A Comprehensive Analysis of Mobility Management in MPLS-Based Wireless Access Networks

Rami Langar, *Member, IEEE*, Nizar Bouabdallah, and Raouf Boutaba, *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 perspective, we propose in this paper a new mobility management scheme designed to track host mobility efficiently so as to minimize both handoff latency and signaling cost. Building on and enhancing Mobile IP and taking advantage of MPLS traffic engineering capability, three mechanisms (FH-, FC- and MFC-Micro Mobile MPLS) are introduced. In order to assess the efficiency of our proposals, all protocols are compared. To achieve this, we develop analytical models to evaluate the signaling cost and link usage for both two-dimensional and one-dimensional mobility models. Additional mathematical models are also provided to derive handoff latency and packet loss rate. Numerical and simulation results show that the proposed mechanisms can significantly reduce the registration updates cost and provide low handoff latency and packet loss rate under various scenarios.

Index Terms—Fast handoff, forwarding chain, micro-mobility, Mobile IP, mobility models, MultiProtocol Label Switching (MPLS), performance analysis, residing area.

I. INTRODUCTION

FUTURE wireless networks are expected to provide IP-based coverage and efficient mobility support with end-to-end quality-of-service (QoS) guarantees. Two enabling factors are considered as crucial: 1) maintaining the network connectivity during node mobility and 2) 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. However, it 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]–[12] to ensure service continuity.

The second major enabling factors identified as crucial for the evolution process of next-generation mobile networks is efficient network resources provisioning. This issue has been largely studied in both wired and wireless environments. For instance, MultiProtocol Label Switching (MPLS [13]) addresses today's network backbone requirements effectively by providing a standardized solution that improves packet forwarding performance and designs high performance QoS guaranteed paths. MPLS strength lies in its ability to accommodate different application requirements and to engineer efficiently traffic tunnels, avoiding thus congestion and enabling an efficient use of the available bandwidth. As a matter of fact, there is an increasing trend towards the introduction of MPLS in wireless environments [14]–[18].

To meet the requirement of next generation mobile networks, in this paper, we propose a new protocol called Micro Mobile MPLS which alleviates the limitations of Mobile IP and in the same time benefits from MPLS resource provisioning capability. We propose three mechanisms that enhance both mobility and QoS support capabilities for next-generation cellular networks. In the first one called FH-Micro Mobile MPLS, we consider the *fast handoff* (FH) mechanism, which anticipates the LSP procedure setup with an adjacent subnet that an MN is likely to visit. This mechanism is proposed to reduce service disruption by using the L2 functionalities. In the second one called FC-Micro Mobile MPLS, a simple *forwarding chain* (FC) concept with loop removal is introduced to track efficiently the host mobility within a domain. The forwarding chain can reduce registration updates cost (i.e., volume of signaling messages exchanged during handoff operations) and provide low handoff latency and QoS support thanks to MPLS capabilities. In the third mechanism called M(Master)FC-Micro Mobile MPLS, we define a new attribute, called *residing area*, for each MN to achieve further reduction of the signaling costs. This mechanism can be seen as an extension of the previous one in a two-dimensional mobility model. To gauge the effectiveness of our proposed mechanisms, we compare the registration updates cost, the link usage cost and the handoff performance metrics (i.e., handoff latency and packet loss rate) using both 2-D and 1-D mobility models. Simulation results show that our proposals can improve significantly the network performances when compared to existing schemes (FMIP [7], MIP-RR [9], Mobile MPLS [14], H-MPLS [16]) under various scenarios.

The rest of this paper is organized as follows. In Section II, we discuss related works and position our own. Section III presents our proposed mobility mechanisms. Section IV describes the system model used to evaluate the performance of our proposals. In Section V, we develop analytical models to derive the signaling cost function of registration updates, the link usage and the handoff performance for all underlying protocols using the 2-D and 1-D mobility models. Numerical and simulation results are given in Section VI. Section VII concludes the paper.

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

N. Bouabdallah is with INRIA/IRISA, 35042 Rennes, France (e-mail: nizar.bouabdallah@inria.fr).

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 (intra-domain mobility) environment, where handoff operations are much more frequent than in the macro-mobility (inter-domain mobility) case. We note that the Mobile IP protocol is, in essence, designed to deal with macro-mobility management rather than micro-mobility management. 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 for Mobile IP is proposed in [6], [7] using both link layer triggers and buffering mechanisms.

Specifically, authors in [6] propose to reduce the delay and the data loss during handoff in two ways: either by using the pre-registration or the post-registration handoff methods. In the former case, layer 3 (L3) handoff is initiated prior to layer 2 (L2) handoff occurrence thanks to L2 triggers. The L2 trigger is a signal from L2 to inform L3 of an imminent L2 handoff. Accordingly, the MN initiates the Mobile IP registration to its Home Agent (HA) through the old Foreign Agent (FA) before the L2 handoff occurrence. In the latter case (i.e., post-registration), the L3 handoff is initiated after the L2 handoff occurrence. Ref. [6] proposes also a systematic way to choose between the two registration methods. The old FA attempts first the pre-registration method before the L2 handoff occurrence, and if it fails, it attempts the post-registration method.

A second alternative to introduce a fast handoff scheme for Mobile IP is proposed in [7]. This scheme, called FMIP, 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 receive data as soon as its attachment is detected by the new access router. This approach has a significant effect on the performance of real-time and QoS sensitive applications. However, the use of a discovery protocol implies modifications to Mobile IP in order to support the discovery of neighboring FAs. 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, authors in [8] propose a distributed dynamic location management scheme. They assume that every FA can act as a Gateway Foreign Agent (GFA). This scheme can be seen as an extension of the IETF regional registration protocol (MIP-RR [9]) 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 [12]. 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 used in [10]–[12]. Accordingly, pointers are setup between the old and new subnets by the MN, reducing hence the local registration cost. However to achieve this, authors in [10] assume that there is a correlation between the com-

munication cost and the geographic distances. As such, the coordinates of each FA along the formed forwarding chain need to be known by the MN.

To achieve the pointer forwarding, authors in [11] and [12] use a caching method and a cascaded tunneling scheme to connect the MN to its HA. However, the authors limited their studies to the simple one dimensional random walk model when analyzing their protocols. Indeed, the use of such schemes in a two-dimensional case will result in the establishment of non-optimal forwarding chains with loop problems leading thus to unacceptable end-to-end delay. In addition, the QoS support has not been addressed.

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 [14]–[18] propose the use of MPLS in IP-based wireless access networks to benefit from its QoS, traffic engineering and reliability capabilities.

Specifically, [14] proposes a scheme to integrate Mobile IP and MPLS protocols. This scheme, called Mobile MPLS, 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 a scheme suffers 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], a signaling framework for intra-domain mobility in a traffic engineered network has been proposed. An enhanced LER called the label edge mobility agent (LEMA) is introduced to support chained 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.

The authors of [16]–[18] 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. However, these works have not taken into account the fact that the signaling delay for registration updates could be very long, which may cause service disruption for real-time services and will result in increasing the registration updates cost and QoS degradation. Indeed, 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 and long service delay. Moreover, 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.

In this paper, we propose a new mobility management scheme for MPLS-based wireless networks that combines the benefits of previous works on mobility management with MPLS capabilities. Our scheme integrates three mobility mechanisms. The first one, called FH, is a hybrid solution combining the pre-registration and post-registration methods with QoS provisioning
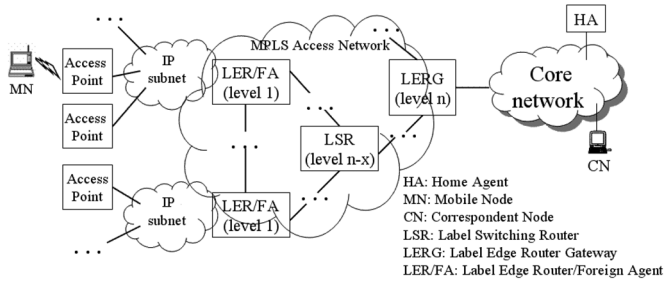
Fig. 1.   Architecture of a Micro Mobile MPLS wireless access network.

and aims at achieving further handoff optimization compared to [6] and [7]. The second mechanism, called FC, is a new implementation alternative of the pointer forwarding concept in an MPLS environment. It overcomes the limitations of [10] and [11] by avoiding the geographical constraints and loop issues, respectively. In addition, this mechanism reduces the resource reservation cost by allowing packets to be forwarded through existing LSPs. Finally, the third mechanism, called MFC, aims to optimize the forwarding chain length and thus the end-to-end delay when MNs move in a two-dimensional space. To the best of our knowledge, this is the first micro-mobility protocol that takes into account the specific mobility propriety of users in a 2-D space.

III. MICRO MOBILE MPLS

In this section, we describe our proposed Micro Mobile MPLS scheme. As mentioned earlier, our ultimate aim is to overcome the limitations of Mobile IP and in the same time benefit from the QoS support capability of MPLS networks. Micro Mobile MPLS is based on the integration of MIP-RR [9] and MPLS [13] protocols. A typical architecture for Micro Mobile MPLS networks [19] is shown in Fig. 1. We assume that an MPLS access network exists between the Label Edge Router Gateway (LERG) and the Label Edge Router/Foreign Agents (LER/FAs). The architecture is based on a two-level hierarchy. At the higher level is the LERG that performs the role of an edge node 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 connectivity. We distinguish here between link-layer 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. In the following, we present the operation of our proposed mechanisms, namely, FH-, FC-, and MFC-Micro Mobile MPLS.

