

Prédiction de mobilité par le mobile ou par le point d'accès : comparaison sur base de traces réelles

Jean-Marc François, Guy Leduc

► **To cite this version:**

Jean-Marc François, Guy Leduc. Prédiction de mobilité par le mobile ou par le point d'accès : comparaison sur base de traces réelles. Colloque Francophone sur l'Ingénierie des Protocoles - CFIP 2006, Eric Fleury and Farouk Kamoun, Oct 2006, Tozeur/Tunisia, 12 p. inria-00110283

HAL Id: inria-00110283

<https://hal.inria.fr/inria-00110283>

Submitted on 20 Nov 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prédiction de mobilité par le mobile ou par le point d'accès : comparaison sur base de traces réelles

Jean-Marc François — Guy Leduc

Université de Liège — Institut Montefiore (B28)

B-4000 Liège Sart-Tilman, Belgique

Tél. : +32-4-3662871

{francois,leduc}@run.montefiore.ulg.ac.be

RÉSUMÉ. Le problème de la prédiction de mobilité se définit comme le fait de deviner quel sera le prochain point d'accès rencontré par un terminal mobile lors de son déplacement dans un réseau sans fil. Les prédictions faites permettent d'améliorer la qualité de service fournie par le réseau en lui permettant de prendre des mesures pro-actives (telles des réservations de ressources). Les agents de prédiction se classent principalement en deux catégories : les agents liés à un mobile particulier (responsables d'anticiper les déplacements de celui-ci) et ceux liés à un point d'accès (prédisant le prochain point d'attache de tous les terminaux y étant connectés). Cet article vise à comparer les deux méthodes à l'aide de traces réelles tirées d'un réseau WiFi de grande taille. Il montre que certains postulats souvent admis (comme le fait que les habitudes de mouvement du week-end sont différentes de celles du reste de la semaine) doivent être revus.

ABSTRACT. The mobility prediction problem is defined as guessing a mobile node's next access point as it moves through a wireless network. Those predictions help take proactive measures in order to guarantee a given quality of service. Prediction agents can be divided into two main categories: agents related to a specific terminal (responsible for anticipating its movements) and those related to an access point (which predict the next access point of all the mobiles connected through it). This paper aims at comparing those two schemes using real traces of a large WiFi network. It shows that commonly admitted hypotheses (such as the different motion behaviour between the week-end and the rest of the week) should be revised.

MOTS-CLÉS : Prédiction, mobilité, modèles markoviens, traces WiFi

KEYWORDS: Prediction, mobility, markovian models, WiFi traces

This work has been partially supported by the Belgian Science Policy in the framework of the IAP program (Motion P5/11 project) and by the European E-Next NoE and IST-FET ANA project.

1. Introduction

Les réseaux mobiles sont actuellement l'objet de deux évolutions concomitantes :

- leur nombre d'utilisateurs augmente rapidement, ainsi que la demande pour des réseaux toujours plus rapides. La tendance actuelle est donc à réduire la zone couverte par chaque émetteur, ce qui provoque de fréquent changement (ou *handovers*) de point d'accès (AP) lors des déplacements d'hôtes mobiles (MH) ;
- la convergence des réseaux de données, de voix et de télédistribution amène une demande pour des réseaux pouvant assurer une garantie de qualité de service (QoS) minimale.

Lorsqu'un terminal mobile se déplace, une des causes majeures de dégradation de son service est le changement de point d'accès au réseau. Celui-ci implique de prendre des mesures permettant d'assurer le re-routage du flux de données émis et reçu par le mobile, procédure dépendant des protocoles utilisés mais qui est, dans tous les cas, cause de délais et de pertes. Prédire le(s) prochain(s) *handover(s)* d'un MH permet donc de prendre des mesures pro-actives réduisant leur impact, un problème désigné sous le terme *prédiction de mobilité*. Dans ce qui suit, on se placera dans le cadre d'un réseau sans fil WiFi et on s'attachera à prédire le prochain AP rencontré par un hôte mobile.

