Proposal and Analysis of Adaptive Mobility Management in IP-Based Mobile Networks

Rami Langar, *Member, IEEE*, Nizar Bouabdallah, *Member, IEEE*, Raouf Boutaba, *Senior Member, IEEE*, and Bruno Sericola, *Senior Member, IEEE*

***Abstract*—Efficient mobility management is one of the major challenges for next-generation mobile systems. Indeed, a mobile node (MN) within an access network may cause excessive signaling traffic and service disruption due to frequent handoffs. The two latter effects need to be minimized to support quality of service (QoS) requirements of emerging multimedia applications. In this paper, we propose a new adaptive micro-mobility management scheme designed to track efficiently the mobility of nodes so as to minimize both handoff latency and total signaling cost while ensuring the MN's QoS requirements. We introduce the concept of *residing area*. Accordingly, the micro-mobility domain is divided into virtual *residing areas* where the MN limits its signaling exchanges within this local region instead of communicating with the relatively far away root of the domain at each handoff occurrence. A key distinguishing feature of our solution is its adaptive nature since the virtual residing areas are constructed according to the current network state and the QoS constraints. To evaluate the efficiency of our proposal, we compare our scheme with existing solutions using both analytical and simulation approaches for the 2-D random walk model as well as real mobility patterns. Numerical and simulation results show that our proposed scheme can significantly reduce registration updates and link usage costs and provide low handoff latency and packet loss rate under various scenarios.**

***Index Terms*—QoS provisioning, mobility management, handoff algorithm, micro-mobility, MPLS, Mobile IP, adaptability, performance analysis.**

I. INTRODUCTION

FUTURE wireless networks are expected to provide IP-based coverage and efficient mobility support with end-to-end QoS guarantees. Two enabling factors are considered as crucial: (i) maintaining the network connectivity during node mobility and; (ii) provisioning the network resources required by the Mobile Node (MN) in all the visited subnetworks.

Mobility management protocols are key for service continuity in mobile networks. Mobile IP [1], the Internet Engineering Task Force (IETF) standard, can serve as the basic mobility management in IP-based wireless networks. According to Mobile IP, a MN can change its point of attachment without changing its IP address. To do so, a MN is assigned with a permanent home address in its home network, and will borrow a temporary care-of address (CoA) in each foreign network. The CoA is the foreign agent (FA) IP address of the currently visited foreign network. In this case, the home agent (HA), residing in the MN's home network, will maintain the mapping between the home address and the CoA. Clearly, such mechanism induces long handoff latency and large signaling load when handoffs occur frequently [2]. In this regard, many enhancements to Mobile IP for MNs with frequent handoffs have been proposed in the literature [3]–[11] to ensure service continuity.

The second major factor identified as crucial for the evolution process of future networks is efficient network resources provisioning. This issue has been largely studied in both wired and wireless environments. For instance, MultiProtocol Label Switching (MPLS [12]) addresses today's network backbone requirements effectively by providing a standardized solution that improves packet forwarding performance and designs high performance QoS guaranteed paths. Specifically, MPLS forwards data using labels that are attached to each data packet instead of traditional IP destination lookup. These labels are distributed using a label distribution protocol, which maintains the coherence of label bindings across the network. MPLS with its traffic engineering (TE) attempts to provide a means to manage and enhance network traffic through rigorous analytical studies. As a matter of fact, there is an increasing trend towards the introduction of MPLS in wireless environments [13]–[18].

To meet the requirement of next generation mobile networks, we propose in this paper a new adaptive micro-mobility management scheme called adaptive Master Residing Area (MRA) which alleviates the limitations of previous works in terms of flexibility and adaptability and in the same time benefits from MPLS resource provisioning capability. The key idea behind our proposal is to manage adaptively the node mobility according to its current state and the QoS constraints. To do so, we introduce a new concept called *Residing Area* (RA). The size of RAs is not the same all the time for a particular MN and is dynamically computed according to the two above conditions (i.e., user's position and the QoS constraints). This allows reducing the signaling cost while respecting the specific QoS MN's needs.

To gauge the effectiveness of our proposed scheme, we develop a new analytical model based on Markov chains. We, explicitly, derive the expressions of the signaling cost function of registration updates, the link usage cost and the handoff performance metrics (i.e., handoff latency and packet loss rate) for a general two-dimensional (2-D) random walk mobility model. In addition, simulations are conducted using

Manuscript received February 25, 2008; revised September 30, 2008 and February 18, 2009; accepted March 30, 2009. The associate editor coordinating the review of this paper and approving it for publication was H. Zheng.

R. Langar is with the Computer Science Laboratory of Paris 6 - LIP6, University of Pierre and Marie Curie, UPMC - Paris Universitas, 75016 Paris, France (e-mail: rami.langar@lip6.fr).

R. Boutaba is with the Computer Science Department, University of Waterloo, ON, N2L3G1 Canada (e-mail: rboutaba@uwaterloo.ca).

N. Bouabdallah and B. Sericola are with INRIA, 35042 Rennes, France (e-mail: {nizar.bouabdallah, bruno.sericola}@inria.fr).

Digital Object Identifier 10.1109/TWC.2009.080166

real mobility patterns in order to evaluate the performance of our proposed scheme in real mobility situations. Numerical and simulation results show that our proposal can improve significantly the network performance when compared to existing schemes (Fast Mobile IP [6], MIP-RR [8], Pointer Forwarding (PF) [10], Mobile MPLS [14] and M-MPLS [16]) under various scenarios.

The remainder of this paper is organized as follows. In section II, we discuss related work and position our own. Section III describes our proposed adaptive micro-mobility management scheme. Section IV introduces the analytical model used to evaluate the performance of our proposal. In section V, a comparison between our proposal and existing solutions is drawn using both analytical and simulation results. Finally, section VI concludes this paper.

II. Related Work

As explained before, the standard network layer solution, Mobile IP [1], has several shortcomings, such as: high handoff latency, high global signaling load and scalability issues. These issues are more pronounced in micro-mobility environment, where handoff operations are much more frequent than in the macro-mobility case. To cope with Mobile IP limitations, several interesting solutions have been proposed in the literature [6]–[18]. In the following we discuss some of the most significant ones.

To alleviate the high handoff latency, a typical issue in Mobile IP, a fast handoff scheme, FMIP, for Mobile IP is proposed in [6]. This scheme enables a MN to quickly discover that it has moved to a new subnet (by using a router discovery protocol or some link-specific event) and receives data as soon as its attachment is detected by the new access router. However, the use of a discovery protocol implies modifications to Mobile IP in order to support the discovery of neighboring FAs. As such, extra information exchanges between access routers are needed. In addition, the registration updates cost in FMIP can be excessive, especially for highly mobile nodes and those located far away from their HAs.

In order to tackle the inherent problem of Mobile IP regarding the high signaling cost, the authors in [7] propose a distributed dynamic location management scheme using hierarchical architecture. This scheme can be seen as an extension of the IETF regional registration protocol (MIP-RR [8]) in order to improve its flexibility and adaptability. However, the main difficulty when running these schemes is the computation of the optimal size of the regional network [11]. Moreover, these schemes do not support QoS provisioning, a vital requirement for next generation mobile networks.

Another approach to reduce the signaling cost is the "pointer forwarding" technique [9]–[11]. Accordingly, a pointer is setup between the old and new visited subnet by the MN. The MN is associated with a chain of nodes (i.e., forwarding paths) that connect to its HA during its movement. However to achieve this, the authors in [9] assume that there is a correlation between the communication cost and the geographic distances. As such, the coordinates of each FA along the forwarding chain need to be known by the MN. In addition, the QoS support has not been addressed in these works.

So far, the discussed works have focused mainly on improving the Mobile IP performance by proposing new extensions. We believe that the role of an efficient mobility management protocol must not be limited to the basic operations regarding the MN's connectivity maintenance during its travel. Instead, mobility management must also be able to provision efficiently the network resources. In this perspective, several works [13]–[18] propose the use of MPLS in IP-based wireless access networks to benefit from its QoS, traffic engineering and reliability capabilities.

Specifically, [13] corresponds to an ITU-T recommendation Y.1281, namely "Mobile IP services over MPLS". This recommendation discusses the architectural details on implementing MPLS for both Mobile IPv4 and v6 in 3G All-IP networks. In addition, authors in [14] propose the Mobile MPLS protocol. This scheme aims at improving the scalability of the Mobile IP data forwarding process by removing the need for IP-in-IP tunneling from the HA to the FA using Label Switched Paths (LSPs). However, such schemes suffer from the non-applicability to micro-mobility, as the scope of Mobile IP is more tailored to global mobility (i.e., macro-mobility).

In [15], the authors provide a taxonomy of the different ways in which IP/MPLS may be used in 3G networks. An enhanced LER called the label edge mobility agent (LEMA) is introduced to support LSP-redirection. The scheme has been shown to be scalable and suitable for QoS support [15]. However, the algorithms for choosing the LEMAs for a particular MN appear to be complex, which affects the reliability and the cost of the proposed scheme.