A. The Fast Handoff Mechanism: FH-Micro Mobile MPLS

The main idea behind FH-Micro Mobile MPLS is to anticipate the L3 handoff using the L2 functionalities and to setup an LSP before the MN moves really into a new subnet to reduce service disruption. In this context, our proposed FH mechanism can be considered as a hybrid solution combining the pre-registration and post-registration methods. Indeed, FH relies on two main operations to accomplish the handoff process: 1) the pre-establishment of a new LSP between the LERG and the new subnet prior to the L2 handoff occurrence (i.e., pre-registration) subject to the desired QoS requirements. As such, the
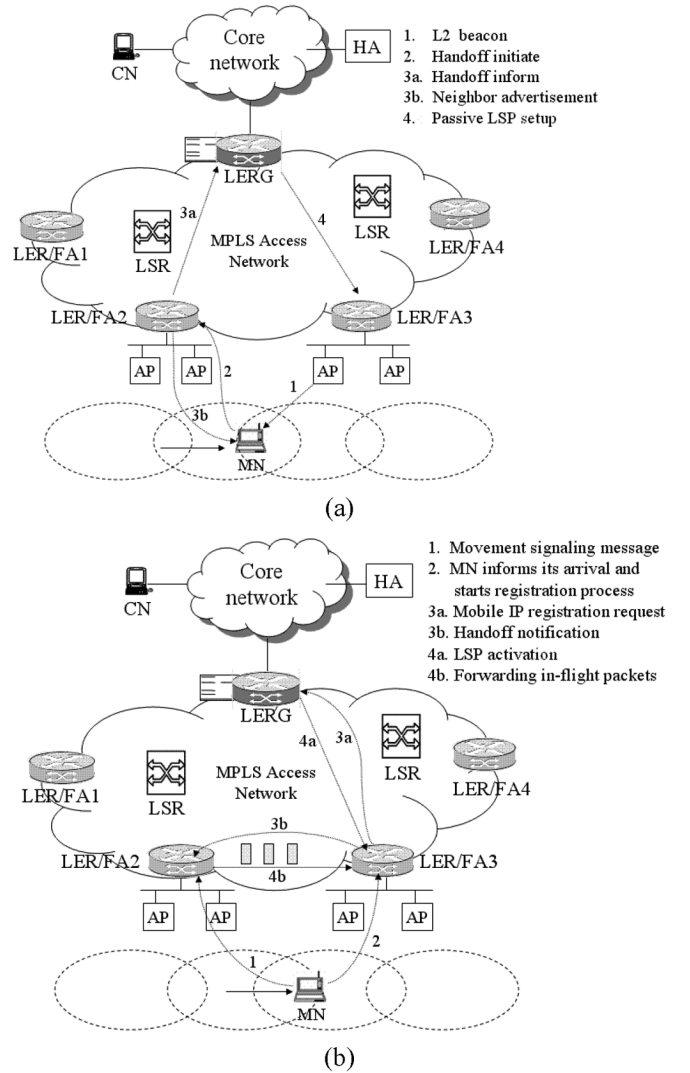

(a)

(b)

Fig. 2.   Operation of the FH-Micro Mobile MPLS scheme in the case of downlink communications. (a) Before handoff: pre-registration step. (b) During handoff: post-registration step.

QoS continuity of the communication in progress is ensured and 2) the initiation of the L3 handoff immediately after the L2 handoff completion through the new FA even before receiving the new Mobile IP advertisement message from the new subnet (i.e., post-registration).

In what follows, we describe the basic operations of our FH mechanism. We consider two types of LSP: active LSP and passive LSP. The active LSP is the one from the LERG to the current serving LER/FA. This LSP is currently used to transfer data. On the other side, a passive LSP is the one from the LERG to the neighboring subnet to which the MN is moving to. This LSP is not currently used until its activation.

The FH mechanism employs a link-layer movement detection to predict the possible MN's next location. In other words, as can be seen in Fig. 2(a), once the MN enters an overlapped area of the boundary cells of two subnets, it receives a L2 beacon from the possible new AP (step 1). Immediately, the MN notifies the current FA for possible handoff by sending a *Handoff Initiate* (HI) signaling message, which contains the MAC address of the new AP (step 2). Note that in practice, the MN is able to sense a set of candidate APs. In this case, it is assumed that the MN

checks continuously the signal strength of these APs and then selects the one with the strongest beacon signal.

Each LER/FA has a *Neighbor Mapping Table* (NMT) that binds between the IP and MAC addresses of all neighboring APs. Hence, when the current FA receives the HI signaling message, it looks into its NMT table to get the new FA's IP address and then informs the LERG (respectively, the new FA) in the case of downlink communications (respectively uplink communications) for the possible L3 handoff (step 3a). A passive LSP with the desired QoS requirements will be then established between the LERG and the new subnet (i.e., pre-registration) using either the RSVP-TE [20] or CR-LDP [21] protocols (step 4). At the same time as step 3a, the current FA informs the MN of the new Regional Care-of-Address (RCoA: the IP address of the new possible LER/FA) by sending a *Neighbor Advertisement* signaling message (step 3b). This allows the MN to be able to start registration process with the LERG before receiving the new Mobile IP advertisement message (L3 beacon) from the new FA. Specifically, the MN's registration to the LERG will be initiated immediately after the L2 handoff completion through the new FA (i.e., post-registration). Indeed, the L2 handoff is launched by the MN when the received signal strength from the current AP falls below the threshold level. In this case, the MN notifies the current FA to initiate the buffering mechanism by sending a Movement signaling message (see Fig. 2(b), step 1). Once the L2 handoff is accomplished (i.e., the MN is attached to the new AP), the L3 handoff is initiated by the MN since it is already aware of the new RCoA (step 2). The new FA forwards the Mobile IP registration request message to the LERG (step 3a) and notifies the old FA for the handoff event (step 3b). As soon as the LERG receives the Mobile IP registration request, it activates the pre-established passive LSP and traffic will be delivered through the activated LSP (step 4a). On the other hand, once the old FA is notified (step 3b), in-flight packets are forwarded to the MN through the new FA along a static MPLS tunnel (step 4b). By using the fast handoff mechanism, we can improve the handoff performance of Micro Mobile MPLS and reduce service disruption.

B. Forwarding Chain Mechanism: FC-Micro Mobile MPLS

This approach is based on the forwarding chain concept (i.e., set of forwarding paths) and can be considered as a new alternative to implement the pointer forwarding technique in an MPLS environment. Indeed, each time that the MN moves to a new subnet, it registers with the old FA instead of the LERG, as shown in Fig. 3. Consequently, the old FA will send a label request message to the new FA, which responds with a label mapping message back to the old FA. This way, the existing LSP (with QoS requirements) between the LERG and the old subnet will be extended to the new one (see Fig. 4). As a result, a forwarding chain of FAs will be constructed for each MN. To do so, each MN keeps a buffer storing IP addresses of the visited FAs. Packets traveling towards this MN will be intercepted by the first FA in the chain (called master FA), taking advantage of the existing LSP between the LERG and the master LER/FA, and then forwarded along the chain of FAs to the MN. It is easy to see that such a scheme may cause unacceptable delays due to long chains. To avoid long forwarding chains, we set a threshold on its length denoted by L_{th}. When the threshold is reached, the
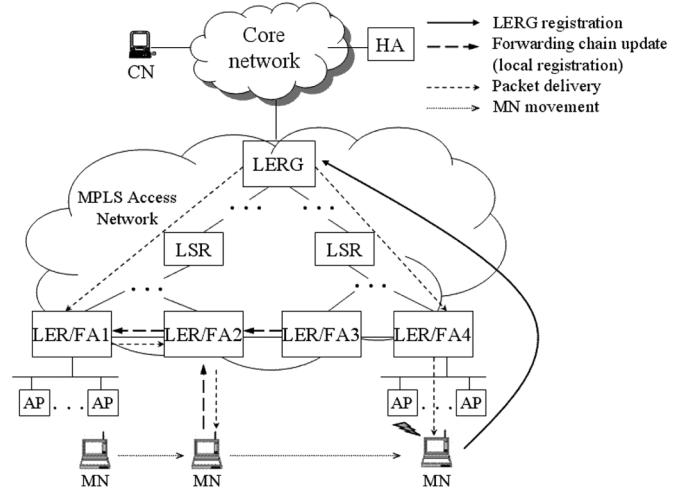

Fig. 3. Operation of the FC-Micro Mobile MPLS scheme.

(a)

In. port	In. label	FEC	Out. port	Out. label
1	20	fec2	-	-
1	25	fec1	-	-
...	...	...	...	...

(b)

In. port	In. label	FEC	Out. port	Out. label
1	20	fec2	-	-
1	25	fec1	2	30
...	...	...	...	...

(c)

In. port	In. label	FEC	Out. port	Out. label
2	30	fec1	-	-
...	...	...	...	...

Fig. 4. Label tables of FA nodes when a MN moves between the LER/FA1 and the LER/FA2, where fec1 corresponds to the five TCP/IP header fields: MN's IP address, MN's port number, protocol number, correspondent node's IP address and correspondent node's port number. (a) Label table of LER/FA1 before handoff. (b) Label table of LER/FA1 after handoff. (c) Label table of LER/FA2.

MN will register to the LERG and delete all the addresses in its buffer. That is, a new LSP between the LERG and the new subnet will be established and the new visited FA becomes the new master FA. This scheme enables a significant reduction of both the registration updates and LSP setup signaling messages sent by the MN to the LERG. As a result, the resource reservation cost will be reduced since packets can be forwarded through the existing LSPs benefiting from existing resource reservation on the old path.

The basic operation of the FC scheme is depicted in Fig. 3, where the MN moves from subnet 1 to subnet 4. We assume that the threshold on the forwarding chain length is equal to three. When the MN moves to subnet 2, it checks first the presence of the new subnet's IP address in its cache memory. Since it is the first time that the MN visits this subnet, it registers with the previous LER/FA1, which is the master FA. Thus, the existing LSP between the LERG and the LER/FA1 nodes will be extended to the LER/FA2 node, as shown in the tables of Fig. 4(a)-(c). Likewise, when the MN moves to subnet 3, it notifies the previous LER/FA2. The threshold of the forwarding chain length (L_{th}) is reached when the MN enters subnet 4. At this time, the MN will

```
%Location registration procedures
Initialize i = 0;
IF (MN enters a new subnet)
     compare the address of the new LER/FA to the addresses in buffer;
     IF (the new address already exists in buffer)
          Loop detected: Extract the rank (rg) of the subnet from buffer;
          i = rg;
          MN deletes the subsequent FA addresses in buffer;
          New LER/FA updates its label table;
     ELSE
          record the new LER/FA address in buffer;
          i = i + 1;
          IF (i < L_th)
               MN registers to the old LER/FA: Old LSP extended;
          ELSE
               MN registers to the LERG: New LSP creation;
               MN notifies the old LER/FA about the handoff;
               MN deletes all the addresses in buffer;
               i = 0;
          ENDIF
     ENDIF
ENDIF
```

Fig. 5. FC-Micro Mobile MPLS scheme.