Les méthodes de prédiction de mobilité se classent en deux familles principales :

- **prédictions intrinsèques** (ou *relatives à un MH*, RMH) : l'agent de prédiction est lié à un MH ; il construit un modèle de déplacements propre à ce mobile ;
- **prédictions extrinsèques** (ou *relatives à un AP*, RAP) : l'agent de prédiction est lié à un AP ; ce point d'accès construit un modèle basé sur les déplacements des mobiles qui le traversent.

Ces modèles sont donc construits au fur et à mesure du déplacement des MH et peuvent à tout moment remplir leur rôle de prédiction.

Les sources d'information permettant d'inférer le futur parcours probable d'un terminal mobile sont variées (par exemple, des coordonnées GPS). Dans ce qui suit, on fait l'hypothèse (assez faible) que cette source d'information est une trace des AP que les MH ont récemment traversés.

Dans cet article, on s'attache à la comparaison (essentiellement quantitative) des méthodes de prédiction intrinsèques et extrinsèques *via* des traces de déplacements réelles extraites d'un réseau WiFi de grande taille. On peut espérer que les conclusions tirées ici soient applicables à d'autres types de réseaux mobiles.

Cet article est organisé comme suit. La section 2 définit les modèles de déplacements utilisés dans cet article. Viennent ensuite les descriptions des données réelles analysées et d'un problème spécifique qu'elles ont posé (section 4). Les sections 5 et 6 comparent les performances obtenues par les deux types d'agents de prédiction (intrinsèque et extrinsèque).

2. Agents de prédiction

Un agent de prédiction construit un modèle des mouvements de mobiles pour utiliser celui-ci à des fins prédictives. Les paramètres du modèle sont ajustés (appris) au fur et à mesure que les MH se déplacent et permettent donc de réaliser des prédictions de plus en plus précises (si le processus de déplacement des mobiles est stationnaire).

2.1. Modèles markoviens

Dans ce qui suit, on s'attache à prédire le prochain point d'accès d'un mobile dans un réseau WiFi. Pour ce faire, on modélise les habitudes de déplacement des MH en étudiant leur *historique* (ou *trace*) de déplacements, soit la séquence de points d'accès traversés tout au long de son parcours. En voyant chaque AP comme un symbole d'un alphabet (fini), l'historique (ou *trace*) de déplacement d'un mobile est une séquence de symboles et une prédiction consiste à deviner le symbole $i + 1$ d'une séquence étant donné les i premiers.

Il a été montré ([SON 04a, SON 04b]) que, dans ce contexte, de simples modèles markoviens donnent des résultats aussi bons que d'autres modèles plus complexes¹ (tels que [BHA 99, YU 02, CLE 84, JAC 00]). Nous nous attachons donc à l'étude de tels modèles dans ce qui suit.

En dénotant l'ensemble des symboles par \mathcal{L} et un historique de déplacement par $L = L_1, L_2, \dots$, l'hypothèse d'un processus stochastique markovien (ou sans mémoire) d'ordre n est :

$$P(L_i = l | L_1, \dots, L_{i-1}) = P(L_i = l | L_{i-n}, \dots, L_{i-1}) \quad \forall l \in \mathcal{L}, i > n \quad [1]$$

Moins formellement, on considère donc qu'une variable aléatoire décrit l'arrivée du symbole suivant ; sa distribution ne dépend que des n symboles précédents. On fait l'hypothèse de stationnarité.

La distribution du prochain point d'accès peut aisément s'apprendre en ligne. On suppose que l'agent de prédiction est régulièrement notifié du déplacement de MH. Soit L^m l'historique des déplacements du mobile m , la règle d'estimation d'un modèle markovien d'ordre n est :

$$P(L_i = l | L_{i-n}, \dots, L_{i-1}) = \frac{\sum_{m \in \mathcal{M}} O(L_{i-n}^m, \dots, L_{i-1}^m, l; L^m)}{\sum_{m \in \mathcal{M}} O(L_{i-n}^m, \dots, L_{i-1}^m; L^m)} \quad [2]$$

où l'opérateur $O(A; B)$ détermine le nombre d'occurrences de "A" dans "B" et \mathcal{M} est l'ensemble des mobiles étudiés. Dans le cas le plus simple, un mobile modélise ses propres déplacements et \mathcal{M} est un singleton. Cette équation montre que la distribution

1. Précisons que certaines de ces méthodes plus complexes permettent non seulement de réaliser des prédictions, mais également du *paging*.

de prochain AP peut s'estimer pour tout *contexte* d'historique de déplacements (c'est-à-dire toute portion de longueur n d'un historique).