Authors in [16] and [17] attempt to improve the performance of Mobile MPLS [14] using different architectures. Commonly, a Foreign Domain Agent (FDA) is introduced into each MPLS domain to support intra-domain mobility (i.e., inside the regional network). However, with a high mobility rate, the system performance is critically affected by frequent registrations and LSPs setup procedure with the FDA, resulting in excessive signaling traffic, long service delay and the loss of a large amount of in-flight packets. Note that in-flight packets are the packets possibly lost during the handoff period. In addition, most of these works have assumed that all base stations are MPLS capable which may not be always desirable as this implies a significant increase in terms of cost and complexity.

To overcome these limitations, we first proposed in [18] a simple architecture that combines MIP-RR [8] and MPLS [12] protocols and achieves further handoff optimization. To do so, we have adapted the L2 triggers [19] and buffering mechanisms in the context of MPLS environment using a one-dimensional mobility model. In this case, the MN can anticipate the Layer 3 handoff by establishing in advance a new LSP with the desired QoS requirements from the root of the domain to the new subnet. This enables low handoff latency and packet loss rate.

In this paper, we propose a new mobility management scheme, namely Adaptive MPLS-enabled MRA, that offers flexibility and adaptability to the network and in the same time benefits from MPLS resource provisioning capability in a two-dimensional space. Our aim is first to reduce the resource reservation cost by allowing packets to be forwarded through existing LSPs (benefiting then from already reserved resources on the old path); and second to manage adaptively the node mobility in a two-dimensional space. This is performed according to the current MN's position and the QoS constraints by using the virtual residing area concept. As such, several issues that are likely to arise when using the traditional chaining methods in a two-dimensional space, such as loop problems [11], are avoided. In addition, adaptively constructing residing areas will improve the network performance (see section V) since the size of each area will be adjusted based on the user's mobility information, the current network state and the QoS constraints. To the best of our knowledge, we are the first to take into account the specific mobility propriety of users in a two-dimensional space with QoS constraints to manage users' mobility.

III. Adaptive MPLS-enabled MRA

A. Proposed architecture

Adaptive MRA relies on our previous work in [18], where we proposed an architecture based on the integration of MIP-RR [8] and MPLS [12] protocols. We assume that an MPLS access network exists between the Label Edge Router Gateway (LERG) and the Label Edge Router/Foreign Agents (LER/FAs) [see Fig. 1]. The network architecture is based on a two-level hierarchy. At the higher level is the LERG that performs the role of an edge Label Switching Router filtering between intra- and inter-domain signaling. At the second level is the LER/FA connected to several access points (APs) that offer link-layer (L2) connectivity. We distinguish between L2 functionalities of the air interface, which are handled by the AP, and IP-layer mobility (L3 handoff), which occurs when the MN moves between subnets served by different LER/FAs. Note that an LER/FA is the first IP-capable network element seen by the MN.

Two types of handoffs are defined in our scheme: Intra-LER and Inter-LER handoffs. An Intra-LER handoff occurs when the MN moves between two APs managed by the same LER/FA. This kind of handoff is basically L2 handoff. On the other hand, an Inter-LER handoff occurs when a new AP and the old AP are under different LER/FAs. This kind of handoff is typically L3 (network-layer) handoff. In this work, we focus on Inter-LER handoff since it has the most important effect on the handoff performance.

B. Handoff operation and Algorithm of Adaptive MRA

As stated before, our approach is based on the adaptive residing area concept and can be considered as a new alternative to track efficiently the mobility of nodes instead of the pointer forwarding technique. Accordingly, the micro-mobility domain will be divided into virtual residing areas where the MN limits its registration updates within this local region, instead of communicating to the far away LERG node. Explicitly, a specific node called Master FA (MFA) will be delegated by the LERG to manage the mobility of nodes inside the current virtual residing area. Initially, the MFA corresponds to the first FA of the subnet the MN visits. Each time the MN moves to a new subnet inside the current residing area, it registers with the MFA instead of the LERG. Consequently,

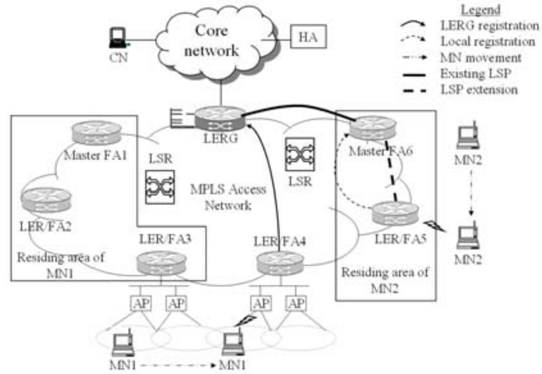


Fig. 1. Architecture of an Adaptive MRA wireless access network

the existing LSP (with QoS requirements) between the LERG and the MFA will be extended to the new visited FA. As such, the resource reservation cost will be reduced since packets can be forwarded through the existing LSPs benefiting from the already reserved resources on the old path. Once the MN goes outside this area, it registers to the LERG. Hence, a new LSP (with new QoS requirements) between the LERG and the new subnet will be established using either the RSVP-TE [20] or CR-LDP [21] protocols and the new visited LER/FA becomes the new MFA of the MN.

As a key distinguishing feature of our handoff mechanism, the virtual residing area around a specific MFA is constructed adaptively according to both the relative position of the current MFA with the LERG and the delay constraint. Indeed, assume that the maximum tolerable delay inside the micro-mobility domain is D_{max}. For the sake of simplicity, D_{max} will be expressed in terms of hops. Each time the MN moves to a new subnet, it compares the length of its indirect path to the LERG through the current MFA with D_{max}. If this distance is equal or less than D_{max}, the MN can register locally to the MFA. In this case, the new visited FA may join the current MFA residing area. Otherwise, it registers directly to the LERG and the new FA becomes the MFA of the new residing area (i.e., the new FA is considered as outside of the previous residing area). Moreover, to minimize the signaling cost, a second condition must be verified. Specifically, a local registration with the MFA is achieved as long as it is cheaper than a LERG registration. Indeed, each time the MN moves to a new subnet, the new LER/FA compares the signaling cost (in terms of hop $\times$ message size) of a registration update to the MFA with that to the LERG. Once the distance between the new visited FA and the LERG is equal or less than the distance between the new FA and the MFA, a LERG registration is preferred.

To illustrate the residing area concept, we consider the simple example presented in Fig. 2, where the LERG node is located at the center of a domain with a radius $R = 3$. We assume also that $D_{max} = 3$. It is worth noting that D_{max} must be at least equal to R. Assume that the current MFA is the subnet S_1^1. The associated residing area will be composed of nine subnets as shown in Fig. 2. These FAs satisfy both conditions regarding the delay and registration cost. Accordingly, as long as the MN remains in this area (i.e., it fulfills the delay constraint and a local registration is cheaper than a LERG one), it carries out a local registration with the MFA. Once it leaves this residing area, it performs a

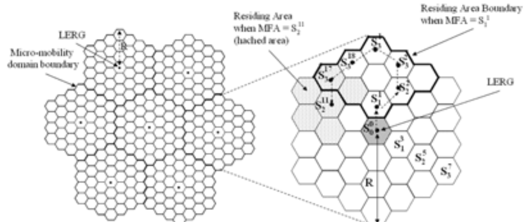


Fig. 2. A model of the simulated cellular network with $R = 3$

LERG registration and the new serving LER/FA becomes the new MN's MFA.

Indeed, the registration updates cost condition is no more verified when the MN enters subnet S_2^{11} (see Fig. 2). Hence, the MN registers to the LERG and updates its address directly to the root of the domain. At the same time, the FA of subnet S_2^{11} becomes the MFA of the new residing area, which will be composed of six subnets as shown in Fig. 2. We can see that the subnet S_3^{17} belongs to both residing areas managed by MFAs S_2^{11} and S_1^1. So using our scheme, the residing areas can overlap. Indeed, the MN can be attached to different MFAs when visiting the subnet S_3^{17}. According to the MN's trajectory, i.e., the tuple (old FA, old MFA), the new visited FA assigns the corresponding MFA to the MN. In our example, when the MN enters to the subnet S_3^{17} with the tuple (S_3^{18}, S_1^1), it registers to the MFA S_1^1. The new state of the MN becomes (S_3^{17}, S_1^1). On the other hand, when the MN visits S_3^{17} while having (S_2^{11}, S_2^{11}) as the current state, it registers to the MFA S_2^{11}. The new MN state is therefore (S_3^{17}, S_2^{11}). This simple example shows clearly the dynamic property of our mobility management scheme.

The basic operations of the adaptive MRA scheme are listed below. When the MN enters for the first time into an MPLS domain, it registers to the LERG through the nearest LER/FA. The latter will be thus configured as the current MFA of the MN. When the MN moves to a new subnet within the same domain, it proceeds as follows.

1) The MN sends a registration message to the new FA. This message contains the IP address of the associated MFA.
2) The new FA checks the existence of the MFA's IP address. If it exists (i.e., the new FA and MFA nodes belong to the same domain), the new FA computes the shortest distance to reach the MFA and the LERG nodes. Accordingly, it performs either a local registration or a LERG registration by verifying both conditions regarding the maximum tolerable delay and the registration cost.
3) Finally, the MN receives a registration reply message either from the LERG or from the MFA, according to the registration type. In the former case, a new LSP will be established between the root of the domain and the new subnet. In this case, the MN's residing area will be renewed, and the new FA becomes the new MFA of the MN. In the latter case, the existing LSP (with already resource reservation) between the LERG and the MFA will be extended to the new subnet. In this case, the new FA will join the existing residing area managed by the MFA. More formally, the adaptive MRA scheme is described by the pseudocode in algorithm 1.