Fig. 6.   Movement path of an MN with $K = 4$ and $L_{\mathrm{th}} = 7$.

register to the LERG and updates its address to the root of the domain directly. At the same time, the new LER/FA4 becomes the master FA of the next forwarding chain.

Note that according to our FC mechanism, the loops possibly formed during the MN's movement will be systematically removed. Indeed, each time the MN moves to a new subnet, it checks first the presence of the new FA's IP address in its cache memory. If the FA is already in the MN's buffer, the MN deletes the generated loop. In this case, the revisited FA updates its label table by associating null values to both the outgoing port and outgoing label of MN's packets, which means that the current FA becomes the egress node of the new LSP.

The FC mechanism avoids also the geographical constraints observed in [10] and alleviates the scalability issue of [12] since only the MN needs to keep a cache of all its visited subnets (i.e., containing the visited FAs' IP addresses and not FAs' coordinates as required in [10]). As such, there is no need to communicate this information to each visited FA as in [12]. The FC scheme is described by the pseudocode in Fig. 5.

C. Master Forwarding Chain Mechanism: MFC-Micro Mobile MPLS

The third mechanism that we propose to handle efficiently local mobility is called Master Forwarding Chain (MFC)-Micro Mobile MPLS and can be seen as an extension of the FC scheme in a two-dimensional space [22].

In this mechanism, the MN registers to the LERG only when its distance, in terms of number of hops (instead of movements) from the master FA, reaches a threshold K. To illustrate the difference between FC and MFC mechanisms, we consider the simple example presented in Fig. 6. Assume that the MN's trajectory is $\{S_1, S_2, S_3, S_4, S_5, S_6\}$. Note that S_1 is the master FA. According to FC, the forwarding chain is $\{S_1, S_2, S_3, S_4, S_5, S_6\}$. Whereas, in the MFC case, the resulting forwarding chain is $\{S_1, S_3, S_6\}$, which is the shortest path between the master FA and the MN. It is easy to see that the MFC scheme reduces the forwarding chain size. In other words, considering the same threshold (i.e., $K = L_{\mathrm{th}}$), the registration updates with the LERG are more frequent in the
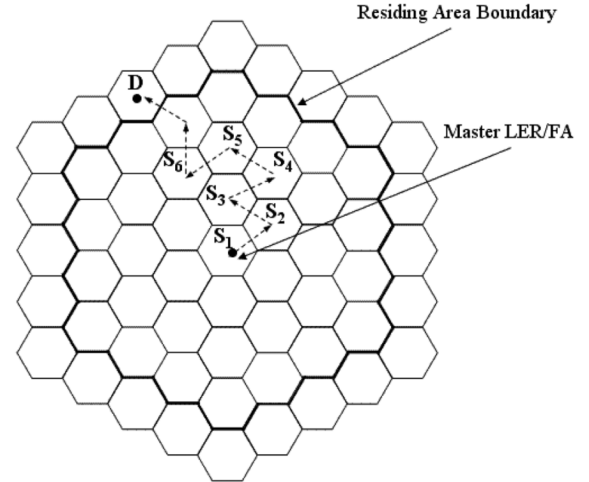
FC case. In addition, the end-to-end delay is reduced with MFC thanks to the relatively shorter forwarding chain.

The second main difference between the two mechanisms is related to the local registration. Considering the MFC scheme, each time that the MN moves to a new subnet, the new RCoA will be registered at the master FA instead of the old FA, as long as the threshold is not reached. Henceforth, this kind of registration will be called local registration as opposed to the LERG registration. Specifically, the local registration is always done through only one hop in the FC case, whereas it can be done through up to $K - 1$ hops in the MFC case.

According to the MFC mechanism, as long as the MN remains in a *residing area* around the master FA, regardless of the taken trajectory, it carries out a local registration (see Fig. 6). Once it leaves this area, it performs a LERG registration and the new serving FA becomes the new MN's master FA. In the example of Fig. 6, the threshold K is set to 4. This figure shows the MN's trajectory. The LERG registration will be performed by the MN at location D since its distance to the master FA reaches the threshold K. Note that, unlike the MFC case, the LERG registration in FC depends mainly on the MN's trajectory.

The basic operations of the MFC mechanism are listed below. When the MN enters for the first time into a domain, it performs the same operations as in FH and FC mechanisms. When the MN moves to a new subnet, it proceeds as follows.

1) The MN sends a registration message to the new LER/FA. This message contains the IP address of the associated master FA.
2) The new LER/FA checks the existence of the master FA's address. If it exists, the new FA computes the shortest distance to the master FA. According to this distance, it performs either a local registration or a LERG registration. Otherwise (if the address does not exist), it performs a LERG registration.
3) Finally, the MN receives a registration reply message either from the LERG or from the master FA, according to the registration type.

It is worth noting that with our proposed MFC mechanism, loops can not be formed during the MN's movement since the
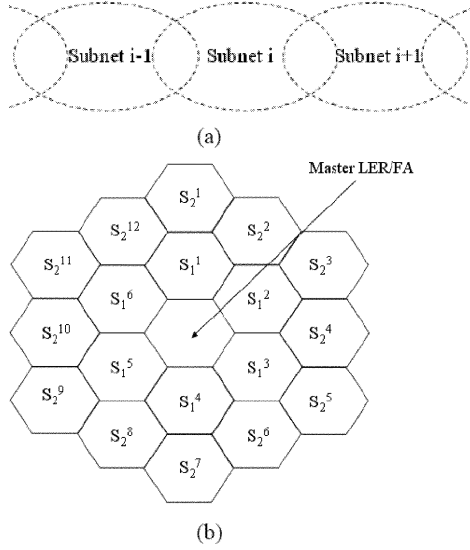
(a)

(b)

Fig. 7. 1-D and 2-D mobility models. (a) Linear model. (b) 2-D model.

shortest path between the Master FA and the new FA nodes is selected each time the MN moves to a new subnet.

IV. SYSTEM MODEL

In this section, we study the MN's mobility. Our aim is to determine the position of an MN with respect to its master FA in order to be able to predict the MN's evolution. To do so, we develop Markovian models. The obtained results will be used, in a later stage, to derive the protocol performance metrics such as the registration updates cost and the link usage.

The wireless network is divided into subnetworks. Each subnet is covered by one LER/FA, called base station in cellular networks. Moreover, each LER/FA manages several access points (APs), which offer link-layer connectivity. The mobility model commonly used when planning the LER/FAs deployment is either the 2-D or the 1-D mobility models. The first (2-D) model fits those applications, where the MN can travel in various directions within the covered region, such as urban areas. The second model (1-D or linear mobility model) is more adapted for those applications where the MN's mobility is limited to a pre-specified unidirectional trajectory. This model is commonly used in access applications along roads, tunnels and trains railways. In this work, we consider both mobility models.

A. 1-D Mobility Model

In this model, each LER/FA has two neighbors. As shown in Fig. 7(a), an MN which is in subnet i, can only move to two neighboring subnets $i+1$ or $i-1$ with equal probability $p(p = 1/2)$; otherwise the connectivity is lost. Note that, considering the linear model, the MN's behavior is the same according to the MFC or FC schemes (when $K = L_{\text{th}}$).

Let $X(t)$ be the distance (in terms of number of hops) between the MN's location at time t and the master FA. The residence time of the MN in each subnet is assumed to be exponentially distributed with the mean $1/\mu$. $\{X(t), t \geq 0\}$ forms a Continuous-Time Markov Chain (CTMC) with state space $E1 = 0, 1, 2, \ldots, (\phi - 1)$ as depicted in Fig. 8, where ϕ is L_{th} in the FC case and K in the MFC case.
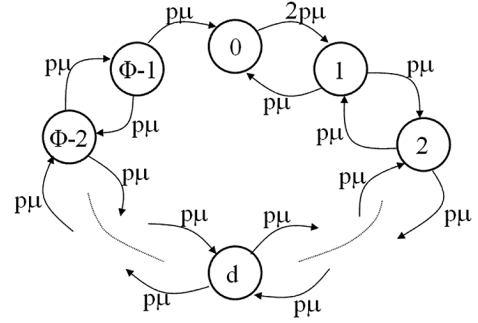

Fig. 8. CTMC for the linear model.

Let $\Pi_d = \lim_{t \to \infty} Prob[X(t) = d]$, $d \in E1$, be the stationary probability distribution of $X(t)$. Based on Fig. 8, the balance equations can be derived as follows:

$$\begin{cases} 2p\mu\Pi_0 = p\mu\Pi_1 + p\mu\Pi_{\phi-1} \\ 2p\mu\Pi_1 = 2p\mu\Pi_0 + p\mu\Pi_2 \\ 2p\mu\Pi_d = p\mu\Pi_{d-1} + p\mu\Pi_{d+1} \quad \forall 2 \leq d \leq (\phi - 2) \\ 2p\mu\Pi_{\phi-1} = p\mu\Pi_{\phi-2} \\ \sum_{d=0}^{\phi-1} \Pi_d = 1. \end{cases} \quad (1)$$

Solving these equations, we obtain

$$\begin{cases} \Pi_0 = \frac{1}{\phi} \\ \Pi_d = \frac{2(\phi-d)}{\phi^2} \quad \forall 1 \leq d \leq (\phi - 1). \end{cases} \quad (2)$$

B. 2-D Mobility Model

In our study, we consider a polygon-based 2-D model. Typically, each LER/FA covers an hexagon area. This model is broadly used in the literature. In this case, each subnet is surrounded by six neighbors [see Fig. 7(b)]. The MN can move to one of the neighboring subnets with equal probability $p(p = 1/6)$. For simplicity, we only present in this section the Markovian chain corresponding to the mobility behavior of an MN in the MFC case. Hence, in the remainder of this section we only consider the MFC case.

Fig. 7(b) represents the *residing area* of an MN for the case $K = 3$. This area contains the master FA subnet surrounded by $2 = K - 1$ 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 current master FA. 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. To generalize, let $i = 0, 1, \ldots, (K-1)$ designate the ith ring away from the master FA. The master FA subnet is denoted by S_0^0. Subnets belonging to ring i are referenced by S_i^j, $1 \leq j \leq 6i$. Note that the ring label represents the distance between the MN and the master FA.