Lorsqu'un modèle est utilisé pour faire une prédiction, le prochain AP le plus probable (étant donné le contexte) est choisi. Si le contexte n'a jamais été vu auparavant, aucune prédiction ne peut être réalisée. Pour rendre cette éventualité peu probable, en plus d'un modèle d'ordre n , on construit $n-1$ modèles d'ordre $n-1, n-2, \dots, 1$. Lors d'une prédiction, si le contexte de longueur n est vu pour la première fois, on procède à une prédiction à l'aide d'un modèle d'ordre inférieur. On ne s'attachera pas ici à prédire l'apparition d'un mobile sur le réseau, ni le moment auquel il le quitte ; ne sont pris en considération que les mouvements proprement dits des terminaux mobiles.

2.2. Méthodes intrinsèques et extrinsèques

Dans cet article, on vise à mettre en exergue les différences entre les deux principales méthodes de prédiction.

Avec la méthode extrinsèque, RAP, les agents de prédiction sont les points d'accès. Le modèle de déplacements, dont l'apprentissage est réalisé par un AP particulier, est basé sur les traces de déplacements de tous les MH qui l'ont traversé. Les mobiles tiennent ici un rôle mineur : ils envoient à leur AP courant un identificateur des derniers AP rencontrés. En faisant l'hypothèse que la prédiction est principalement utile à l'infrastructure du réseau (qui peut par exemple l'utiliser pour mettre en place anticipativement une réservation de ressources), on constate que cette solution semble fort naturelle.

La méthode intrinsèque, RMH, est plus distribuée : chaque MH fait l'apprentissage de ses propres déplacements. Puisque plus spécifique dans son approche, on peut attendre de cette technique de meilleurs résultats : le mouvement d'un mobile particulier n'est sans doute pas réductible au mouvement des MH dans leur ensemble. Par contre, elle s'insère assez mal dans le paradigme habituel des réseaux sans fil, où chaque terminal est supposé disposer d'une faible quantité de mémoire et d'une puissance de calcul médiocre (par exemple, un GSM d'entrée de gamme), donc mal adapté à l'exécution d'un algorithme d'apprentissage.

3. Historiques de déplacements

Les historiques de déplacements utilisés ont été collectés par l'Université de Dartmouth dans le cadre du projet CRAWDAD ([CRA]) ; il s'agit d'un recensement des événements générés par le réseau sans fil WiFi du campus. Des données *Syslog* et *SNMP* ont été enregistrées pendant 2 années et couvrent 6202 utilisateurs et 575 AP ([KOT 02]).

Ces mêmes données ont été analysées ([SON 04b]) pour en extraire les traces de déplacements proprement dites (soit, une séquence de points d'accès pour chaque

MH). Un point d'accès spécial, dénoté *OFF*, signale qu'un mobile s'est 'désauthenticé' ou n'a plus présenté d'activité depuis 30 minutes ; on considère alors qu'il s'est déconnecté du réseau.

Ces données ont été divisées en deux parties : cinq sixièmes servent d'ensemble d'apprentissage, le reste d'ensemble de test (uniquement utilisé à la section 6).

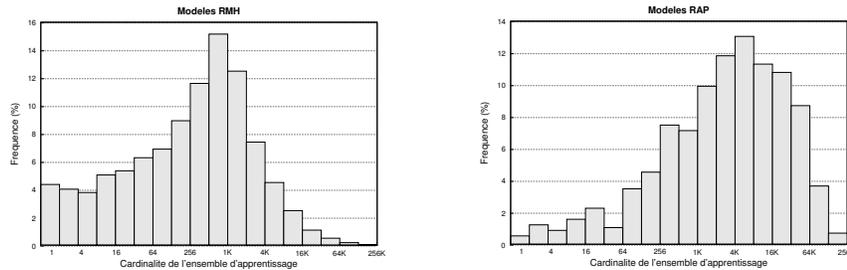


Figure 1. Distribution des cardinalités des ensembles d'apprentissage pour RMH (gauche) et RAP (droite). L'axe x est logarithmique.