Algorithm 1 Adaptive MPLS-enabled MRA algorithm

```
%Location registration procedures
if (MN enters a new subnet) then
    New LER/FA checks the existence of the MFA's IP address in its routing table;
    if (the address exists) then
        New FA and MFA nodes belong to the same domain: intra-domain mobility;
        New FA computes the shortest distance to the MFA and to the LERG;
        if (d(newFA, MFA) < d(newFA, LERG)) & (d(newFA, MFA) + d(MFA, LERG) ≤ D_max) then
            Perform a local registration: new FA registers to the MFA;
        else
            Perform a LERG registration: new FA registers to the LERG;
            Master FA renewal: new FA becomes the new MFA;
        end if
    else
        New FA and current MFA belong to different domains: inter-domain mobility;
        Perform both LERG and home registrations;
        Master FA renewal: new FA becomes the new MFA;
    end if
end if
```

It is worth noting that our proposed mobility scheme needs all the FAs to maintain the information regarding the distances between them. This information can be computed using two methods. We can either setup these distances in advance or use the coordinates of FAs. In the former case, as the FAs are not mobile components, each one maintains a static table containing the distance that separates from each of the remaining FAs of the same micro-mobility domain. In the latter case, each FA keeps only the information regarding the coordinates of the remaining FAs of the micro-mobility domain. Accordingly, it determines the distance to each of its destinations (in terms of number of hops) through the shortest path. In our work, we have adopted the second method since it is more flexible. Notably, in case of FA failure, the identity of the failed FA needs only to be notified to all the FAs, whereas using the first method, all the distances should be calculated again according to the new topology and notified to each FA.

IV. Analytical Model

In this section, we develop a new analytical model using Markov chains to evaluate the performance of our adaptive MRA scheme. Specifically, we derive the registration updates cost, the link usage and the handoff performance (i.e., handoff delay and packet loss rate) when the mobility of nodes is managed using the adaptive MRA scheme. The elaborated model will be also used to derive the performance metrics for the existing solutions FMIP, Mobile MPLS, MIP-RR and M-MPLS. In the PF case, we will use simulations.

In our study, we consider first a general 2-D random walk model [22] [23]. The wireless network is divided into hexagon-based cells as shown in Fig. 2. Typically, each cell is covered by one LER/FA, called base station in cellular networks. In this case, each subnet is surrounded by six neighbors. The MN can move to one of the neighboring subnets with equal probability p ($p = \frac{1}{6}$). Note that our model is not restricted only to the hexagon cell configuration. It can be used indeed for any arbitrary shaped subnet. In this case, we define the connectivity degree $\delta(S_i)$ of a subnet S_i as the number of neighboring subnets that are one hop away from S_i. Consequently, under the 2-D random walk model, the MN connected to S_i moves to one of the neighboring subnets with equal probability $p = 1/\delta(S_i)$. We note also that, by using different values of the probability p such that $\sum p_j = 1$ ($1 \leq j \leq \delta(S_i)$), different non-random mobility patterns can be generated.

Figure 2 represents a micro-mobility domain with a radius $R = 3$ in a two dimensional space. The domain contains the LERG node surrounded by 3 rings of subnets. Each subnet is referenced by the ring label and its position inside that ring, which determines the exact MN's position with respect to the LERG of the domain. For example, subnets belonging to ring 1 are referenced by S_1^j, $1 \leq j \leq 6$, those belonging to ring 2 are referenced by S_2^j, $1 \leq j \leq 12$, and so on and so forth. To generalize, let i ; $i = 0, 1, \ldots, R$ designate the ith ring away from the LERG node. The LERG node subnet is denoted by S_0^0. Subnets belonging to ring i are referenced by S_i^j, $1 \leq j \leq 6\,i$. Note that the ring label represents the distance between the MN and the LERG.

Let $X(t)$ be the MN's state within the micro-mobility domain at time t defined by the tuple (S_i^j, S_n^m), where S_i^j is the current subnet location and S_n^m is the current MN's MFA. The sojourn time of a MN in a subnet S_i^j has a general distribution (not necessary exponential) with a mean $1/\mu$. Moreover, the sojourn times of a MN in different subnets are independent and have the same mean. Since the MN evolves as a 2-D random walk, the process $X = \{X(t), t \geq 0\}$ is a homogeneous semi-Markov process with state space $\mathcal{S} = \{\ (S_i^j, S_n^m) \ \mid \ 0 \leq i \leq R \ \ , 1 \leq j \leq 6\,i \ \ , S_n^m \in E_{S_i^j}\}$, where $E_{S_i^j}$ is the set of possible MFAs that a MN can register to when it is located in the subnet S_i^j. In other words, $S_n^m \in E_{S_i^j}$ if and only if the subnet S_i^j belongs to the residing area managed by S_n^m, i.e., it satisfies the following relation:

$$S_n^m \in E_{S_i^j} \ \text{ if and only if} \begin{cases} d(S_i^j, S_n^m) < d(S_i^j, \mathit{LERG}) \\ \qquad\qquad \& \\ d(S_i^j, S_n^m) + d(S_n^m, \mathit{LERG}) \leq D_{max} \end{cases} \tag{1}$$

where $d(x, y)$ denotes the shortest path distance (in terms of number of hops) between subnets x and y. Note that the first condition in (1) ensures that a local registration cost is cheaper than a LERG one. The second condition ensures that the MN fulfills the delay constraint.

It is worth noting that for performance analysis purpose, both data and signaling messages are assumed to be routed through the shortest path.

According to our adaptive MRA scheme, the MN's state at time t is completely defined by the tuple $X(t) = (S_i^j, S_n^m)$. Using that information, we can predict exactly the MN's evolution. According to its next location (i.e., visited subnet), the MN can perform either a local registration or a LERG registration. In the latter case, the MFA will be updated and its associated residing area will be created.

We denote by T_0, T_1, T_2, ... the successive times of transitions for X and by Z_0, Z_1, Z_2, ... the successive states visited between these transitions, i.e., for every $k \geq 0$, $Z(k) = X(t)$ if $T_k \leq t < T_{k+1}$, where $T_0 = 0$. According to [24], the embedded process $Z = \{Z(k), k \geq 0\}$ in the transition instants of X is a discrete time homogeneous Markov chain with state space $\mathcal{S}$ and transition probability matrix denoted by P whose transition probabilities are given below.

To get insight into $Z(k)$ and $E_{S_i^j}$, let us revisit the example of Fig. 2. The MN moves from subnet S_0^0 (i.e., the LERG) to subnet S_2^{11}. In this case, the MN begins its trajectory at subnet S_0^0, which represents also the MFA of the current residing area. Hence $Z(0) = (S_0^0, S_0^0)$. When the MN moves to S_1^1, it performs a LERG registration and the new FA of subnet S_1^1 becomes the MFA of the new residing area. As such, $Z(1) = (S_1^1, S_1^1)$. As stated before, the MN achieves always a local registration to the current MFA S_1^1 when it moves to subnets S_2^2, S_3^2, S_3^1, S_3^{18} and S_3^{17} as it remains in the current MFA residing area. So, $Z(k)$ gets the following values: $Z(2) = (S_2^2, S_1^1)$, $Z(3) = (S_3^2, S_1^1)$, ..., $Z(6) = (S_3^{17}, S_1^1)$. In this case, we can see that S_1^1 belongs to $E_{S_1^1}$, ..., $E_{S_3^{17}}$.

In the following, we calculate the transition probability matrix P of the process Z. To derive the transition probabilities when leaving a generic state (S_i^j, S_n^m), we distinguish between two cases: either the current MN's subnet S_i^j is located at the ith ring far away from the LERG node with $1 \leq i \leq R-1$ or it is located at the boundary of the micro-mobility domain (i.e., $i = R$). In the latter case, the MN may leave the current domain and enters a new one. The transition probabilities regarding each case are derived as follows.

1) Case1: S_i^j is not a boundary subnet (i.e., $i \neq R$)

Let $S_{i'}^{j'}$ denote the next visited subnet by the MN. $S_{i'}^{j'}$ is one of the six neighbors that surrounds the current subnet S_i^j. Hence, the MN moves to subnet $S_{i'}^{j'}$ with a probability p. According to whether $S_{i'}^{j'}$ belongs or not to the current residing area managed by the MFA S_n^m, we can identify the next MN's state. Specifically, if it is the case (i.e., $S_n^m \in E_{S_{i'}^{j'}}$), the MN will transit to the state $(S_{i'}^{j'}, S_n^m)$. In this case, the MN performs a local registration to the current MFA S_n^m. Henceforth, we denote by $\mathcal{A}$ the event that $S_n^m \in E_{S_{i'}^{j'}}$ (see (1)). On the other hand, if $\mathcal{A}$ is not satisfied (i.e., $\bar{\mathcal{A}} = S_n^m \notin E_{S_{i'}^{j'}}$), the MN registers to the LERG and the new FA becomes the MFA of the new residing area. As such, the MN transits to state $(S_{i'}^{j'}, S_{i'}^{j'})$. Consequently, for $i \neq R$, we have:

$$\begin{cases} P\Big((S_i^j, S_n^m), (S_{i'}^{j'}, S_n^m)\Big) = p \cdot 1_{\mathcal{A}} \\ P\Big((S_i^j, S_n^m), (S_{i'}^{j'}, S_{i'}^{j'})\Big) = p \cdot 1_{\bar{\mathcal{A}}} \end{cases}$$

where $1_{\mathcal{A}}$ (respectively $1_{\bar{\mathcal{A}}}$) is the indicator function of the condition $\mathcal{A}$ (respectively $\bar{\mathcal{A}}$), i.e., it is equal to 1 if the condition $\mathcal{A}$ (respectively $\bar{\mathcal{A}}$) is true and 0 otherwise.