Let $X(t)$ be the MN's location within the residing area at time t. The residence time of an MN in each subnet S_i^j is assumed to be exponentially distributed with the mean $1/\mu$. $\{X(t), t \geq 0\}$ is, therefore, a Markov process with continuous time and finite state space $E2 = \{S_i^j | 0 \leq i \leq (K-1), 1 \leq j \leq 6i\}$. Recall that our main objective is to determine the MN's position within the *residing area* in order to predict its evolution. According to its next location, the MN can perform either a local registration

or a LERG registration. In the latter case, the master FA will be updated and its associated *residing area* will be created.

The resolution of the Markovian chain, as defined above, is time-consuming. Moreover, this chain suffers from the state-space explosion problem, mainly when the threshold K takes high values. To avoid this issue, we extract a new chain $Y(t)$ from $X(t)$ by aggregating its states. In other words, all the sates where the MN exhibits exactly the same behavior will be aggregated. Hence, the size of the state space $E2$ will be drastically reduced. To achieve this, we profit from the symmetric property of the 2-D model. Let us revisit the example of Fig. 7(b). The six states of the first ring are symmetric. Wherever the MN's location is in the first ring, it exhibits the same behavior. It moves to the master FA subnet with probability p, remains in the same ring with probability $2p$, and goes to ring 2 with probability $3p$. In this case, all these states can be grouped into one aggregate state. In the second ring, we have to segregate between two cases. The first aggregated state contains the following subnets $\{S_2^1, S_2^3, S_2^5, S_2^7, S_2^9, S_2^{11}\}$, whereas the second case comprises $\{S_2^2, S_2^4, S_2^6, S_2^8, S_2^{10}, S_2^{12}\}$ subnets. Indeed, the MN exhibits different behavior according to each aggregated state. For instance, in the first case, the MN leaves the residing area with probability $3p$ instead of $2p$ in the second case. In what follows, we describe the algorithm to perform state aggregation.

1) As before, S_i^j denotes the MN's position in the *residing area*. As presented in Fig. 7(b), the state S_i^1 is chosen to be the one at the top of state S_{i-1}^1. Afterwards, each ring i consists of $6i$ subnets labeled in a clockwise direction as $S_i^1, \ldots, S_i^{6i}$. Let S_i^{j*} denote the new aggregated state, where i always designates the ring reference, and $j*$ the state label inside the ring.

2) Start with i = 1;
   until $(i = K - 1)$
   Repeat {
    set $S_i^{0*} = S_i = \bigcup_{0 \le n \le 5} S_i^{n \times i + 1}$
    If $(i \ne 1)$ {
        set $L = \lceil (i-1)/2 \rceil$;
        For $m = 1$ to $m = L$
            set $S_i^{m*} = \bigcup_{0 \le n \le 5} S_i^{n i + m + 1}$
            $+ \bigcup_{1 \le n \le 6} S_i^{n i - m + 1}$
    }
    $i = i + 1$;
   }

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

For instance, for $K = 5$, we obtain the following aggregated states.

$S_0^{0*} = S_0 = \{S_0^0\}$,
$S_1^{0*} = S_1 = \{S_1^1, S_1^2, S_1^3, S_1^4, S_1^5, S_1^6\}$,
$S_2^{0*} = S_2 = \{S_2^1, S_2^3, S_2^5, S_2^7, S_2^9, S_2^{11}\}$,
$S_2^{1*} = \{S_2^2, S_2^4, S_2^6, S_2^8, S_2^{10}, S_2^{12}\}$,
$S_3^{0*} = S_3 = \{S_3^1, S_3^4, S_3^7, S_3^{10}, S_3^{13}, S_3^{16}\}$,
$S_3^{1*} = \{S_3^2, S_3^3, S_3^5, S_3^6, S_3^8, S_3^9, S_3^{11}, S_3^{12}, S_3^{14}, S_3^{15}, S_3^{17}, S_3^{18}\}$,
$S_4^{0*} = S_4 = \{S_4^1, S_4^5, S_4^9, S_4^{13}, S_4^{17}, S_4^{21}\}$,
$S_4^{1*} = \{S_4^2, S_4^4, S_4^6, S_4^8, S_4^{10}, S_4^{12}, S_4^{14}, S_4^{16}, S_4^{18}, S_4^{20},$
        $S_4^{22}, S_4^{24}\}$,
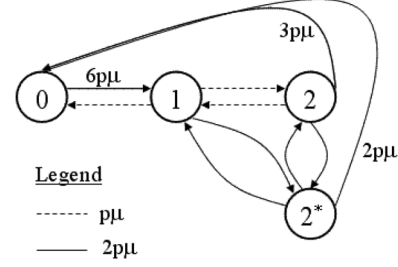$S_4^{2*} = \{S_4^3, S_4^7, S_4^{11}, S_4^{15}, S_4^{19}, S_4^{23}\}$

Fig. 9. State transition diagram of the aggregated Markov chain for $K = 3$.

Theorem:

Let $F = \{S_0, S_1, S_i, S_i^{1*}, \ldots, S_i^{m*}, \ldots, S_i^{L*}, S_{i+1}, \ldots\}$ $(i = 2, 3, \ldots, K-1)$ designate the state space of the new chain $Y(t)$ obtained by aggregation of the initial Markovian chain $X(t)$. The resulting aggregated process $Y(t)$ is also Markovian.

Proof:

For convenience, we denote by F_i each state of the set F and by $M = K + \sum_{i=2}^{K-1} \lceil (i-1)/2 \rceil$ the set size. Let Q_{init} designate the generator matrix of the initial Markov chain $X(t)$. We arrange the states of $E2$ according to the space F partitions (i.e.,

$$E2 = \{ \underbrace{S_0^0}_{F_1}, \underbrace{S_1^1, S_1^2, S_1^3, S_1^4, S_1^5, S_1^6}_{F_2}, \underbrace{S_2^1, S_2^3, S_2^5, S_2^7, S_2^9, S_2^{11}}_{F_3},$$
$$\underbrace{S_2^2, S_2^4, S_2^6, S_2^8, S_2^{10}, S_2^{12}}_{F_4}, \ldots\}).$$

In this case, the infinitesimal matrix Q_{init} can be written as a $N \times N$ symmetric matrix $(N = 1 + \sum_{i=1}^{K-1} 6i)$ and has the following form:

$$Q_{init} = (B_{mn})_{1 \le m, n \le M}$$

where B_{mn} is the block matrix corresponding to the transition probabilities between each element of the set F_m and the set F_n. In addition, these blocks verify the constant-row sum property [23]. In other words, we have

$$\forall i, \sum_j (B_{mn})_{ij} \text{ is a constant, denoted by } c_{mn}(c_{mn} \ge 0).$$

Thus, according to [23], the resulting aggregated process is Markovian. To illustrate this result, let us revisit the example of Fig. 7(b) where $K = 3$. As explained before, and according to the state aggregation algorithm, the state space $E2$ of the initial Markov chain $X(t)$ can be arranged with respect to the partition F. The generator matrix Q_{init} of the process $X(t)$ can be written as in (3), shown at the bottom of the next page, where $a = p\mu$ and $b = -6p\mu$. It is easy to see that each block matrix justifies the constant-row sum property. The new generator matrix Q_{aggr} of the aggregated Markov chain $Y(t)$ is, therefore, given by

$$Q_{aggr} = \begin{bmatrix} S_o & S_1 & S_2 & S_2^{1*} \\ b & 6a & 0 & 0 \\ a & (b+2a) & a & 2a \\ 3a & a & b & 2a \\ 2a & 2a & 2a & b \end{bmatrix} \quad (4)$$

In Fig. 9, we represent the state transition diagram of this aggregated process (i.e., for $K = 3$). Fig. 10 represents the case
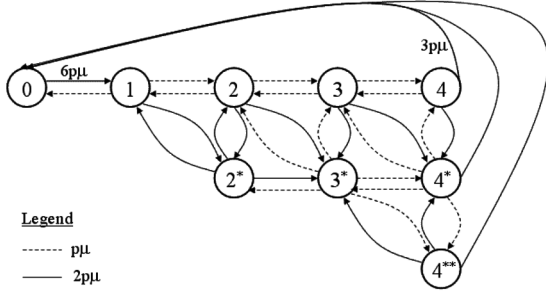
Fig. 10. State transition diagram of the aggregated Markov chain for $K = 5$.

where $K = 5$. For ease of use, the aggregate states were assigned numbers, as follows: $S_0 \to 0, S_1 \to 1, S_2 \to 2, S_2^{1*} \to 2^*, S_3 \to 3, S_3^{1*} \to 3^*, S_4 \to 4, S_4^{1*} \to 4^*, S_4^{2*} \to 4^{**}$.

Steady-State Probabilities:

Based on the state transition diagram of the aggregated Markov chain, we can obtain the steady-state probability for state F_i, $(i = 1, \ldots, M)$. Denote by Π_i and $\Pi_i^{(m)}$ $(i = (0, 1, \ldots, K - 1)$ and $m = (1, \ldots, L))$ the stationary probability of the system for the aggregated state S_i and S_i^{m*}, respectively. The balance equations for the aggregated Markov chain are obtained recursively as follows:

$$
\begin{cases}
\Pi_0 = p\Pi_1 + 3p\Pi_{K-1} + 2p \sum_{j=1}^{\lceil \frac{K-2}{2} \rceil} \Pi_{K-1}^{(j)} \\
\Pi_1 = 6p\Pi_0 + 2p\Pi_1 + p\Pi_2 + 2p\Pi_2^{(1)} \\
\Pi_2 = p\Pi_1 + p\Pi_3 + 2p\Pi_2^{(1)} + p\Pi_3^{(1)} \\
\Pi_{K-1} = p\Pi_{K-2} + p\Pi_{K-1}^{(1)} \\
\forall \quad 3 \le i \le K - 2 \\
\Pi_i = p\Pi_{i-1} + p\Pi_{i+1} + p\Pi_i^{(1)} + p\Pi_{i+1}^{(1)}
\end{cases}
\tag{5}
$$