Dans ce qui suit, cet ensemble d'apprentissage est utilisé aussi bien pour l'apprentissage RMH que RAP. Plus formellement, à chaque mouvement d'un MH m , ses n derniers AP $L_{i-n}^m, \dots, L_{i-1}^m$ et le point d'accès suivant L_i^m sont utilisés pour faire l'apprentissage des paramètres d'un modèle markovien d'ordre n . Ces couples "contexte, AP suivant" sont les éléments des ensembles d'apprentissage. RMH utilise m comme agent de prédiction alors qu'il s'agira de L_{i-1}^m dans le cas de RAP. Notons que les contextes fournis à L_{i-1}^m ont toujours " L_{i-1}^m " comme dernier élément ; la complexité d'un modèle d'ordre n pour RAP est donc similaire à celle d'un modèle d'ordre $n + 1$ pour RMH.

La figure 1 compare les distributions des cardinalités des ensembles d'apprentissage dans le cas de RMH et RAP. Beaucoup de MH bougent très peu : 12% effectuent moins de 8 mouvements (figure de gauche, somme des pourcentages correspondant aux trois premières barres). Par contre, on constate que peu de points d'accès sont peu populaires : 15% sont traversés par moins de 256 MH. Le phénomène de sous-apprentissage devrait donc être plus marqué dans le cas d'une méthode intrinsèque.

4. Prédiction du prochain point d'accès

4.1. Exactitude des prédictions

La table 1 donne les pourcentages de prédictions correctes pour chacune des deux méthodes présentées et pour des modèles de différents ordres.

| Méthode\Ordre | 0 | 1 | 2 | 3 | 4 | 5 |
|---------------|------------------|-------|-------|-------|-------|-------|
| RAP | N/A ² | 40.9% | 64.5% | 64.9% | 65.1% | 64.9% |
| RMH | 38.7% | 68.4% | 72.9% | 73.0% | 73.0% | 72.8% |

Tableau 1. Exactitude des prédictions. *Chaque entrée correspond au rapport du nombre de prédictions correctes et du nombre total de prédictions.*

La première constatation est qu'une simple approche statistique³ donne de mauvais résultats. Ils s'améliorent toutefois très vite : ils sont quasi-optimaux avec des modèles d'ordre 2. Ils n'augmentent presque plus pour arriver à leur apogée lorsque l'ordre est égal à 4. Les résultats relatifs à RMH sont en accord avec ceux de [KOT 02]. Des modèles d'ordres supérieurs présentent des résultats moins bons, signe d'un léger sur-apprentissage.

La méthode intrinsèque présente de meilleures performances. On pourrait penser qu'il s'agit d'un indice —attendu— des différences qui marquent nos comportements personnels. Cette affirmation est certainement vraie, mais nous montrons ci-dessous qu'un artefact technique est également en grande partie la cause de ces différences.

4.2. L'impact du ping-ponging

Le *ping-ponging* se définit comme un passage répété entre deux point d'accès (au moins) de façon oscillante. Un périphérique se connecte normalement au point d'accès caractérisé par le signal le plus fort, ce qui dépend à la fois du déplacement du mobile et des conditions de propagations. Un mécanisme d'hystérésis —qui varie de terminal en terminal— limite normalement ce type de comportement.

La prédiction de mobilité s'intéresse au mouvement physique des terminaux mobiles et non à ce type d'artefact rapide. Si la prédiction a pour but une réservation de bande passante B par exemple, une stratégie simple serait d'allouer un débit $B/2$ aux deux AP entre lesquels le phénomène de ping-pong apparaît ; ils ne sont plus alors considéré que comme une seule entité. Cela n'implique pas que prédire le ping-ponging soit une question sans intérêt, mais plutôt qu'elle est assez éloignée du problème de la prédiction de mobilité et n'est pas abordée ici.

Étant donné une trace de déplacements L_1, \dots, L_n , le mouvement vers L_n est classifié comme ping-ponging si $L_{n-2} = L_n$. Cette règle conduit certainement à quelques faux positifs, par exemple quand un MH fait physiquement l'aller-retour d'un point d'accès à un autre. Toutefois, de tels mouvements sont assez rares, comme le confirme les résultats ci-dessous.