2) Case2: S_i^j is a boundary subnet (i.e., $i = R$)

In this case, the MN may leave the current mobility domain. Accordingly, we distinguish between two sub-cases:

2.a: The MN remains in the same domain: In this case, the MN behaves in a similar way to case1 and we get the same transition probabilities.

2.b: The MN moves to an adjacent domain: This happens with a probability $f \times p$, where f is the number of neighboring subnets to S_R^j located in the adjacent micro-mobility domains. According to our hexagon subnet model, f can take two values: 2 or 3. In our study, we assume that all the domains have the same radius R. So, like the old subnet S_R^j, the new visited subnet will be R hops far away from the new LERG (i.e., again at the boundary of the new domain). As such, the MN will be assumed to enter the subnet S_R^j of the new domain. When the MN enters the new domain, it registers to the new LERG and the new visited subnet becomes the MFA. As a result, the MN transits to the state (S_R^j, S_R^j) with a probability $f \times p$, i.e., $P\Big((S_R^j, S_n^m), (S_R^j, S_R^j)\Big) = f \times p$.

Based on the different cases listed above, we derive the transition probability matrix $P = [p_{ij}]$. The Markov chain Z being irreductible and aperiodic and the sojourn times of X having the same mean, we have (see for instance [24]), for every $h, s \in \mathcal{S}$,

$$\lim_{t \longrightarrow \infty} \Pr\{X(t) = s | X(0) = h\} = \Pi_s$$

where $\Pi = [\Pi_s]$ is the steady state distribution of the Markov chain Z, which satisfies:

$$\Pi P = \Pi \;\text{ and }\; \sum_{s \in \mathcal{S}} \Pi_s = 1 \tag{2}$$

As such, we get the steady state probabilities of the process X. Using these results, we evaluate hereafter the performance of our proposed adaptive MRA scheme. Building on these results, we can also evaluate analytically the existing solutions (i.e., FMIP, Mobile MPLS, MIP-RR and M-MPLS) except the PF scheme, which will be studied using simulations as shown in the next section. We will derive first the signaling cost of registration updates as well as the link usage cost. Moreover, we will calculate the handoff performance (i.e., handoff delay and packet loss rate).

A. Link Usage Cost

Let LU denote the link usage in the micro-mobility domain, which is the average number of links used for packet delivery between the MN and the LERG. In our adaptive MRA case, packets exchanged between the MN and the LERG have to pass-through the MFA. Hence the LU can be written as follows:

$$\begin{aligned} LU(\text{Adaptive MRA}) &= \sum_{s \in \mathcal{S}} \Pi_s \Big(d(\textit{Subnet(s)}, \textit{MFA(s)}) + d(\textit{MFA(s)}, \textit{LERG}) \Big) \\ &= V_{LU} \times \Pi \end{aligned} \tag{3}$$

where $s = (\textit{Subnet(s)}, \textit{MFA(s)}) = (S_i^j, S_n^m)$ and V_{LU} denote the link usage vector of all states $s \in \mathcal{S}$.

In PF, packets have to traverse both the connection binding the LERG to the MFA and the forwarding chain binding the MFA to the MN. The LU metric in PF will be derived through simulations since the movement of MNs in a 2-D area is not a Markovian process as the MN's evolution depends on its mobility history. In FMIP, Mobile MPLS, MIP-RR and M-MPLS schemes, packets are delivered using the shortest path routing between the LERG and the MN. Hence, the link usage is the same and can be given by:

$$\begin{aligned} &LU(\text{FMIP / Mobile MPLS / MIP-RR / M-MPLS}) \\ &\quad = \sum_{0 \le i \le R} \; \sum_{1 \le j \le 6\,i} \pi(S_i^j) \times d(S_i^j, \textit{LERG}) \\ &\quad = \sum_{0 \le i \le R} \; \sum_{1 \le j \le 6\,i} i \times \pi(S_i^j) \end{aligned} \tag{4}$$

where $\pi(S_i^j)$ denotes the steady probability that the MN is located at the physical subnet S_i^j. $\pi(S_i^j)$ can be expressed as follows, using the steady state probabilities $\Pi_s = \Pi(S_i^j, S_n^m)$:

$$\pi(S_i^j) = \sum_{S_n^m \in E_{S_i^j}} \Pi(S_i^j, S_n^m) \tag{5}$$

B. Registration Updates Cost

Let C_u denote the signaling cost of registration updates when a L3 handoff occurs. It represents the traffic load of signaling messages (hop $\times$ message size) exchanged in the network when the MN moves to a new subnet. In adaptive MRA, a local registration followed by a LSP procedure setup between the MFA and the new FA are required as long as the MN remains in the same residing area. Otherwise, a LERG registration with a new LSP setup is performed. Additional registration to the MN's HA is also needed, each time the MN moves to a new domain. In this regard, the average registration updates cost when transiting to a state $s = (\textit{Subnet(s)}, \textit{MFA(s)}) = (S_i^j, S_n^m)$ can be written as follows, using the transition probability matrix P and the steady state probability vector Π.

$$cost(s) = \begin{cases} \dfrac{\sum_{i \in \mathcal{S}} P(i,s) \times \Pi_i \times \mathcal{C}(i,s)}{\sum_{i \in \mathcal{S}} P(i,s) \Pi_i} & \text{if } \textit{Subnet(s) = MFA(s)} \\ C_{local}(s) & \text{otherwise.} \end{cases} \tag{6}$$

where

$$\mathcal{C}(i,s) = \begin{cases} C_{lerg}(s) + C_{home} & \text{if } \textit{Subnet(s) = Subnet(i)} \\ C_{lerg}(s) & \text{if } (\textit{Subnet(s)} \neq \textit{Subnet(i)}) \\ & \;\text{and } (\textit{MFA(s)} \neq \textit{MFA(i)}) \\ 0 & \text{otherwise.} \end{cases}$$

with $C_{home} = 2\, m_u\, d(\textit{LERG}, \textit{HA})$, $C_{lerg}(s) = 2\,(m_u + m_l) \times d\big(\textit{Subnet(s)}, \textit{LERG}\big)$, and $C_{local}(s) = 2\,(m_u + m_l)\, d\big(\textit{Subnet(s)}, \textit{MFA(s)}\big)$, where m_u is the average size of signaling messages for the registration updates and m_l is the average size of a label message for LSP setup.

The total registration updates cost can be thus written as:

$$C_u(\text{Adaptive MRA}) = \sum_{s\in\mathcal{S}} \Pi_s cost(s) = V_{C_u} \times \Pi \quad (7)$$

where $V_{C_u} = [cost(s)]$ denote the registration updates cost vector of all states $s \in \mathcal{S}$.

Considering the competing schemes, the average registration updates cost can be expressed as follows using (4) and (5). In PF, simulations will be used to evaluate the registration cost.

$$\begin{cases} C_u(\text{FMIP}) = 2\ m_u\ LU(\text{FMIP}) + C_{home} \\ C_u(\text{Mobile MPLS}) = 2\ (m_u + m_l)\ LU(\text{Mobile MPLS}) \\ \qquad\qquad +C_{home} + L_{home} \\ C_u(\text{MIP-RR}) = 2\ m_u\ LU(\text{MIP-RR}) + \alpha\ C_{home} \\ C_u(\text{M-MPLS}) = 2\ (m_u + m_l)\ LU(\text{M-MPLS}) + \alpha\ C_{home} \end{cases} \quad (8)$$

where $L_{home} = 2\ m_l\ d(LERG, HA)$ is the signaling cost needed to setup a LSP between the HA and the LERG, and $0 \leq \alpha \leq 1$ denotes the probability of moving out of a micro-mobility domain. This probability can be derived using the elaborated model as shown below.

$$\alpha = \sum_{s\in\mathcal{S}} \Pi_s H(s) \quad (9)$$

where $H(s)$ denotes the probability of entering a new domain when transiting to the state s. It corresponds also to the probability of performing a home registration in the adaptive MRA case when the MN moves to the state s. $H(s)$ can be written as:

$$H(s) = \begin{cases} \dfrac{\sum_{i\in\mathcal{S}} P(i,s) \times \Pi_i \times h(i,s)}{\sum_{i\in\mathcal{S}} P(i,s)\Pi_i} & \text{if } \textit{Subnet(s) = MFA(s)} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

where

$$h(i,s) = \begin{cases} 1 & \text{if } \textit{Subnet(s) = Subnet(i)} \\ 0 & \text{otherwise.} \end{cases}$$

C. Average handoff time

For convenience, we define the following parameters for our handoff analysis.