$$
\begin{cases}
\Pi_2^{(1)} = 2p\Pi_2 + 2p\Pi_1 + p\Pi_3^{(1)} \\
\Pi_3^{(1)} = 2p\Pi_3 + 2p\Pi_2 + 2p\Pi_2^{(1)} + p\Pi_3^{(1)} \\
\quad + p\Pi_4^{(1)} + 2p\Pi_4^{(2)} \\
\Pi_4^1 = 2p\Pi_4 + 2p\Pi_3 + 2p\Pi_4^2 + p\Pi_3^1 + p\Pi_5^1 + p\Pi_5^2 \\
\forall \quad 5 \le i \le K - 1 \\
\Pi_i^{(1)} = 2p\Pi_i + 2p\Pi_{i-1} + p\Pi_i^{(2)} + p\Pi_{i-1}^{(1)} \\
\quad + \alpha p\Pi_{i+1}^{(1)} + \alpha p\Pi_{i+1}^{(2)}
\end{cases}
\tag{6}
$$

$$
\begin{cases}
\forall \quad 6 \le i \le K - 1 \text{ and } 2 \le j \le \lceil \frac{i-1}{2} \rceil - 1 \\
\Pi_i^{(j)} = p\Pi_i^{(j-1)} + \beta_1 p\Pi_i^{(j+1)} + p\Pi_{i-1}^{(j-1)} + p\Pi_{i-1}^{(j)} \\
\quad + \beta_2 p\Pi_{i+1}^{(j)} + \beta_2 p\Pi_{i+1}^{(j+1)}
\end{cases}
\tag{7}
$$

$$
\begin{cases}
\forall \quad 2 \le n \le \frac{K-1}{2} \\
\Pi_{2n}^{(n)} = p\Pi_{2n}^{(n-1)} + p\Pi_{2n-1}^{(n-1)} + \gamma_1 p\Pi_{2n+1}^{(n)}
\end{cases}
\tag{8}
$$

$$
\begin{cases}
\forall \quad 2 \le n \le \frac{K-2}{2} \\
\Pi_{2n+1}^{(n)} = p\Pi_{2n+1}^{(n)} + p\Pi_{2n+1}^{(n)} + p\Pi_{2n}^{(n-1)} \\
\quad + 2p\Pi_{2n}^{(n)} + \gamma_2 p\Pi_{2n+2}^{(n)} + \gamma_2 2p\Pi_{2n+2}^{(n+1)}
\end{cases}
\tag{9}
$$

where

$$
\alpha = \begin{cases} 1 & \text{if } 5 \le i \le K - 2 \\ 0 & \text{if } i = K - 1 \end{cases}
$$

$$
\beta_1 = \begin{cases} 1, & \text{if } i \text{ is odd} \\ 1, & \text{if } i \text{ is even and } 2 \le j \le (\lceil (i-1)/2 \rceil - 2) \\ 2, & \text{if } i \text{ is even and } j = (\lceil (i-1)/2 \rceil - 1) \end{cases}
$$

$$
\beta_2 = \begin{cases} 1, & \text{if } 6 \le i \le K - 2 \\ 0, & \text{if } i = K - 1 \end{cases}
$$

$$
\gamma_1 = \begin{cases} 0, & \text{if } n = (K - 1)/2 \\ 1, & \text{otherwise} \end{cases}
$$

$$
Q_{init} =
\begin{bmatrix}
\begin{array}{c|cccccc|cccccc|cccccc}
S_0^0 & S_1^1 & S_1^2 & S_1^3 & S_1^4 & S_1^5 & S_1^6 & S_2^1 & S_2^3 & S_2^5 & S_2^7 & S_2^9 & S_2^{11} & S_2^2 & S_2^4 & S_2^6 & S_2^8 & S_2^{10} & S_2^{12} \\
b & a & a & a & a & a & a & & & & & & & & & & & & \\
\hline
a & b & a & & & & a & a & & & & & & a & & & & & a \\
a & a & b & a & & & & & a & & & & & a & a & & & & \\
a & & a & b & a & & & & & a & & & & a & a & & & & \\
a & & & a & b & a & & & & & a & & & & a & a & & & \\
a & & & & a & b & a & & & & & a & & & & a & a & & \\
a & a & & & & a & b & & & & & & a & & & & a & a & \\
\hline
3a & a & & & & & & b & & & & & & a & & & & & a \\
3a & & a & & & & & & b & & & & & a & a & & & & \\
3a & & & a & & & & & & b & & & & a & a & & & & \\
3a & & & & a & & & & & & b & & & & a & a & & & \\
3a & & & & & a & & & & & & b & & & & a & a & & \\
3a & & & & & & a & & & & & & b & & & & a & a & \\
\hline
2a & a & a & & & & & a & a & & & & & b & & & & & \\
2a & & a & a & & & & & a & a & & & & & b & & & & \\
2a & & & a & a & & & & & a & a & & & & & b & & & \\
2a & & & & a & a & & & & & a & a & & & & & b & & \\
2a & & & & & a & a & & & & & a & a & & & & & b & \\
2a & a & & & & & a & a & & & & & a & & & & & & b \\
\end{array}
\end{bmatrix}
\tag{3}
$$

Legend

------- $p\mu$

——— $2p\mu$

$$\gamma_2 = \begin{cases} 0, & \text{if } n = (K-2)/2 \\ 1, & \text{otherwise} \end{cases}$$

and

$$\sum_{i=0}^{K-1} \Pi_i + \sum_{i=2}^{K-1} \sum_{m=1}^{\left\lceil \frac{i-1}{2} \right\rceil} \Pi_i^{(m)} = 1. \qquad (10)$$

Given the balance equations ((5)–(9)) and the normalization equation (10), the steady state probabilities of the aggregated Markov chain can be derived. Note that obtained results will be used in the next section to derive the signaling cost function of registration updates, the link usage and the handoff performance.

V. PERFORMANCE EVALUATION AND ANALYSIS

In this section, we develop analytical models to derive the link usage cost, the handoff performance and the cost function of registration updates for both the 2-D and the linear mobility models. This analysis includes our three mobility mechanisms (FH, FC and MFC) and competing schemes (FMIP [7], MIP-RR [9], Mobile MPLS [14] and H-MPLS [16]). The following parameters are used in our analysis.

Parameters:

t_s average session connection time;
t_r average FA resident time;
T_{ad} time interval for an 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;
s_u average size of a signaling message for the registration update;
s_l average size of a label message for LSP setup;
h_{x-y} average number of hops between x and y in the wired network;
C_{fh} location update cost between an FA and the HA (hop × message size);
C_{fg} location update cost between an LER/FA and the LERG (hop × message size);
C_{ff} location update cost between two neighboring LER/FAs (hop × message size);
l_{fh} traffic load related to LSP setup procedure between an FA and the HA (hop × message size);
l_{fg} traffic load related to LSP setup procedure between an LER/FA and the LERG (hop × message size);
l_{ff} traffic load related to LSP setup procedure between two neighboring LER/FAs (hop × message size).

The C_i and l_i parameters can be written as

$$\begin{cases} C_{fh} = 2s_u h_{FA-HA}; & l_{fh} = 2s_l h_{FA-HA} \\ C_{fg} = 2s_u h_{FA-LERG}; & l_{fg} = 2s_l h_{FA-LERG} \\ C_{ff} = 2s_u h_{FA-FA}; & l_{ff} = 2s_l h_{FA-FA} \end{cases} \qquad (11)$$

A. Link Usage in the MPLS Access Network: Linear Model

In the linear mobility model, we consider a full binary tree with the LERG as a root. The depth ℓ of a full binary tree with

N nodes is $\lfloor \log_2 N \rfloor + 1$. That is, a full binary tree of depth $\ell \geq 1$ has $N = 2^\ell - 1$ nodes (including the LERG, LSRs and LER/FAs) and the number of subnets or leaf nodes (i.e., LER/FAs) is $2^{\ell-1}$.

Let LU denote the link usage in the MPLS access network, which is the number of links used for packet delivery between the MN and the LERG. Recall that, in FMIP, MIP-RR, Mobile MPLS, H-MPLS, and FH schemes, packets are delivered using the shortest path routing. Hence, packets exchanged between the LERG and any FA traverse $\ell - 1 = \delta$ hops. However, in both FC and MFC, packets have to traverse the connection binding the LERG to the master FA and the forwarding chain binding the master FA to the MN. Recall that, in the linear mobility model, the MN moves similarly in both FC and MFC mechanisms. The mean value of LU for both FC and MFC can thus be given by

$$LU\left((\text{FC/MFC})\text{-Micro Mobile MPLS}\right) = \delta + \left(\sum_{d=1}^{\phi-1} d\Pi_d \right) \qquad (12)$$

where the first term $\delta = \ell - 1$ is the number of links from the LERG to the master FA and the second term is the mean path length used to forward packets from the master FA to the current serving FA. Note that the stationary probabilities Π_d are calculated in Section IV-A.

B. Link Usage in the MPLS Access Network: 2-D Model

As before, we assume that the distance between the LERG and any FA is the same and equal to δ. The link usage in FMIP, MIP-RR, Mobile MPLS, H-MPLS, and FH schemes is thus equal to δ. In both FC and MFC cases, we have to take also into account the mean forwarding chain size. Based on the analysis of Section IV-B, the mean forwarding chain size for the MFC case is given by

$$f(K) = 6p\Pi_0 + \sum_{i=1}^{K-2} \Pi_i(i+2p) + \sum_{i=2}^{K-2} \sum_{m=1}^{\left\lceil \frac{i-1}{2} \right\rceil} i\Pi_i^{(m)}$$

$$+ 2p(2K-3) \sum_{m=1}^{\left\lceil \frac{K-2}{2} \right\rceil} \Pi_{K-1}^{(m)} + p(3K-4)\Pi_{K-1} \qquad (13)$$

Hence, the mean value of LU for MFC-Micro Mobile MPLS is given by

$$LU(\text{MFC} - \text{Micro Mobile MPLS}) = \delta + f(K) \qquad (14)$$

In the FC case, 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 other words, the MN's next registration (i.e., LERG or local registration) depends not only on its current position, but also on its entire trajectory since it has left the master FA. This increases the complexity of analysis. Therefore, the link usage parameter of the FC mechanism will be evaluated only through simulations.