2. En effet, l'ensemble d'apprentissage d'un modèle d'ordre n est formé de contextes de longueur n . Dans RAP, l'agent de prédiction est le dernier AP de ce contexte, qui n'est pas défini dans le cas de contexte de longueur 0.

3. C'est-à-dire d'ordre 0 ou 1 pour respectivement RMH et RAP.

La figure 2 et la table 2 montrent que le ping-ponging est un phénomène fréquent et le corrèle aux fabricants des interfaces sans fil. La table 2 montre le rapport des tailles d'ensembles d'apprentissage sans et avec les éléments considérés comme ping-ponging. On peut en tirer différents enseignements.

Premièrement, notre critère de classification semble pertinent puisque les résultats varient de fabricant en fabricant : il filtre des handovers déclenchés par une cause technologique et non par le véritable mouvement des MH.

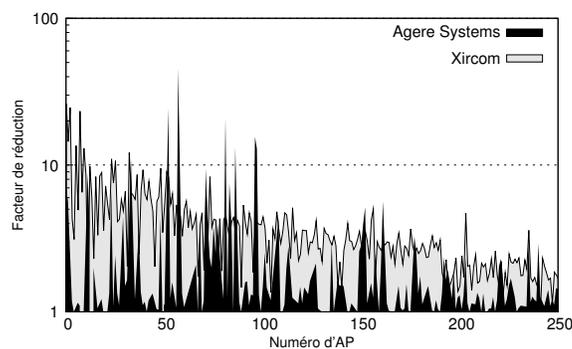


Figure 2. Ratios des cardinalités des ensembles d'apprentissage avant et après la suppression du ping-ponging pour deux fabricants d'interfaces sans fil. Les AP sont triés par ratio moyen.

Deuxièmement, la réduction de cardinalité des ensembles d'apprentissage due au ping-ponging est très élevée : environ 3 mouvements sur 10 sont considérés comme tel. Cela explique sans doute les résultats de [SON 04b] montrant qu'il est inutile de construire des modèles complexes (par exemple basés sur Lempel-Ziv) : des modèles simples (markoviens de faibles ordres) suffisent en effet à prédire le ping-ponging. Il serait intéressant de supprimer ce biais et répéter ces expériences.

Finalement, les écarts-types montrent la diversité des situations rencontrées par les MH : certains connaissent un grand nombre de ping-ponging, d'autres très peu.

| Fabricant | Sans pondération | Pondéré |
|------------------|------------------|---------------|
| Agere Systems | 44.9% (23.8%) | 63.1% (27.3%) |
| Xircom | 52.6% (27.4%) | 65.4% (25.0%) |
| Aironet Wireless | 58.9% (26.1%) | 71.3% (26.0%) |
| Apple Computer | 72.3% (20.9%) | 79.0% (20.1%) |

Tableau 2. Facteur de réduction des cardinalité d'ensemble d'apprentissage par fabricant. Les écarts-types sont donnés entre parenthèses. Les pourcentages de la colonne de gauche considèrent que tous les mobiles ont la même importance ; la colonne de droite pondère ces chiffres par la longueur des historiques de chaque MH. Seuls les fabricants représentant 200 MH au moins sont listés.

Les résultats qui suivent sont tous obtenus après que les mouvements classifiés comme ping-pong aient été enlevés.

La table 3 donne la fiabilité des méthodes RAP et RMH. On y voit que les performances de RMH sont devenues (pratiquement) indépendantes du fabricant et que, pour des modèles d'ordre assez grand, les différences entre RAP et RMH sont assez faibles (passant de 55.2% à 59.8%). Les performances sont aussi moins bonnes que celles présentées table 1, ce qui montre que les mouvements répétitifs de ping-pong sont facilement prévisibles.

| Ordre | RAP | RMH | | | | |
|-------|-------|---------|-------|--------|---------|-------|
| | | Moyenne | Agere | Xircom | Aironet | Apple |
| 1 | 28.7% | 39.4% | 38.8% | 40.0% | 39.9% | 38.6% |
| 2 | 47.6% | 58.4% | 56.6% | 59.3% | 59.3% | 58.9% |
| 3 | 53.0% | 59.5% | 57.4% | 60.4% | 60.3% | 60.1% |
| 4 | 55.2% | 59.8% | 57.5% | 60.7% | 60.7% | 60.4% |
| 5 | 55.4% | 59.6% | 57.5% | 60.6% | 60.5% | 60.2% |

Tableau 3. Exactitude des prédictions *sans tenir compte des mouvements de ping-pong*.

