Fast Handover for Multiple-interface Mobile Devices Connecting to a Single Foreign Agent in MIPv4 with MIH Support

Beizhong Chen Ivan Marsic
Rutgers University, Piscataway, NJ 08854
{bzchen, marsic}@caip.rutgers.edu

David Faucher Cezar Purzynski Ray Miller Sharma Sameer
Bell Labs, Murray Hill, NJ 07974
{ dfaucher, cezar, rbmiller,svsharma }@alcatel-lucent.com

Abstract–**Growing number of mobile clients are equipped with multiple interfaces that can receive services from multiple service providers, possibly via different access technologies. Media Independent Handover (MIH, IEEE 802.21) is an ongoing, evolving effort to provide a generic solution for mobility across access networks. Collaborating with Layer 3 Mobility Management protocol, MIH makes it possible to achieve better performance with respect to latency and packet loss rate than the existing mechanisms. Currently, Mobile IPv4 (MIPv4) is the dominant mechanism and is expected to persist into the future. However, MIPv4 and its existing improvements do not perform well when the current active interface goes down unexpectedly and the second interface is connected via the same foreign agent (FA). To achieve better performance, we first introduce a new concept of Interface Simultaneous/Backup FA Binding between a mobile node (MN) and a single FA. Based on this, we propose a novel mobile IP extension that realizes fast handover. Compared with the existing mechanism, our experiments show that our algorithm achieves much better performance with MIH support. Our method also allows FA-bicasting, which can help improve transmission reliability by combining traffic from different links.**

Keywords- MIH; Mobile IP; Fast Handover; WiFi; Ethernet; 2G/3G; WiMax; Multiple Interface; IEEE 802.21;Mobility

I. INTRODUCTION

In the past decade, different wireline/wireless technologies, e.g., Ethernet, WiFi, 3GPP, 3GPP2, Wi-Max, have evolved down distinct paths, each providing services with different characteristics. For example, WiFi is mainly deployed indoors providing short-range connectivity while 3GPP provides a much longer-range service outdoors. In order to provide better coverage and improved user experience, some vendors offer proprietary dual-mode mobile terminals that allow users to access two different networks, e.g., WiFi and GSM. IEEE 802.21 (Media Independent handover, MIH) [5] aims to provide a generic solution for intelligent handover between heterogeneous technologies. Fig.1 shows the placement of the MIH function in the protocol stack, sandwiched between Layer 2 (L2) and Layer 3 (L3). MIH is intended to assist in dissemination of handover-related information (e.g., link states, handover progress, etc.) to the upper layers (via event service), enabling them to monitor and control the handover (via command service).

Client mobile IP (MIP) [2] is the most popular L3 mobility management protocol. A mobile node (MN) entering a MIP-enabled network searches for a Foreign Agent (FA) that can act as the MN's Care of Address (CoA). Through a registration procedure initiated by the MN, a tunnel is established between the FA and the Home Agent (HA). All data from Correspondent Nodes (CN) destined for the MN are intercepted by the HA and forwarded over this tunnel to the current CoA. The FA is then responsible for delivering the data to the MN. The MIP mechanism provides a transparent solution for mobility management, which means that the CN need not know where the MN resides (e.g., home or mobile). When a MN roams from one network to another, a handover must be completed to allow for session continuity.

Vertical handover occurs when a MN roams into a network served by a different access technology, (e.g., from a WLAN network to a 3GPP network), while a handover between networks served by the same access technology is termed *horizontal handover*. Handovers can also be classified as Layer 2 or Layer 3. Typically, a L3 handover occurs after a L2 handover is completed. A handover involves many operations, e.g., physical reconnection, protocol negotiation, reconfiguration, etc., and it is inevitable to have performance deterioration during the handover. Typically, during a handover process, latencies are introduced that include L2 disconnection detection, L2 reconnection, L3 re-connection, etc. Fig. 2 shows the detailed handover timing diagram for the client Mobile IP. The followings are descriptions of the variables in Fig.2 (each variable represents a time interval except for T15):