C. Registration Updates Cost: Linear Model

Let C_u denote the signaling cost of registration updates when a L3 handoff occurs. It is the traffic load of signaling messages (hop × message size) exchanged in the network when the MN

moves to a new subnet. In FMIP, the MN only performs a home registration update with the HA. In Mobile MPLS, we have to take into consideration the additional cost associated to the LSP procedure setup with the new FA. In MIP-RR, only a LERG registration update with the root of the domain is required. Additional cost, associated to the LSP procedure setup with the new FA, is to be considered in H-MPLS. In FH, a LERG registration update and a handoff notification to the old FA (in order to forward in-flight packets) are performed at every L3 handoff. Note that in this case, the LSP binding the LERG and the new FA already exists and only needs activation. In both FC and MFC, a local registration followed by an LSP procedure setup between two neighboring FAs along the extended path is required as long as the forwarding chain length does not reach the threshold. Otherwise, a LERG registration and a new LSP between the root of the domain and the new subnet are performed. This happens in the linear model when the distance of the current serving FA from the master FA is $(\phi - 1)$, and the MN moves in the direction that increases this distance (which happens with probability p). We summarize the expression of registration updates cost for all underlying protocols as follows:

$$\begin{cases} C_u(\text{FMIP}) = C_{fh} \\ C_u(\text{MIP} - \text{RR}) = C_{fg} \\ C_u(\text{Mobile MPLS}) = C_{fh} + l_{fh} \\ C_u(\text{H} - \text{MPLS}) = C_{fg} + l_{fg} \end{cases} \quad (15)$$

$$\begin{cases} C_u(\text{FH}) = C_{fg} + C'_{ff} \\ C_u(\text{FC}) = 2p\Pi_0 C'_{ff} + \left(\sum_{d=1}^{K-2} p\Pi_d C'_{ff} \right) + p\Pi_{L_{\text{th}}-1}(C'_{fg} + l_{fg}) \\ \qquad = p\left(1 + \Pi_0 - \Pi_{L_{\text{th}}-1}\right)C'_{ff} + p\Pi_{L_{\text{th}}-1}(C'_{fg} + l_{fg}) \\ C_u(\text{MFC}) = 2p\Pi_0 C'_{ff} + \left(\sum_{d=1}^{K-2} p\Pi_d \left((d+1)C_{ff} + l_{ff}\right) \right) \\ \qquad + p\Pi_{K-1}(C'_{fg} + l_{fg}) \\ \qquad = p(1 + \Pi_0 - \Pi_{K-1})C'_{ff} + pC_{ff}\left(\sum_{d=1}^{K-2} \Pi_d \right) \\ \qquad + p\Pi_{K-1}(C'_{fg} + l_{fg}) \end{cases}$$

$$(16)$$

where $C'_{ff} = C_{ff} + l_{ff}$, and $C'_{fg} = C_{fg} + l_{fg}$.

D. Registration Updates Cost: 2-D Model

The expressions of registration updates cost for the FMIP, MIP-RR, Mobile MPLS, H-MPLS, and FH schemes remains the same as in the linear model. In FC, the registration updates cost will be derived through simulations. In MFC, the MN reports its location to the LERG only when it crosses the *residing area* boundary. Using the results of Section IV-B, the expression of registration updates cost for the MFC mechanism can be written as

$$C_u(\text{MFC-Micro Mobile MPLS})$$
$$= \Pi_0 6p C'_{ff} + \sum_{i=1}^{K-2} \Pi_i C'_{ff} \left(p(i-1) + 2pi + 3p(i+1) \right)$$
$$+ \sum_{i=1}^{K-2} \sum_{m=1}^{\lceil \frac{i-1}{2} \rceil} \Pi_i^{(m)} C'_{ff} \left(2p(i-1) + 2pi + 2p(i+1) \right)$$

$$+ \sum_{m=1}^{\lceil \frac{K-2}{2} \rceil} 2p\Pi_{K-1}^{(m)} \left((K-2)C'_{ff} + (K-1)C'_{ff} + C'_{fg} \right)$$
$$+ \Pi_{K-1} \left(p(K-2)C'_{ff} + 2p(K-1)C'_{ff} + 3pC'_{fg} \right). \quad (17)$$

E. Average Handoff Time

For convenience, let $t(s, h_{x-y})$ denote the time that takes a message of size s to be forwarded from x to y via both the wired and wireless links. $t(s, h_{x-y})$ can be expressed as follows:

$$t(s, h_{x-y}) = c + h_{x-y} \times \left(\frac{s}{B_w} + L_w \right) + (h_{x-y} + 1) \times P_t$$

where

$$c = \begin{cases} (s/B_{wl}) + L_{wl}, & \text{if } x = \text{MN} \\ 0 & \text{otherwise.} \end{cases}$$

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 spends an MN without connection to any FA during the handoff process. 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:

$$T_d = \begin{cases} 0, & \text{if } T_{\text{overlap}} \geq T_{\text{ad}} \\ \frac{T_{\text{ad}}}{2} + \frac{T_{\text{overlap}}^2}{2T_{\text{ad}}} - T_{\text{overlap}}, & \text{otherwise} \end{cases} \quad (18)$$

where T_{overlap} denotes the time spent by the MN in the overlapping area. Note that with FMIP and FH schemes, the disruption time (T_d) corresponds to the physical disconnection from the old AP until the connection to the new one. As soon as the MN establishes a physical connection with the new AP, it starts receiving in-flight packets through the new FA.

Completion Time (T_c): It is the time to complete the registration update. We summarize the (T_c) value corresponding to each scheme as follows:

$$\begin{cases} T_c(\text{FMIP}) = 2t(s_u, h_{MN-HA}) \\ T_c(\text{MIP} - \text{RR}) = 2t(s_u, h_{MN-GFA}) \\ T_c(\text{Mobile MPLS}) = 2t(s_u, h_{MN-HA}) + 2t(s_l, h_{FA-HA}) \\ T_c(\text{H} - \text{MPLS}) = 2t(s_u, h_{MN-LERG}) \\ \qquad + 2t(s_l, h_{FA-LERG}) \end{cases}$$

$$(19)$$

$$\begin{cases} T_c(\text{FH}) = 2t(s_u, h_{MN-LERG} + h_{FA-FA}) \\ T_c(\text{FC}) = 2t(s_u, h_{MN-FA} + h_{FA-FA}) \\ \qquad + 2t(s_l, h_{FA-FA}) + \frac{2}{f_c(L_{\text{th}})} \\ \qquad \times \left[t(s_u, h_{MN-LERG}) + t(s_l, h_{FA-LERG}) \right] \\ T_c(\text{MFC}) = 2t(s_u, h_{MN-FA} + h_{FA-mFA}) \\ \qquad + 2t(s_l, h_{FA-mFA}) + \frac{2}{g(K)} \\ \qquad \times \left[t(s_u, h_{MN-LERG}) + t(s_l, h_{FA-LERG}) \right] \end{cases} \quad (20)$$

where $f_c(L_{\text{th}}) = \circ(L_{\text{th}})$ denotes the average number of L3 handoffs that takes place between two consecutive renewals of the forwarding chain; and $g(K) = \circ(K)$ denotes the average number of visited subnets during a cycle (i.e., before that the master FA and its associated *residing area* are renewed).

F. 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 MIP-RR, Mobile MPLS and H-MPLS, all in-flight packets will be lost during the handoff time due to the lack of any buffering mechanism. In FMIP, FH, FC, and MFC, in-flight packets would be lost till the buffering mechanism is initiated. Note that this mechanism is initiated at the old FA level in our mobility mechanisms while it is initiated at the new FA level (or new access router) in the FMIP scheme. Pkt_loss for each scheme can be expressed as follows:

$$
\begin{cases}
Pkt\_loss(\text{FMIP}) = t(s_u, h_{MN-FA} + h_{FA-FA}) \times \lambda \times N_h \\
Pkt\_loss(\text{MIP} - \text{RR}) = T_h(\text{MIP} - \text{RR}) \times \lambda \times N_h \\
Pkt\_loss(\text{Mobile MPLS}) = T_h(\text{Mobile MPLS}) \times \lambda \times N_h \\
Pkt\_loss(\text{H} - \text{MPLS}) = T_h(\text{H} - \text{MPLS}) \times \lambda \times N_h \\
Pkt\_loss((\text{FH/FC/MFC})\text{-Micro Mobile MPLS}) \\
\quad = t(s_u, h_{MN-FA}) \times \lambda \times N_h
\end{cases}
$$

$$(21)$$

G. Buffer Size Requirement

According to our mobility mechanisms, a buffer is required at the old LER/FA to store in-flight packets during each handoff operation. As stated before, the buffering mechanism is activated when the current LER/FA receives the Movement signaling message. This message notifies an imminent L2 handoff occurrence. On the other side, the buffering mechanism is disabled when the current LER/FA is notified by the MN through the new FA to forward in-flight packets. In both FH and MFC, the signaling message used to notify the old FA corresponds to the handoff notification message. In FC, it corresponds to the forwarding chain update message. The buffer size requirement (Buf_size) for FMIP, FH, FC, and MFC is listed as follows:

$$
\begin{cases}
Buf\_size(\text{FMIP}) = [T_d + t(s_u, h_{MN-FA})] \times \lambda \\
Buf\_size((\text{FH/FC/MFC})\text{-Micro Mobile MPLS}) \; . \\
\quad = [T_d + t(s_u, h_{MN-FA} + h_{FA-FA})] \times \lambda
\end{cases}
$$

$$(22)$$

VI. NUMERICAL AND SIMULATION RESULTS

In this section, we compare all underlying protocols using both analytical and simulation approaches. The parameter settings in our experiments are listed in Table I.

Fig. 11 represents the link usage cost of all underlying protocols for both the linear and the 2-D mobility models. We can see that FC and MFC have the same cost in both mobility models when $K = L_{\text{th}}$. This cost is higher compared to the remaining protocols. This slight difference is due to the additional cost introduced by the forwarding chain.

