. Zhang, Z. Feng, C. Chi, P. Zhang, and M Muck, “CPC-assisted Network Selection Strategy,” *16th IST Mobile and Wireless Communications Summit*, July 2007, pp. 1 – 5.

[ZA07] A. H. Zahran and B. Liang, “A generic framework for mobility modeling and performance analysis in next generation heterogeneous wireless networks,” *IEEE Commun. Mag.*, vol. 45, no. 9, pp. 92 - 99, Sept. 2007.

[ZA08] A. H. Zahran and B. Liang, “Mobility Modeling and Performance Evaluation of Heterogeneous Wireless Networks,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 8, pp. 1041 - 1056, Aug. 2008.

[ZW04] W. Zhang, “Handover decision using fuzzy MADM in heterogeneous wireless networks,” *IEEE WCNC 2004*, pp. 653 - 658.