- *T1*: L2 disconnection detection
- *T2*: L2 connection up (e.g., power up, association)
- *T3*: MN solicitation
- *T4*: FA advertisement
- *T5*: MN registration request
- *T6*: FA processing and MN-FA authentication
- *T7*: Registration forwarding
- *T8*: Home Agent (HA) processing
- *T9*: Authentication

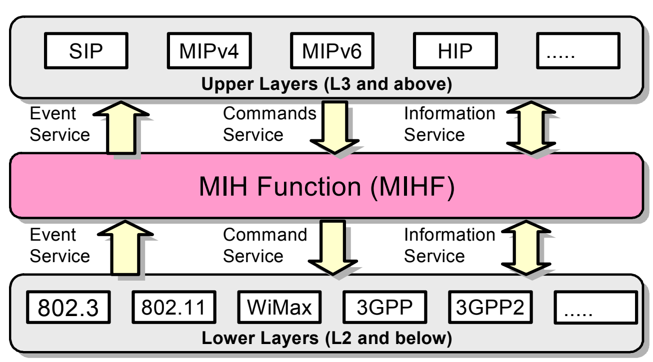

Fig. 1. MIH function

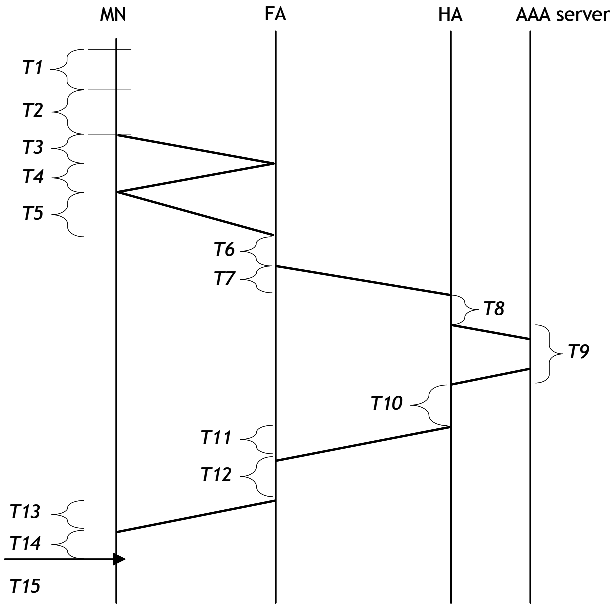

Fig. 2. Timing diagram for handover using Mobile IP

- *T10*: HA processing
- *T11*: Registration reply
- *T12*: FA processing
- *T13*: Registration reply
- *T14*: MN update (e.g., routing table update)
- *T15*: Connection resumed

Many methods have been proposed to reduce the various types of handover latencies, mostly through switching the order of L2/L3 handover, or reducing the absolute latency value by localizing the L3 registration processing. As an example of the first approach, Low-latency Handover [1] proposes a method called *pre-registration* that allows the MN to register with a new FA through the current FA before attaching to a new network. Hence, the MN does not need to wait until the L2 handover is completed to start the L3 handover. Ideally, the L3 handover latency, *T3 ~ T14* in Fig. 2, can be removed and therefore a quicker handover is achieved. Regional Registration [3] suggests a hierarchical architecture that alleviates the high latency of MN-HA registration and authentication by adding a regional GFA (Gateway Foreign Agent) to provide a quicker registration locally. All of the abovementioned algorithms are designed for mobiles with a single interface registering with two different FAs. However, for vertical handover on the same FA, which is our focus, these methods are not directly applicable without extensive modifications. Note that different access technologies typically use different authentication mechanisms.