Fig. 12(a) plots the different registration updates costs at every L3 handoff for the linear mobility model. The FH scheme exhibits a lower registration cost than FMIP, Mobile MPLS and H-MPLS protocols, since the required LSP between the LERG and the new FA already exists. However, it has a higher registration cost than MIP-RR, due to the extra signaling messages sent to the old subnet to forward in-flight packets. FC and MFC, on the other hand, provide the lowest registration cost since some expensive LERG registration updates are replaced by low-cost local registrations.
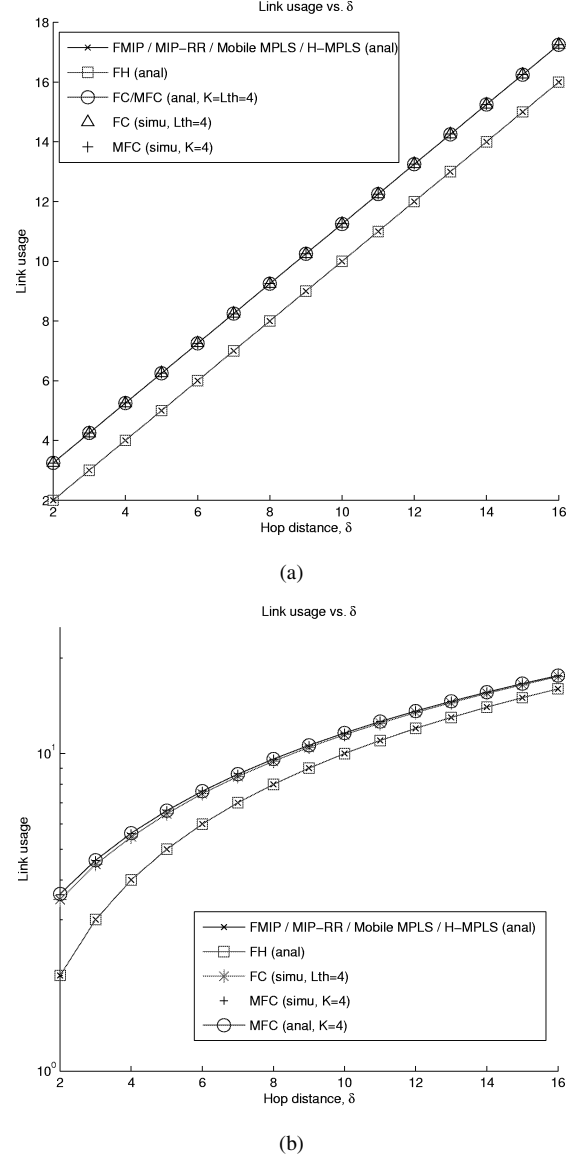

Fig. 11. Link usage cost. (a) Linear model. (b) 2-D model.

TABLE I
PARAMETER SETTINGS

Parameter	Value	Parameter	Value
t_s	1000 sec	$h_{HA-LERG}$	4
t_r	$5 \sim 50$ sec (default 20)	P_t	10^{-6} sec
T_{ad}	1 sec	B_w	100 Mbps
L2 Beacon	100 msec	B_{wl}	11 Mbps
s_u	48 bytes	L_w	1 msec
s_l	28 bytes	L_{wl}	2 msec
δ	$2 \sim 16$ (default 9)	ℓ	$3 \sim 17$
ϕ (L_{th} or K)	$1 \sim 15$ (default 4)	λ	64 Kbps

Fig. 12(b) shows the 2-D case. As before, both FC and MFC schemes exhibit the smallest costs. However, this figure shows that minimum registration cost is obtained by different strategies according to the value of δ. In this particular case (i.e., $K = L_{\text{th}} = 4$), the FC scheme stands out as the best choice when $\delta \leq 7$, otherwise the MFC scheme provides the best cost. Recall that considering the same threshold (i.e., $K = L_{\text{th}}$), the expensive registration updates with the LERG are more frequent in the FC case. However, the local registration cost in the FC strategy
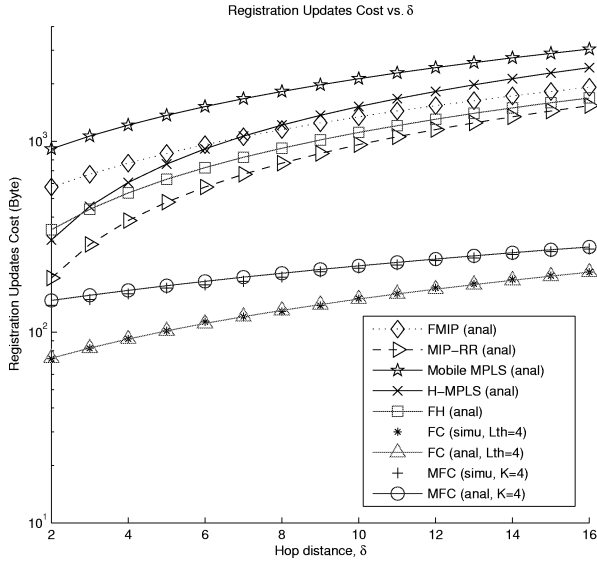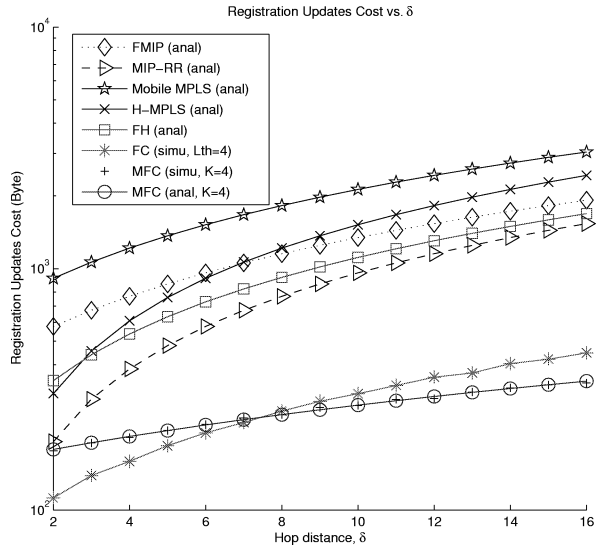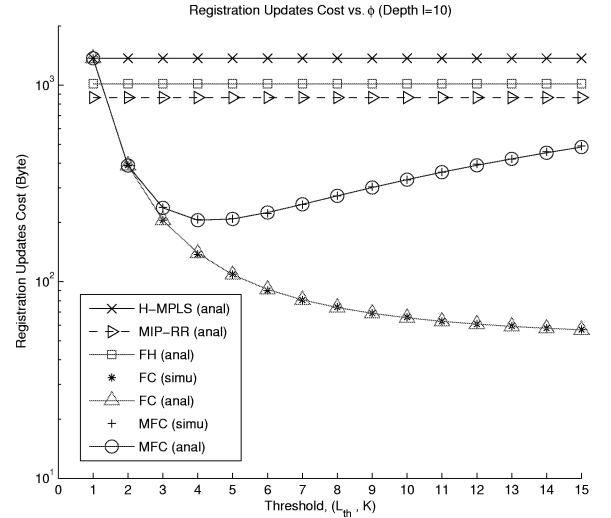
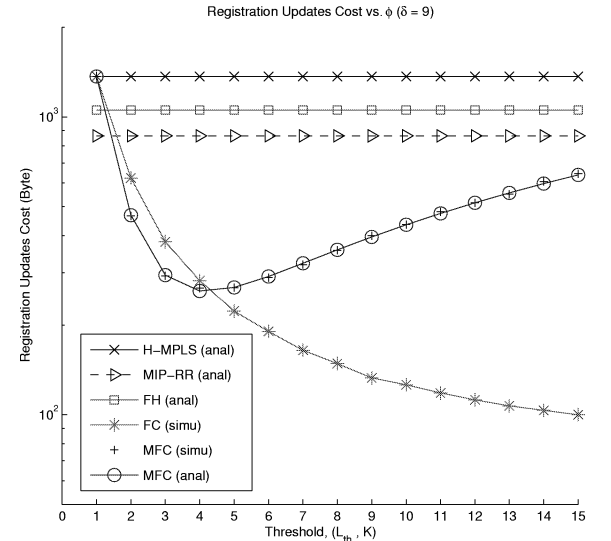Fig. 12. Registration updates cost at every L3 handoff. (a) Linear model. (b) 2-D model.

Fig. 13. Effect of L_{th} and K on the registration updates cost in (FC/MFC)-Micro Mobile MPLS. (a) Linear model. (b) 2-D model.

is cheaper than the MFC one. In this regard, when δ is large, the LERG registration cost is a dominant cost. Hence, the MFC mechanism stands out as the best choice. Otherwise, when δ is relatively small, FC becomes the best choice.

Fig. 13 depicts the cost of registration updates for FC and MFC as a function of their respective thresholds and for both mobility models. We can observe that the cost of registration updates for MFC is a convex function of K, where the minimum cost is obtained for K_{opt} (in these figures $K_{\text{opt}} = 4$). In fact, the LERG registration frequency decreases with the increase of the threshold $\phi = K$, since more and more expensive LERG registrations are replaced by local registrations. On the other hand, the local registration cost increases with the threshold, since the average distance between the MN and the master FA increases. In view of this, the optimal cost is a trade-off between

these two opposite requirements. Specifically, the registration updates cost of MFC can be written as follows:

$$
\begin{aligned}
C_u(&\text{MFC} - \text{Micro Mobile MPLS}) \\
&= f(K) \times (\text{registration cost between two neighboring FAs}) \\
&\quad + \frac{1}{g(K)} \times (\text{LERG registration cost}) \\
&= f(K)C'_{ff} + \frac{1}{g(K)}C'_{fg}
\end{aligned}
$$

where $f(K) = \circ(K)$ denotes the average distance (in terms of number of hops) between the MN and the master FA and $g(K) = \circ(K)$ denotes the average number of visited subnets during a cycle (i.e., inside the *residing area*). This formula exhibits clearly the convex behavior of the MFC registration updates cost. Finally, we notice that analytical results, for both 2-D and linear mobility models, practically coincide with the simulation results, which illustrates the accuracy of our study.
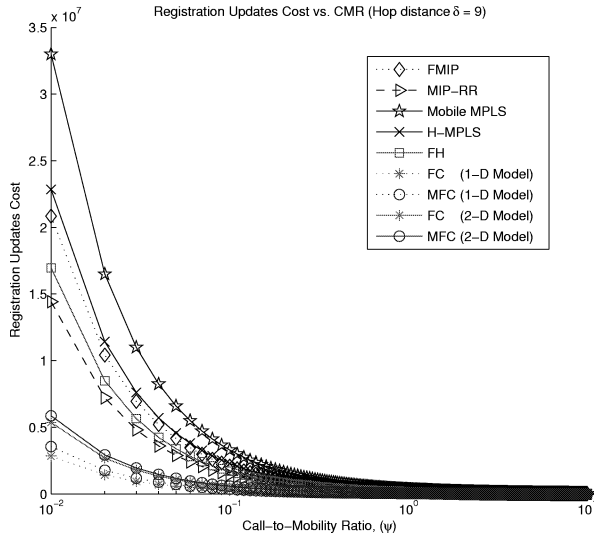
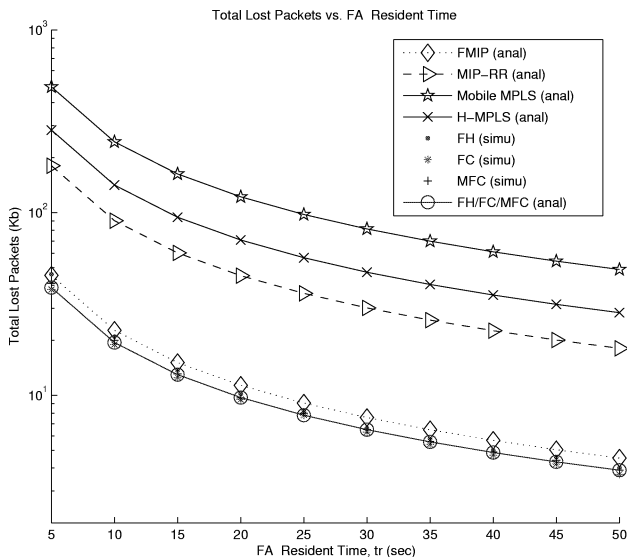Fig. 14. Effect of CMR on the registration updates cost.