Parameters:

- t_s average session connection time;
- t_r average FA resident time;
- T_{ad} Time interval for a FA to send agent advertisements;
- N_h average number of L3 handoff during a session (i.e., $N_h = t_s/t_r$);
- B_w Bandwidth of the wired link;
- B_{wl} Bandwidth of the wireless link
- L_w Latency of the wired link (propagation delay);
- L_{wl} Latency of the wireless link (propagation delay);
- P_t Routing or label table lookup and processing delay;
- λ downlink packet transmission rate;

Let $t\big(m, d(x,y)\big)$ denote the time that takes a message of size m to be forwarded from x to y via both wired and wireless links. Since these links are assumed to be not congested, $t\big(m, d(x,y)\big)$ can be expressed as follows:

$$t\big(m, d(x,y)\big) = c + d(x,y) \times (\frac{m}{B_w} + L_w) + \big(d(x,y)+1\big) \times P_t \quad (11)$$

$$\text{where } c = \begin{cases} \frac{m}{B_{wl}} + L_{wl} & \text{if } x = \text{MN} \\ 0 & \text{otherwise.} \end{cases}$$

The average handoff time (T_h) can be expressed as the sum of two terms: disruption time (T_d) and completion time (T_c).

Disruption time (T_d): It is the average time that a MN spends without connection to any LER/FA during the handoff process. In other words, it is the time between the moment that the MN disconnects from the old FA to the moment that it connects to the new one. It is easy to see that the disruption time becomes null when the overlapping area is large enough. The worst case value for this quantity is equal to the L3 beacon period (T_{ad}). T_d can be given by the following expression:

$$T_d = \begin{cases} 0 & \text{if } T_{overlap} \geq T_{ad} \\ \frac{1}{T_{ad}} \int_0^{T_{ad}-T_{overlap}} f(t)\, dt & \text{otherwise} \end{cases} \quad (12)$$

where $T_{overlap}$ denotes the time spent by the MN in the overlapping area and $f(t) = T_{ad} - T_{overlap} - t$. Hence, T_d is equal to:

$$T_d = \begin{cases} 0 & \text{if } T_{overlap} \geq T_{ad} \\ \frac{T_{ad}}{2} + \frac{T^2_{overlap}}{2T_{ad}} - T_{overlap} & \text{otherwise} \end{cases} \quad (13)$$

Note that the disruption time has the same expression for all underlying protocols as expressed in (13) except for FMIP and adaptive MRA. For the two latter schemes, this time corresponds only to the physical disconnection from the old AP until the connection to the new one (i.e., L2 handoff delay) thanks to the use of L2 triggers [19].

Completion time (T_c): It is the time to complete the registration update. The expression of T_c when considering the proposed adaptive MRA as well as the competing schemes can be written as (14), shown at the top of the next page, where $0 \leq \alpha \leq 1$ is, as defined in subsection IV-B, the probability of performing an inter-domain mobility, $0 \leq \beta \leq 1$ denotes the probability of renewing the forwarding chain when a L3 handoff occurs in the PF case. It is also the probability to register to the HA at each L3 handoff in the PF case. This probability will be derived through simulations. And finally, $0 \leq \gamma \leq 1$ denotes the probability of performing a LERG registration at a L3 handoff in the adaptive MRA case. Using the elaborated model, this probability can be given by:

$$\gamma = \sum_{s\in\mathcal{S}} \Pi_s G(s) \quad (15)$$

where $G(s)$ denotes the probability of performing a LERG registration in the adaptive MRA when the MN moves to the state s. $G(s)$ can be derived as:

$$G(s) = \begin{cases} \dfrac{\sum_{i\in\mathcal{S}} P(i,s) \times \Pi_i \times g(i,s)}{\sum_{i\in\mathcal{S}} P(i,s)\Pi_i} & \text{if } \textit{Subnet(s) = MFA(s)} \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

$$\begin{cases} T_c(\text{FMIP}) = 2\, t\big(m_u, d(\text{MN}, \text{HA})\big) \\ T_c(\text{Mobile MPLS}) = 2\, t\big(m_u, d(\text{MN}, \text{HA})\big) + 2\, t\big(m_l, d(\text{FA}, \text{HA})\big) \\ T_c(\text{MIP-RR}) = 2\, t\big(m_u, d(\text{MN}, \text{GFA})\big) + \ 2\, \alpha\, t\big(m_u, d(\text{GFA}, \text{HA})\big) \\ T_c(\text{M-MPLS}) = 2\, t\big(m_u, d(\text{MN}, \text{LERG})\big) + 2\, t\big(m_l, d(\text{FA}, \text{LERG})\big) + \ 2\, \alpha\, t\big(m_u, d(\text{LERG}, \text{HA})\big) \\ T_c(\text{PF}) = 2\, (1-\beta)\, t\big(m_u, d(\text{MN}, \text{FA}) + d(\text{old FA}, \text{new FA})\big) + \ 2\, \beta\, t\big(m_u, d(\text{MN}, \text{HA})\big) \\ T_c(\text{Adaptive MRA}) = 2\, (1-\gamma)\, \Big[t\big(m_u, d(\text{MN}, \text{FA}) + d(\text{FA}, \text{MFA})\big) + \ t\big(m_l, d(\text{FA}, \text{MFA})\big) \Big] \\ \qquad\qquad + 2\, \gamma\, \Big[t\big(m_u, d(\text{MN}, \text{LERG})\big) + t\big(m_l, d(\text{FA}, \text{LERG})\big) \Big] + \ 2\, \alpha\, t\big(m_u, d(\text{LERG}, \text{HA})\big) \end{cases} \tag{14}$$

with

$$g(i,s) = \begin{cases} 0 & \text{if } \textit{Subnet(s)} \neq \textit{Subnet(i)} \\ & \text{and } \textit{MFA(s)} = \textit{MFA(i)} \\ 1 & \text{otherwise.} \end{cases}$$

D. Total packet loss during a session

The total packet loss (Pkt_loss) during a session is defined as the sum of lost packets per MN during all handoffs. In Mobile MPLS, MIP-RR, M-MPLS and PF all in-flight packets will be lost during the handoff time due to the lack of any buffering mechanism. In both FMIP and adaptive MRA, meanwhile, in-flight packets would be lost until the buffering mechanism is initiated [18]. Note that this mechanism is initiated at the old FA level in our mobility protocol while it is initiated at the new FA level in the FMIP scheme. Pkt_loss for each scheme can be expressed as follows:

$$\begin{cases} Pkt\_loss(\text{FMIP}) = t\big(m_u, d(\text{MN}, \text{FA}) + d(\text{FA}, \text{FA})\big) \times \lambda \times N_h \\ Pkt\_loss(\text{Adaptive MRA}) = t\big(m_u, d(\text{MN}, \text{FA})\big) \times \lambda \times N_h \end{cases} \tag{17}$$

For the remaining schemes, Pkt_loss can be written as:

$$Pkt\_loss(.) = T_h(.) \times \lambda \times N_h \tag{18}$$

E. Buffer size requirement

According to our proposed adaptive MRA scheme, a buffer is required at the old LER/FA to store in-flight packets during each handoff operation. Indeed, the buffering mechanism is activated when the current LER/FA receives the Movement signaling message [18]. This message notifies an imminent L2 handoff occurrence. On the other side, the buffering mechanism is disabled when the old LER/FA is notified by the MN through the new FA to forward in-flight packets. The buffer size requirement (Buf_size) for FMIP and adaptive MRA is determined as follows:

$$\begin{cases} Buf\_size(\text{FMIP}) = \big[\ T_d + t\big(m_u, d(\text{MN}, \text{FA})\big)\ \big] \times \lambda \\ Buf\_size(\text{Adaptive MRA}) \\ \qquad = \big[\ T_d + t\big(m_u, d(\text{MN}, \text{FA}) + d(\text{FA}, \text{FA})\big)\ \big] \times \lambda \end{cases} \tag{19}$$

V. Numerical & Simulation Results

In this section, we compare our proposal with respect to the FMIP [6], MIP-RR [8], PF [10], Mobile MPLS [14] and M-MPLS [16] schemes through both simulations and analytical approaches. To evaluate the link usage and registration updates cost by simulations, we develop our own discrete-event simulator. We also derive by simulations the handoff latency and packet loss rate for the different protocols. In this case, the ns-2 simulator [25] is used.