5. Performances

5.1. Performance des agents de prédiction vs taille d'ensemble d'apprentissage

On l'a vu, les phases d'apprentissage de paramètres des modèles markoviens et de prédiction sont ici *entrelacées*. Les taux de prédiction correcte sont donc fonction du temps d'apprentissage, lui-même fonction de la méthode utilisée —RAP ou RMH.

Les courbes figure 3 (traits gras) montrent les taux instantanés de prédictions correctes en fonction de la cardinalité de l'ensemble d'apprentissage. Elles nous apprennent qu'un ensemble d'apprentissage composé de quelques centaines d'éléments permet déjà d'atteindre des prédictions exactes à plus de 40%, que les agents de prédictions soient liés aux mobiles ou aux points d'accès. Les deux méthodes présentent ensuite des profils différents. Dans le cas de RAP, les performances s'améliorent doucement pour atteindre 55% lorsque 3000 éléments d'apprentissage ont été observés. RMH voit ses performances augmenter beaucoup plus vite pour atteindre 60% après 1000 éléments et 73% après 2000. Ce dernier chiffre semble peu réaliste et est donc étudié plus attentivement ci-dessous.

Les courbes en traits fins montrent le pourcentage des agents de prédictions disposant d'un ensemble d'apprentissage d'une cardinalité supérieure à une valeur donnée. (Par exemple, pour RMH, environ 20% des mobiles ont un ensemble d'apprentissage comportant plus de 500 éléments.) La courbe RMH décroît de façon beaucoup plus abrupte que celle de RAP : pratiquement tous les modèles ont un ensemble d'appren-

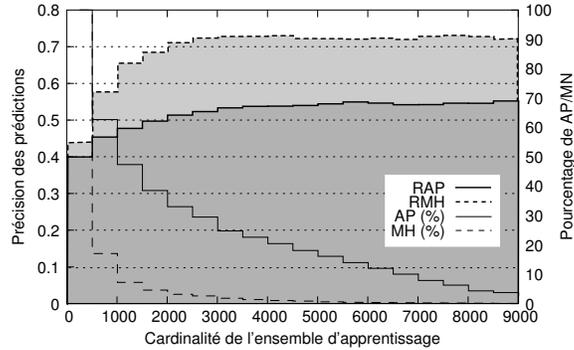


Figure 3. Taux de prédictions correctes (traits gras, axe de gauche) et fraction des agents de prédictions (traits fins, axe de droite) en fonction de la cardinalité des ensembles d'apprentissage *pour des modèles d'ordre 3*.

tissage d'une cardinalité inférieure à 3000. La courbe relative à RAP a une allure très différente : des tailles de plus de 8000 éléments sont encore fréquentes.

Un très petit nombre de MH sont donc responsables de prédictions d'une efficacité supérieure à 70%. Une recherche manuelle des historiques de déplacement laisse penser que le ping-pong entre 3 AP (ou plus) est la cause de cet anomalie. Il semble difficile de retirer automatiquement des mouvements de ping-pong entre plus de 2 AP : il est délicat de les différencier de mouvements normaux. Heureusement, leur petit nombre n'influe que marginalement sur les résultats.

5.2. Comparaison des entropies résiduelles

La connaissance des derniers AP traversés (le *contexte*) par un terminal mobile ne permet bien entendu pas, en général, de déduire sans risque d'erreur son point d'accès suivant. On peut par contre décrire ce point d'accès suivant par une distribution de probabilité qui correspond elle-même à une entropie donnée.

Plus formellement, si \mathcal{C} dénote l'ensemble des contextes, $N_a(c)$ le nombre d'occurrences du contexte c vu par l'agent de prédiction a et \mathcal{L} l'ensemble des AP, l'entropie associée à un agent de prédiction est donnée par :

$$H_a = - \frac{1}{\sum_{c \in \mathcal{C}} N_a(c)} \sum_{c \in \mathcal{C}} N_a(c) \sum_{l \in \mathcal{L}} P_a(l|c) \log_2 P_a(l|c) \quad [3]$$

$P_a(l|c)$ est la probabilité estimée par l'agent a que l soit le prochain AP connaissant le contexte c ; elle peut être calculée via [2]. La table 4 reprend les mesures d'entropie.