Fig. 15. Total lost packets during a session.

On the other side, in FC, the cost of registration updates decreases with the threshold $\phi = L_{\text{th}}$. Indeed, the expensive LERG registrations become less frequent. However, we note that the threshold value will be limited by delay constraint. Typically, delay sensitive applications, such as video or voice services, will require small values of L_{th} to ensure acceptable end-to-end delay. Finally, it is worth noting that the variation of ϕ (i.e., K or L_{th}) does not affect the performance of the other studied protocols (FMIP, MIP-RR, Mobile MPLS, H-MPLS, and FH). Consequently, the results presented in Figs. 11 and 12 for these schemes are ϕ-independent.

Fig. 14 shows the effect of the call-to-mobility ratio (CMR) on the registration updates cost for different schemes under both mobility models and using the optimal value $K_{\text{opt}} = 4$. CMR is defined as the ratio of the packet arrival rate λ to the mobility rate. In this figure, we observe that when the CMR is small (i.e., when the MN handoffs frequently), the FH scheme generates less signaling traffic than the FMIP, Mobile MPLS and H-MPLS schemes, which are more suitable for mobile users with high CMR. We notice also that both FC and MFC can significantly reduce the registration updates cost mainly when the CMR value is

TABLE II
AVERAGE HANDOFF TIME IN MILLISECONDS

FMIP	30.198
Mobile MPLS	56.256
MIP-RR	22.159
H-MPLS	40.199
FH-Micro Mobile MPLS	24.171
FC-Micro Mobile MPLS	18.136
MFC-Micro Mobile MPLS	20.145

low. Our results demonstrate that these schemes can reduce the MIP-RR signaling cost by 72% and 57% when the CMR is small for the linear and 2-D mobility models, respectively. Note that, this considerable gain is obtained with respect to the MIP-RR scheme, which exhibits the best cost among the existing protocols. According to these results, FC and MFC mechanisms are the best strategies for MNs with high mobility rate.

The average handoff time values for different schemes are listed in Table II. Each value was obtained by averaging 100 consecutive simulations. The simulated access network consists of a full binary tree of depth $\ell = 10$ and $K = L_{\text{th}} = 4$. Every leaf node (LER/FA) is connected to one AP. During simulations, the MN moves periodically between neighboring APs and receives downlink packets. The simulations are run using the Network Simulator NS-2.[1] As can be seen, FH improves the handoff performance when compared with FMIP, Mobile MPLS and H-MPLS since the registrations are performed within a local domain and the required active LSP between the root and the new subnet already exists. Furthermore, FC provides the lowest average handoff time, since in this case the registrations are often carried out with the previous FA instead of the LERG, which enables shorter delay to complete the registration updates. The handoff delay increases slightly with MFC compared to FC, since the MN performs registrations with the master FA instead of the previous FA.

Fig. 15 shows the amount of packet losses during the whole connection session for different schemes assuming the same scenario. 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. In contrast, FH, FC, and MFC provide 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 more than 30 thousands of MNs.

VII. CONCLUSION

This paper described a new mobility management scheme, called Micro Mobile MPLS, that supports both mobility management and QoS resource provisioning in IP/MPLS-based wireless access networks. Micro Mobile MPLS integrates three mobility management mechanisms: FH, FC, and MFC. With FH, we anticipate the LSP establishment before the MN moves into a new subnet to reduce service disruption by using the L2 functionalities. With FC and MFC, we track the MN's movement using the concepts of forwarding chains and residing area, respectively. We demonstrated how our mechanisms reduce the cost of registration updates and provide low handoff

[1]ns2 home page. http://www.isi.edu/nsnam/ns

latency and small packet loss rate. In particular, we showed that the proposed mechanisms provide short handoff delay, with a slight advantage for the FC protocol. Also we found the FC mechanism to yield the smallest registration updates cost always when considering linear mobile movements. In the 2-D mobility case, we found that the MFC protocol stands out as the best choice for delay sensitive applications; otherwise the FC mechanism provides the lowest cost. Using both analytical and simulation approaches, we compared our proposed mechanisms with existing solutions (FMIP, MIP-RR, Mobile MPLS and H-MPLS). We analytically derived the registration updates cost, the link usage and the handoff performance for all protocols and we found that our proposed mechanisms achieve a substantial signaling cost reduction and improve the handoff performance at the price of a slight increase in link usage cost.

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, A. G. Valko, C.-Y. Wan, and Z. R. Turanyi, "Design, implementation, and evaluation of cellular IP," *IEEE Pers. Commun. Mag.*, vol. 7, no. 4, pp. 42–49, Aug. 2000.
[4] R. Ramjee *et al.*, "HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 3, pp. 396–410, Jun. 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, no. 3, pp. 138–145, Mar. 2002.
[6] K. El Malki, "Low latency handoffs in Mobile IPv4," IETF, draft-ietf-mobileip-lowlatency-handoffs-v4-11.txt (work in progress), Oct. 2005.
[7] R. Koodli, "Fast handovers for Mobile IPv6," RFC 4068, Jul. 2005.
[8] J. Xie and I. F. Akyildiz, "A distributed dynamic regional location management scheme for mobile IP," in *Proc. IEEE INFOCOM*, 2002, vol. 2, pp. 1069–1078.
[9] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IPv4 regional registration," IETF, draft-ietf-mip4-reg-tunnel-01.txt (work in progress), Nov. 2005.
[10] Y. Bejerano and I. Cidon, "An anchor chain scheme for IP mobility management," *Wireless Netw.*, vol. 9, no. 5, pp. 409–420, 2003.
[11] C. Chu and C. Weng, "Pointer forwarding MIPv6 mobility management," in *Proc. IEEE Globecom*, Nov. 2002, vol. 3, pp. 2133–2137.
[12] W. Ma and Y. Fang, "Dynamic hierarchical mobility management strategy for mobile IP networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 4, pp. 664–676, May 2004.
[13] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture," IETF, Internet RFC 3031, Jan. 2001.
[14] Z. Ren, C. Tham, C. Foo, and C. Ko, "Integration of mobile IP and multi-protocol label switching," in *Proc. IEEE ICC*, 2001, vol. 7, pp. 2123–2127.
[15] F. M. Chiussi, D. A. Khotimsky, and S. Krishnan, "Mobility management in third-generation all-IP networks," *IEEE Commun. Mag.*, vol. 40, no. 9, pp. 124–135, Sep. 2002.
[16] T. Yang, Y. Dong, Y. Zhang, and D. Makrakis, "Practical approaches for supporting micro mobility with MPLS," in *Int. Conf. Telecommunications (ICT 2002)*, Beijing, China, 2002.
[17] K. Xie, V. Wong, and V. Leung, "Support of micro-mobility in MPLS-based wireless access networks," in *Proc. IEEE WCNC*, Mar. 2003, vol. 2, pp. 1242–1247.
[18] 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*, 2003, vol. 1, pp. 250–255.
[19] R. Langar, S. Tohme, and N. Bouabdallah, "Mobility management support and performance analysis for wireless MPLS networks," *ACM-Wiley Int. J. Netw. Manag.*, vol. 16, no. 4, pp. 279–294, Jul. 2006.
[20] D. Awduche, L. Berger, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP tunnels," IETF, Internet RFC 3209, Dec. 2001.
[21] B. Jamoussi *et al.*, "Constraint-based LSP setup using LDP," IETF, Internet RFC 3212, Jan. 2002.
[22] R. Langar, N. Bouabdallah, S. Tohme, and R. Boutaba, "Mobility modeling and handoff analysis for IP/MPLS-based cellular networks," in *Proc. 49th IEEE Globecom*, San Francisco, CA, Nov. 2006.
[23] G. Rubino and B. Sericol, "A finite characterization of weak lumpable Markov processes. Part II: The continuous time case," *Stochastic Processes and Their Applications*, vol. 45, 1993.

Rami Langar (M'06) received the B.S. degree in telecommunications engineering from Sup'Com, Tunis, Tunisia, in 2001, and the M.S. degree from the University of Paris VI, France, in 2002 and the Ph.D. degree from ENST-Paris, France, in 2006, both in network and computer science. In 2006, he was also with ENS-Cachan, France, as a teaching and research assistant.

He is currently a Postdoctoral Research Fellow in the School of Computer Science, University of Waterloo, Waterloo, ON, Canada. His research interests include 4G visions, mobility management, resource management, wireless mesh networks, performance evaluation and quality of service support.

Nizar Bouabdallah received the B.S. degree in telecommunications engineering from Sup'Com, Tunis, Tunisia, in 2001, and the M.S. and Ph.D. degrees in network and computer science from the University of Paris VI, 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, Rennes, France. His research interests include optical networking, wireless and sensor networks, performance evaluation, network planning and modeling, as well as control and management architectures.

Raouf Boutaba (M'93–SM'01) received the M.Sc. and Ph.D. degrees in computer science from the University of Paris VI, 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, and 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.