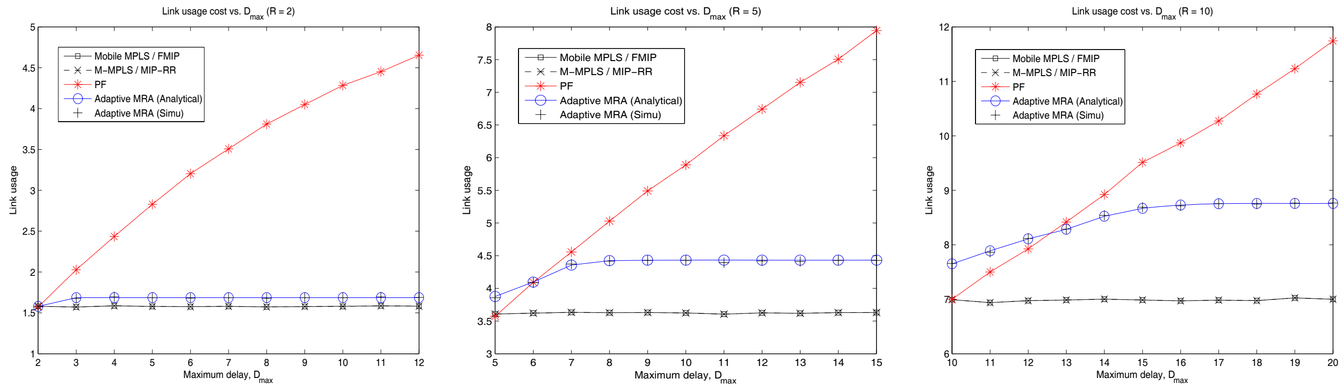
The simulation environment consists of a cellular network formed by adjacent micro-mobility domains with the same radius R, as shown in Fig. 2. The mobility of nodes is simulated using either a random walk model or real mobility patterns. In the former case, the random walk is generated based on the ns-2 code in [26]. In the latter case, we used the mobility traces of the Dartmouth College campus Wifi network [27]. Two scenarios are considered. In the first one, only one MN moves between neighboring subnets and receives 64 kb/s downlink packets. In this case, there is no background traffic in all visited subnets. That is, only one connection exists in each subnet at any time. In the second scenario, we investigated the case of multiple connections in each subnet, which is more likely to be the case in real networks. Specifically, a MN moving at a low speed (i.e., 1 m/s) is assumed to be present at each subnet and receives 64 kb/s background traffic. Then, we study the performance of a MN moving between such subnets at a speed of 4 m/s and receiving 64 kb/s downlink packets.

According to each mobility scheme, the average registration cost per handoff, the link usage and the handoff performance are calculated. All the simulation results given below were obtained with very narrow 97.5% confidence intervals. The parameter settings in our experiments are listed in table I, as in [28].

To get an insight into the accuracy of our results under the 2-D random walk mobility model and the computation gain achieved through the analytical study, let us consider table II. This table shows the analytical and simulation results regarding the link usage and registration updates costs in the adaptive MRA case when $D_{max} = R$ and when using the 2-D random walk. In the table, R is varied from 2 to 10. The simulation for each value of R is run until a relative error less than 0.1% is obtained with respect to the exact analytical results. The computation time for both analytical and simulation results is listed in table II. These times are computed using a PC with 3.2 GHz of CPU and 2.00 GB of RAM. The reported results show that the analytical model achieves great time savings compared to the simulation approach.

In all figures, we can see that the analytical and simulation curves regarding the 2-D random walk model for the adaptive MRA scheme coincide, which illustrates the accuracy of our models. We note also that in the remaining schemes, simulation results coincide with analytic data too. In view of this, we only present the analytic curves for these schemes.

Figure 3 depicts the link usage cost of all underlying proto-



(a) (b) (c)

Fig. 3. Link usage cost vs. D_{max} using the 2-D random walk mobility model: (a) Radius $R = 2$, (b) Radius $R = 5$, (c) Radius $R = 10$

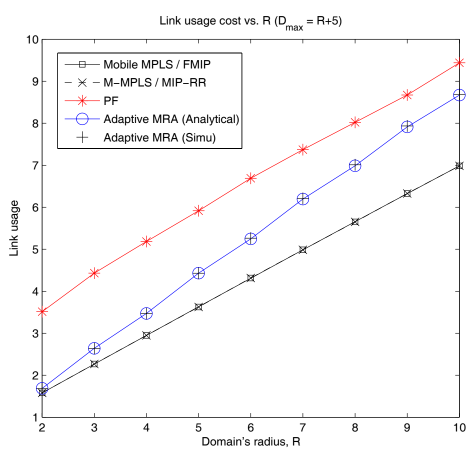


Fig. 4. Link usage cost vs. R using the 2-D random walk mobility model

TABLE I
PARAMETER SETTINGS

Parameter	Value	Parameter	Value
t_s	1000 sec	$d(HA, LERG)$	10
t_r	$5 \sim 50$ sec	P_t	10^{-6} sec
T_{ad}	1 sec	B_w	100 Mbps
m_u	48 bytes	B_{wl}	11 Mbps
m_l	28 bytes	L_w	1 msec
R	$2 \sim 10$	L_{wl}	2 msec
D_{max}	$2 \sim 20$	λ	64 Kbps

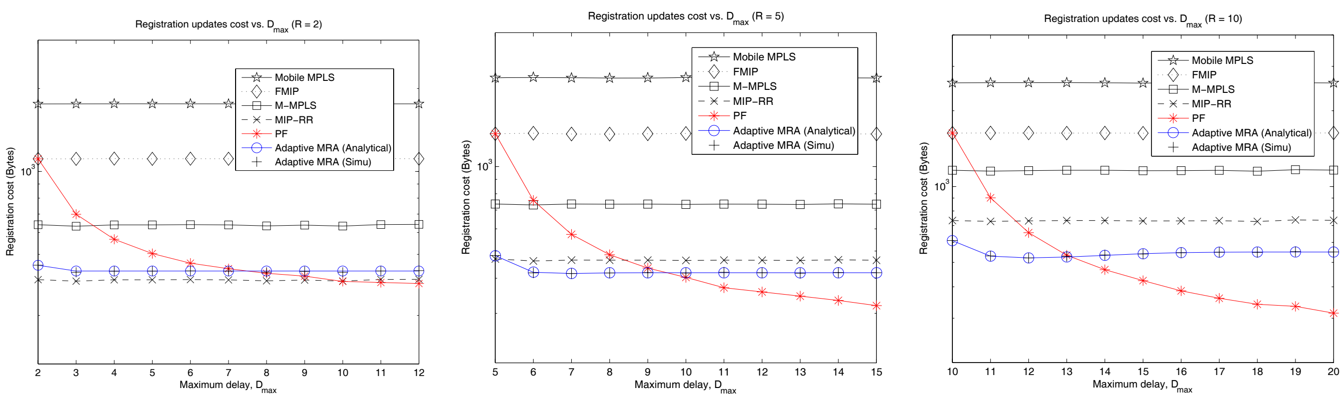
cols as a function of D_{max} under the 2-D random walk model. We considered three values of the radius R in our simulations: $R = 2$, $R = 5$ and $R = 10$, as shown in Figs. 3(a), 3(b) and 3(c), respectively. These values of R are representative of small, medium and large micro-mobility domains. We can observe that the link usage cost with FMIP, Mobile MPLS, MIP-RR and M-MPLS schemes is the same and insensitive to D_{max} since, in these cases, packets are delivered using the shortest path from the LERG node to the current serving LER/FA. In the PF case, the link usage cost increases due to the additional cost introduced by the forwarding chain. This increase grows dramatically with D_{max}. In our adaptive MRA case, the link usage cost is reduced considerably, compared to the PF scheme, notably when D_{max} is large. Indeed, the additional cost introduced by the residing area is less important than the one introduced by the PF forwarding chain, whose length can reach $D_{max} - R$. In contrast, the maximum length introduced by the residing area is $R-1$. As such, the longest path between the MN and the LERG in the MRA case is $2R-1$. In view of this, the link usage cost in MRA remains constant for values of $D_{max} \geq 2R-1$ as shown in Fig. 3.

Figure 4 shows the link usage cost as a function of the micro-mobility domain radius R, when using the 2-D random walk model. The value of D_{max} is set equal to R, $R+5$ and $R+10$ in our simulations. These values of D_{max} are representative of stringent delay sensitive, delay sensitive and delay tolerant applications, respectively. Due to space limitation, only the case of $D_{max} = R+5$ is represented. Two main finding can be identified. First, the link usage in PF and MRA schemes is again larger than the optimal cost of remaining schemes. Second, this increase is insignificant in the MRA case compared to the PF case. The gain achieved by the MRA policy over the PF scheme increases with D_{max}. As a result, the PF scheme is penalizing in term of link usage cost in the case of insensitive delay applications. This issue is alleviated thanks to the MRA scheme. Based on the results of Figs. 3 and 4, we can conclude that the mobility tracking achieved in the MRA case is more efficient from link usage perspective than the PF scheme.