On constate une différence importante entre les deux méthodes, à l'encontre des résultats présentés dans la table 3 qui montrent certes une différences entre RAP et

RMH, mais beaucoup moins marquée. On peut donc en conclure que RMH prédit très sensiblement mieux quels peuvent être les prochains points d'accès d'un mobile, mais que ce problème est très différent de celui étudié ici, qui ne s'attache qu'au choix *du prochain AP le plus probable*.

5.3. Division temporelle

Il est fréquent (par exemple [CHO 98]) de supposer désirable le tronçonnage d'un ensemble d'apprentissage en périodes de temps homogènes ; il semble par exemple raisonnable de faire l'hypothèse d'habitudes de mouvements différentes entre le week-end et le reste de la semaine, et donc de construire des modèles différents pour ces deux périodes. En utilisant cette méthode, on s'expose par contre aux problèmes potentiels de performances dégradées à cause du début de la courbe d'apprentissage ou d'ensembles d'apprentissage de trop petite taille.

La table 5 donne les résultats obtenus en expérimentant différents types de tronçonnages. Ils sont assez surprenants : aucune amélioration n'est apportée dans aucun cas. Deux hypothèses peuvent l'expliquer : soit les comportements des MH sont sensiblement les mêmes pour chacune de ces périodes de temps, soit le contexte de déplacement capture ces changements de comportement.

6. Prédiction de mobilité de MH externes

Un terminal mobile rencontre régulièrement des zones d'un réseau qu'il n'a jamais rencontrées ; un touriste utilisant son GSM est généralement dans cette situation. La méthode RMH n'est ici d'aucun secours.

Pour simuler cette situation, nous divisons les MH en deux ensembles : un ensemble d'apprentissage (EA, 5005 éléments) et un ensemble de test (ET, 1197 éléments). L'ensemble d'apprentissage permet la construction des modèles, comme précédemment. Ceux-ci permettent de prédire le prochain point d'accès des éléments de l'ensemble de test.

La table 6 donne les résultats obtenus précédemment (table 3) avec le modèle RAP (première ligne), comparés avec les performances réalisées lorsque l'ensemble de test est le même que l'ensemble d'apprentissage (deuxième ligne) ou avec un ensemble de test indépendant (troisième ligne).

| Méthode\Ordre | 1 | 2 | 3 |
|---------------|-------------|-------------|-------------|
| RAP | 1.86 (0.95) | 1.72 (1.18) | 1.58 (0.49) |
| RMH | 0.98 (0.50) | 0.82 (0.66) | 0.91 (0.37) |

Tableau 4. Entropie des distributions de prochaines cellules. *Les écarts-types sont donnés entre parenthèses.*

Les très bons résultats obtenus lorsque l'ensemble d'apprentissage est utilisé comme ensemble de test montrent que des modèles aussi simples que des modèles markoviens sont susceptibles de faire l'apprentissage de caractéristiques individuelles ; un corollaire est que l'utilisation de ces modèles pour prédire le trajet de mobiles qui n'avaient jamais rejoint le réseau auparavant donne des résultats significativement moins bons.

7. Conclusions

Il existe deux grandes classes d'agents de prédictions, ici dénomées RAP (ou extrinsèques) et RMH (ou intrinsèques). Il est assez surprenant de constater qu'aucune étude ne s'attachait jusqu'ici à en faire une comparaison quantitative.

Cet article comble partiellement ce manque grâce à une étude basée sur des modèles markoviens. L'apprentissage des paramètres de ces modèles est fait *via* l'analyse d'une base de données reprenant les traces réelles d'un réseau WiFi.

Il montre que :

- extraire des traces du réseau le mouvement réel des utilisateurs n'est pas aisé : le phénomène de *ping-ponging* génère un grand nombre de mouvements parasites. Ne pas en tenir compte revient à construire des modèles qui sont pour une grande part le reflet de cet artefact et non du mouvement physique des utilisateurs.

- alors qu'on pourrait penser beaucoup plus efficace de construire un modèle pour chaque mobile, on constate que les efficacités des deux approches RAP et RMH ne diffèrent que de quelques pour cent (typiquement 55% pour l'une, 59% pour l'autre) ;

- le taux de prédictions correctes est dans tous les cas bas (moins de 60%) ; ceci doit être mis en perspective vu l'entropie intrinsèque du déplacement des mobiles WiFi (on peut supposer cette entropie plus basse pour d'autres types de réseau, par exemple GSM) ;

- étonnamment, construire des modèles différents pour différentes périodes de temps (semaine/week-end, matin/après-midi) n'apporte pas d'amélioration ;

- dans le cas de RAP, les modèles construits peuvent servir à prédire le mouvement de MH qui n'ont pas été utilisés lors de la phase d'apprentissage ; les performances

| Granularité | RAP | RMH |
|-------------------------|-------|-------|
| Aucun tronçonnage | 55.2% | 59.8% |
| Week-end/Semaine | 54.2% | 58.4% |
| Matin/Après-midi | 54.1% | 58.2% |
| Uniquement janvier 2003 | 53.8% | 57.0% |
| Périodes de deux heures | 53.1% | 53.4% |
| Jours de la semaine | 52.0% | 54.6% |

Tableau 5. Efficacité des prédictions pour des modèles d'ordre 4 en utilisant différents tronçonnages.

| Ordre | 1 | 2 | 3 | 4 | 5 |
|-------------|-------|-------|-------|-------|-------|
| EA | 28.7% | 47.6% | 53.0% | 55.2% | 55.4% |
| EA comme ET | 28.8% | 49.4% | 58.9% | 66.8% | 71.5% |
| ET | 23.4% | 39.0% | 45.9% | 48.5% | 49.0% |

Tableau 6. Performances obtenues par différents ensembles de tests.

obtenues sont alors très sensiblement moins bonnes.

Cette comparaison peut être poursuivie de différentes manières. En particulier, l'utilisation de modèles plus complexes (utilisant, par exemple, des mesures d'angles d'arrivée) et l'analyse de types de réseaux différents (par exemple, GSM) sont deux voies de recherche intéressantes.

8. Bibliographie

- [BHA 99] BHATTACHARYA A., DAS S., « LeZi-Update : an information-theoretic approach to track mobile users in PCS networks », *Proc. of ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99)*, Seattle, August 1999.
- [CHO 98] CHOI S., SHIN K. G., « Predictive and adaptive bandwidth reservation for hand-offs in QoS-sensitive cellular networks », *ACM SIGCOMM Computer Communication Review*, vol. 28, n° 4, 1998, p. 155–166, ACM Press.
- [CLE 84] CLEARY J. G., WITTEN I. H., « Data compression using adaptive coding and partial string matching », *IEEE Transactions on Communications*, vol. 32, n° 4, 1984, p. 396–402.
- [CRA] « CRAWDAD, a Community Resource for Archiving Wireless Data At Dartmouth », <http://crawdad.cs.dartmouth.edu/>.
- [JAC 00] JACQUET P., SZPANKOWSKI W., APOSTOL I., « An universal predictor based on pattern matching, preliminary results », chapitre 7, p. 75–85, Daniele Gardy and Abdelkader Mokkadem editors, Birkhauser, 2000.
- [KOT 02] KOTZ D., ESSIEN K., « Analysis of a Campus-wide Wireless Network », *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (MobiCom)*, September 2002, p. 107–118, Revised and corrected as Dartmouth CS Technical Report TR2002-432.
- [SON 04a] SONG L., KOTZ D., JAIN R., HE X., « Evaluating location predictors with extensive Wi-Fi mobility data », rapport n° TR2004-491, February 2004, Dept. of Computer Science, Dartmouth College.
- [SON 04b] SONG L., KOTZ D., JAIN R., HE X., « Evaluating location predictors with extensive Wi-Fi mobility data », *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 2, March 2004, p. 1414–1424.
- [YU 02] YU F., LEUNG V., « Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks », *Computer Networks*, vol. 38, n° 5, 2002, p. 577–589.