Figure 5 plots the different registration updates cost at every L3 handoff as a function of D_{max}. As before, we consider three values of R (i.e., 2, 5 and 10). Likewise the link usage, the registration update cost in FMIP, Mobile MPLS, MIP-RR and M-MPLS is insensitive to D_{max}. In these cases the shortest path between the LERG and the MN is always used to forward the registration update packets at each L3 handoff. Mobile MPLS (respectively M-MPLS) has a higher cost than FMIP (respectively MIP-RR), due to the additional signaling cost needed to establish a LSP between the HA (respectively LERG) and the new visited FA at each L3 handoff. Adaptive MRA, on the other hand, reduces the registration cost since some expensive LERG registrations are replaced by low-cost local registrations, even though we have accounted for both MPLS and handoff signaling in computing the total signaling cost. In this case, the registration updates for MRA is a convex function of D_{max}, where the minimum cost is obtained for a given D^{opt}_{max}. For instance, when $R = 10$, we found from our simulations that $D^{opt}_{max} = 12$. For $R = 5$, $D^{opt}_{max} = 7$ as shown

TABLE II
AVERAGE COST OF C_u AND LU FOR ADAPTIVE MRA: ANALYTICAL AND SIMULATION RESULTS

Radius R	2	3	4	5	6	7	8	9	10
LU (Analytical)	1.5789	2.3274	3.1433	3.8765	4.6216	5.3863	6.1445	6.8842	7.6498
LU (Simu)	1.5802	2.3265	3.1419	3.877	4.6244	5.3864	6.1387	6.8904	7.6461
C_u (Analytical)	456.3963	449.2111	460.7756	481.2784	505.6283	530.6132	556.5164	583.3888	609.6725
C_u (Simu)	456.0715	449.0087	460.6031	481.5721	505.8367	531.0735	556.6528	583.3818	609.8191
Time (Analytical) (sec)	0.53	2.41	11.5	49.77	132.13	412.47	1516.11	2953.63	5400.75
Time (Simu) (sec)	158.06	359.22	804.5	1663.55	2502.53	2759.75	4145.11	12796.98	14040.14



(a) (b) (c)

Fig. 5. Registration updates cost vs. D_{max} using the 2-D random walk mobility model: (a) Radius $R = 2$, (b) Radius $R = 5$, (c) Radius $R = 10$

in Fig. 5. Specifically, using (9) and (14), the registration updates cost of adaptive MRA can be written as follows:

$$\begin{aligned} C_u(\text{Adaptive MRA}) &= (1-\gamma)\,\bar{C}_{local} + \gamma\,\bar{C}_{LERG} + \alpha\,C_{home} \\ &= (1-\gamma)\,\bar{d}_{local}\,m_{sig} + \gamma\,\bar{d}_{LERG}\,m_{sig} + \alpha\,C_{Home} \quad (20) \end{aligned}$$

where $m_{sig} = 2 \times (m_u + m_l)$, $\bar{d}_{local}$ (respectively $\bar{d}_{LERG}$) denotes the average distance used by the MN to perform a local registration to its MFA (respectively a LERG registration to the root of the domain), and γ and α denote the probability to register to the LERG and HA, respectively, when a L3 handoff occurs. Using (16), $\bar{d}_{local}$ and $\bar{d}_{LERG}$ can be written as:

$$\bar{d}_{local} = \frac{1}{1-\gamma} \sum_{s \in \mathcal{S}} \Pi_s \left(1 - G(s)\right)\, d\big(\mathit{Subnet(s)}, \mathit{MFA(s)}\big) \quad (21)$$

$$\bar{d}_{LERG} = \frac{1}{\gamma} \sum_{s \in \mathcal{S}} \Pi_s\, G(s)\, d\big(\mathit{MFA(s)}, \mathit{LERG}\big) \quad (22)$$

The expression in (20) exhibits clearly the convex behavior of the MRA registration updates cost with D_{max}. On the first hand $(1-\gamma)\,\bar{d}_{local}$ increases with D_{max}. On the other hand $\gamma\,\bar{d}_{LERG}$ decreases with D_{max}. Hence, the trade-off between these two amounts leads to the optimal registration updates cost. Note that the third quantity in expression (20) (i.e., $\alpha\,C_{Home}$), is insensitive to D_{max}. The rationale behind such finding is as follows. Assume that maximal tolerable delay inside the micro-mobility domain is $D_{max} > D_{max}^{opt}$. In this case, it is better for the network administrator to regulate its MRA management protocol according to D_{max}^{opt} instead of D_{max}. In doing so, it respects the D_{max} constraint since $D_{max}^{opt} < D_{max}$. Moreover, it minimizes the registration updates cost as well as the link usage cost. Recall that the link usage cost increases with D_{max} as shown in Fig. 3.

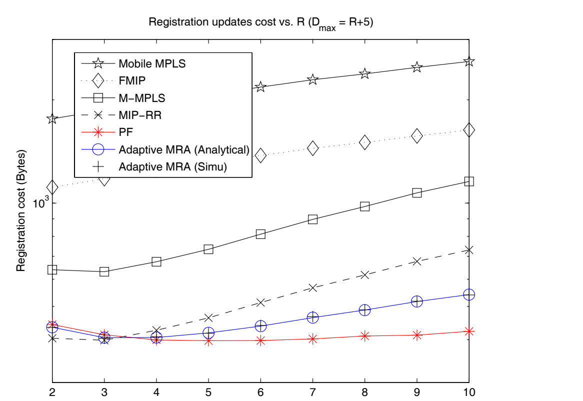


Fig. 6. Registration updates cost vs. R using the 2-D random walk mobility model

Figure 6 shows the effect of the domain radius R on the registration updates cost for different schemes, when using the 2-D random walk model. Again, we presented only the case of $D_{max} = R+5$ due to space limitation. We can observe that the cost of registration updates for FMIP, Mobile MPLS, MIP-RR and M-MPLS schemes increases with R, since the average distance between the MN and the LERG increases with R. Considering the adaptive MRA scheme, the registration cost is reduced except when $R = 2$, where the MIP-RR scheme allows the lowest cost. In such very small size domain context, the residing area concept has no effect. In this case, the slight increase of the signaling cost compared to MIP-RR is due to the additional cost introduced by the use of MPLS.

Again, we observe in this figure a convex behavior of the MRA registration cost with R and the minimum cost is obtained for R_{opt} (in our example, $R_{opt} = 3$). Indeed, the home registration frequency decreases with the increase of

TABLE III
AVERAGE HANDOFF TIME IN MSEC

	Single MN						Multiple MNs					
	2-D random walk			Real mobility patterns			2-D random walk			Real mobility patterns		
D_{max}	10	12	20	10	12	20	10	12	20	10	12	20
Mobile MPLS	69.8016	70.3187	71.4239	83.7984	85.0537	85.9492	87.6760	87.2635	89.0330	105.2601	106.7338	107.5957
FMIP	38.9828	38.2416	38.7950	54.9900	55.6185	56.0669	55.9942	55.6139	56.8428	66.2171	67.2634	67.9569
M-MPLS	35.3039	35.2712	35.1563	51.8685	52.3506	52.3248	44.8673	43.6608	44.5600	67.6214	68.0131	67.6372
MIP-RR	21.3693	21.4086	21.2694	33.1410	33.4190	33.3799	30.9388	30.1746	30.7321	48.1822	48.9734	48.3556
PF	40.1176	19.8281	12.5569	45.4063	28.0241	22.9650	59.0367	28.9158	17.9659	67.0067	40.3667	33.7904
Adaptive MRA	22.1752	19.4276	20.2310	40.1309	31.1483	30.5384	28.3896	26.7574	25.9681	55.7446	44.4585	43.8784

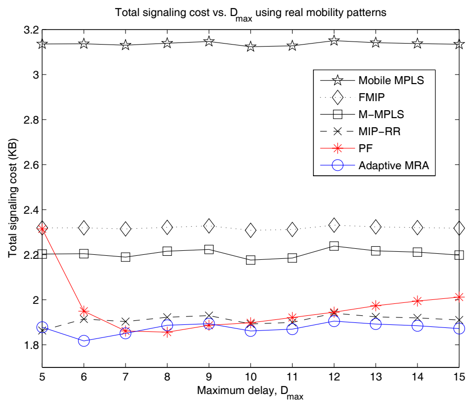


Fig. 7. Total signaling cost vs. D_{max} using real mobility patterns

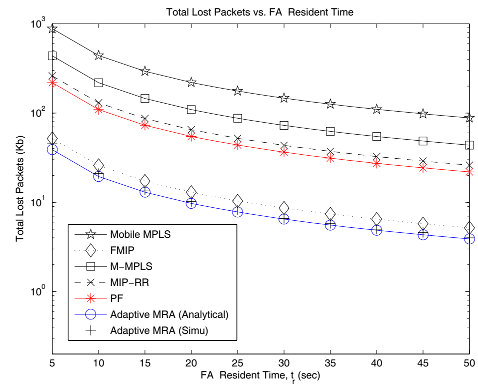


Fig. 8. Total lost packets during a session

R, since the probability that a MN roams under the same LERG increases. On the other hand, both LERG and local registrations cost increases with R, since the average distance between the MN and both the LERG and the MFA increases. In view of this, the optimal cost turns out as a trade-off between these two opposite requirements.

Figure 7 plots the total signaling cost of different schemes as a function of D_{max}, when using real mobility patterns. From this figure, we can notice that our proposal surpasses previous approaches performance wise, especially for low and high values of D_{max}. The achieved gain can attain 20% and 7% compared to PF and MIP-RR, respectively.

The average handoff time values for different schemes are reported in table III for both mobility patterns (i.e., random walk and real mobility) and both scenarios (i.e., single MN and multiple MNs). Each value was obtained by averaging 100 consecutive simulations. From this table, we can see that both adaptive MRA and PF schemes provide the lowest average handoff time since registration updates are often carried out locally (i.e., with the MFA in adaptive MRA and with the previous FA in PF). This enables the MRA scheme to outperform the remaining solutions although an extra time is needed in this case to setup LSP paths at each L3 handoff. Note that the handoff delay in PF is lower than in the MRA scheme when D_{max} takes large values. In this case, the local registrations become more frequent. However, this happens at the cost of considerable increase in the link usage cost as shown in Figs. 3 and 4.

It is important to note that the handoff delay is increased when using multiple MNs for all studied schemes and for both mobility patterns. This is related to the increase of the time needed to deliver packets to the MNs, since the available wireless links will be shared by all users. However, this increase is less pronounced in adaptive MRA for two main reasons. First, the use of residing areas allows further handoff optimization. Second, MNs that are located in the same subnet can use the same established LSP. So in this case, the time needed to establish the required path will be saved.

Figure 8 shows the total packet losses for the session duration with respect to different schemes. We can observe that the total packet losses for all approaches increases when the MN handoffs frequently (i.e., when the FA resident time is short). Notice that Mobile MPLS has the largest amount of packet losses since it has the longest handoff delay. In contrast, adaptive MRA provides the smallest amount of packet losses thanks to the buffering mechanism. In this case, the maximum buffer size requirement for each MN is about 4.016 KB. This means that a memory of size 128 MB can handle over than 30 thousands of MNs.

VI. CONCLUSION

This paper described a new micro-mobility management scheme, called adaptive MRA, that supports both mobility and QoS resource provisioning in IP-based mobile networks. Our proposal uses the concept of residing areas in an MPLS access network. This area is constructed dynamically according to the current MN's position and the delay constraints. Using both analytical and simulation approaches, we compared our proposal with existing solutions (FMIP, Mobile MPLS, MIP-RR, M-MPLS and PF) using the 2-D random walk mobility model as well as real mobility patterns. To do so, we analytically derived the registration updates cost, the link usage and the handoff performance for all underlying protocols. Numerical and simulation results showed that our proposed scheme achieves substantial signaling cost and link usage reduction and improves the handoff performance, which are crucial for supporting real-time applications. In particular, we found that our scheme provides the lowest registration cost and handoff latency when the maximum tolerable delay inside a micro-mobility domain has moderate or small values. As such, our protocol stands out as the best choice for delay sensitive applications.

REFERENCES

[1] C. Perkins, IP Mobility Support for IPv4, RFC 3220, Jan. 2002.
[2] I. F. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *IEEE Wireless Commun. Mag.*, vol. 11, no. 4, pp. 16-28, Aug. 2004.
[3] A. Campbell, J. Gomez, S. Kim, Z. Turanyi, A. Valko, and C.-Y. Wan, "Design, implementation and evaluation of cellular IP," *IEEE Personal Commun. Mag.*, Aug. 2000.
[4] R. Ramjee, K. Varadhan, L. Salgarelli, *et al.*, "HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, pp. 396-410, June 2002.
[5] A. Misra, S. Das, A. Dutta, A. Mcauley, and S. K. Das, "IDMP-based fast handoffs and paging in IP-based 4G mobile networks," *IEEE Commun. Mag.*, vol. 40, pp. 138-145, Mar. 2002.
[6] R. Koodli, ed., "Fast handovers for mobile IPv6," RFC 4068, July 2005.
[7] J. Xie and I. F. Akyildiz, "A distributed dynamic regional location management scheme for mobile IP," *IEEE INFOCOM*, 2002.
[8] E. Gustafsson, *et al.*, "Mobile IPv4 Regional Registration," draft-ietf-mip4-reg-tunnel-01.txt, Nov. 2005, work in progress.
[9] Y. Bejerano and I. Cidon, "An anchor chain scheme for IP MOBILITY MANAGement," *ACM-Baltzer J. Wireless Networks*, vol. 9, no. 5, pp. 409-420, 2003.
[10] C. Chu and C. Weng, "Pointer forwarding MIPv6 mobility management," *IEEE GLOBECOM*, 2002.
[11] 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.
[12] E. Rosen, *et al.*, "Multiprotocol label switching architecture," Internet IETF RFC 3031, Jan. 2001.
[13] ITU-T Recommendation Y.1281: Mobile IP services over MPLS.
[14] Z. Ren, C. Tham, C. Foo, and C. Ko, "Integration of mobile IP and multi-protocol label switching," *IEEE ICC*, June 2001.
[15] F. M. Chiussi, D. A. Khotimsky, and S. Krishnan, "Mobility management in third-generation all-IP networks," *IEEE Commun. Mag.*, pp. 124-135, Sept. 2002.
[16] V. Vassiliou, H. L. Owen, D. Barlow, J. Sokol, H.-P. Huth, and J. Grimminger, "M-MPLS: micromobility-enabled multiprotocol label switching," in *Proc. IEEE ICC*, vol. 1, pp. 250-255, May 2003.
[17] K. Xie, V. Wong, and V. Leung, "Support of micro-mobility in MPLS-based wireless access networks," *IEICE Trans. Commun.*, vol. E88-B, no. 7, pp. 2735-2742, July 2005.
[18] R. Langar, S. Tohme, and N. Bouabdallah, "Mobility management support and performance analysis for wireless MPLS networks," *ACM Int. J. Network Management*, vol. 16, no. 4, pp. 279-294, July 2006.
[19] K. El Malki, "Low Latency Handoffs in Mobile IPv4," IETF, draft-ietf-mobileip-lowlatency-handoffs-v4-11.txt, Oct. 2005.
[20] D. Awduche, L. Berger, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: extensions to RSVP for LSP tunnels," Internet IETF RFC 3209, Dec. 2001.
[21] B. Jamoussi, L. Anderson, R. Callon, R. Dantu, *et al.*, "Constraint-based LSP setup using LDP," Internet IETF RFC 3212, Jan. 2002.
[22] K.-H. Chiang and N. Shenoy, "A 2-D random-walk mobility model for location-management studies in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 53, no. 2, pp. 413-424, Mar. 2004.
[23] D. N. Alparslan and K. Sohraby, "Two-dimensional modeling and analysis of generalized random mobility models for wireless ad hoc networks," *IEEE/ACM Trans. Networking*, vol. 15, no. 3, pp. 616-629, June 2007.
[24] E. Cinlar, *Introduction to Stochastic Processes*. Prentice Hall, 1975.
[25] ns2 home page. [Online]. Available: http://www.isi.edu/nsnam/ns
[26] S. PalChaudhuri, J.-Y. Le Boudec, and M. Vojnovic, "Perfect simulations for random trip mobility models," in *Proc. 38th Annual Simulation Symposium*, San Diego, CA, Apr. 2005.
[27] "A community resource for archiving wireless data at Dartmouth. [Online]. Available: http://crawdad.cs.dartmouth.edu
[28] S.-C. Lo, G. Lee, W.-T. Chen, and J.-C. Liu, "Architecture for mobility and QoS support in all-IP wireless networks," *IEEE J. Select. Areas Commun.*, vol. 22, pp. 691-705, May 2004.

Rami Langar received the B.S. degree in telecommunications engineering from the Ecole Superieure des Communications (Sup'Com), Tunisia, in 2001. He received the M.S. degree from the University of Pierre and Marie Curie - Paris 6, France, in 2002, and the Ph.D. degree from the Ecole Nationale Superieure des Telecommunications de Paris, France, in 2006, both in Network and Computer Science.

In 2007 and 2008, he was with the School of Computer Science at the University of Waterloo, Ontario, Canada as a Postdoctoral research fellow. He is currently an Associate Professor of Computer Science at the University of Pierre and Marie Curie, UPMC - Paris Universitas. His research interests include post-IP architecture, mobility management, resource management, wireless mesh and vehicular ad-hoc networks, performance evaluation as well as quality of service support.

Nizar Bouabdallah received the B.S. degree in telecommunications engineering from Sup'Com, Tunisia, in 2001, and the M.S. and Ph.D. degrees in network and computer science from the University of Pierre and Marie Curie - Paris 6, France, in 2002 and 2004, respectively. He joined Alcatel Research Laboratories, Marcoussis, France, in 2002, while working on his Ph.D. degree.

In 2005, he was with the North Carolina State University, Raleigh, as a Postdoctoral Fellow. He is currently a researcher at INRIA (Institut National de Recherche en Informatique et en Automatique). In 2007, he spent six months as a Visiting Researcher at the School of Computer Science, University of Waterloo, Waterloo, ON, Canada. His research interests include wireless mesh and sensor networks, optical networking, resource allocation under QoS, network planning and modeling, as well as performance evaluation.

Raouf Boutaba received the M.Sc. and Ph.D. degrees in computer science from the University of Pierre and Marie Curie - Paris 6, France, in 1990 and 1994 respectively. He is currently a Full Professor of computer science at the University of Waterloo, Waterloo, ON, Canada. His research interests include network, resource and service management in multimedia wired and wireless networks.

Dr. Boutaba is the founder and Editor-in-Chief of the electronic publication IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and is on the editorial boards of several other journals. He is currently a distinguished lecturer of the IEEE Communications Society, the chairman of the IEEE Technical Committee on Information Infrastructure and the IFIP Working Group 6.6 on Network and Distributed Systems Management. He has received several best paper awards and other recognitions including the Premier's research excellence award.

Bruno Sericola received the Ph.D. degree in computer science from the University of Rennes I in 1988. He has been with INRIA (Institut National de Recherche en Informatique et Automatique, a public research French laboratory) since 1989. His main research activity is in computer and communication systems performance evaluation, dependability and performability analysis of fault-tolerant architectures and applied stochastic processes.