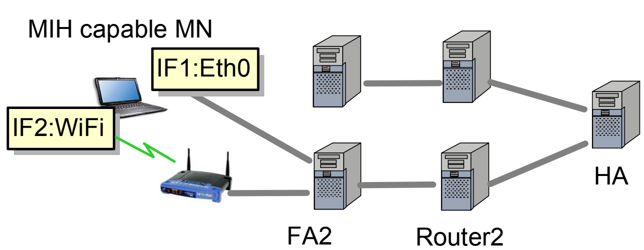

Fig. 3. MN with 2 interfaces connecting to a single FA

II. Multiple-interface MN Connecting to a FA

A. Existing Problem

Here we review the traditional client MIP and identify an existing problem. Imagine a scenario where a user with an Ethernet- and WiFi-equipped laptop connects via the same FA regardless of which interface is active (Fig. 3). We assume that MIP service is available and that the same FA is used for both Ethernet and WiFi in the same building. While in the office, the user prefers to use Ethernet, which typically provides better connection quality than wireless link. When leaving the office the connection must be transferred to WiFi. Generally, unplugging an Ethernet cable is not pre-detectable and therefore, the L2 disconnection is an unexpected event. If the user is using the client MIP without MIH, she will lose her connection for a relatively long period of time until the MIP detects the disconnection by methods defined in [2]. For the case where MIH is available, a quicker handover can be achieved because the MIH can notify the MIP that the L2 connection is lost. Using IF2, the MIP client in MN starts a L2/L3 handover process that goes through the *T1~T15* stages in Fig. 2, and the L3 handover occurs after L2 handover is completed. If we can have IF2 ready for handover at all times, regardless of when the IF1 connection is lost, the MN can switch to IF2 and therefore eliminate most of the *T2~T15* stage and a faster handover can be achieved.

In the client MIP, when a new registration is authenticated, the FA updates its binding record. The new registration record replaces the current one. Specifically, assume that the MN is using IF1 (Eth0) and it sends a new registration request over IF2 (WiFi). The FA forwards the request to the HA. From the HA's perspective, the registration request is no different from the one sent over Eth0 (i.e., it looks like a re-registration rather than a move) and only the expiration time is updated. However, in the FA, the link-layer address in the binding record (visitor list) will be replaced by the link-layer address of IF2 once the HA reply is received. As a result, the active connection will be transferred to IF2 even if IF1 can still adequately support the connection. The desired behavior would be to maintain current FA registration and add a secondary one in preparation for a handover. However, in the current client MIP implementation there is no mechanism to add a secondary registration and, more importantly, to control the state of the registered connections.

B. Interface Simultaneous/Backup FA Binding and a Novel MIP Extension

In order to reduce the latency of handovers caused by unpredictable events, it is desired to have a handover-ready backup interface. To that goal, we introduce a concept of Interface Simultaneous/Backup FA Binding, which allows

TABLE I
EXAMPLE OF FA VISITOR TABLE

	IP	MAC	Active	Lifetime	Auth
IF1	xx.xx.xx.xx	xx:xx:xx:xx	1	xx	1
IF2		xx:xx:xx:xx	0	xx	0

the FA to maintain records of multiple registered interfaces (from the same MN). Based on this, we propose a novel MIP extension. The first step is to allow simultaneous interface binding by expanding the FA visitor list to include records of all registered interfaces, along with their link-layer addresses and active/backup status. Active interfaces are involved in forwarding the traffic to the MN. The second step is to define a new MIP extension (Fig. 4), called *simultaneous FA binding extension,* which is denoted as SFB-MIP hereafter. Note that we use the conventional term of bi-casting even if ≥ 2 interfaces can be used to send duplicate packets.

C. Basic Flow

In this section, we show the basic flow to set up a simultaneous FA binding. Assume the MN is connected to a FA using IF1 and it begins to register IF2.

1. *The MN sends a registration request with the new extension over IF2.*
2. *The FA checks the registration and finds this extension. From the MN home IP, FA knows the request comes from the same MN as IF1.*
   a. *The FA forwards the registration to the HA for authentication and timer update, removing the simultaneous FA binding extension.*
   b. *The FA updates the visitor list. Table I shows an example.*
3. *The HA sends reply.*
4. *The FA updates its visitor list again based on the reply. If the HA accepts the request, the FA marks the corresponding entry in the visitors list as authenticated (auth). If the HA denies the request, IF2 needs to be removed from the visitor list.*
5. *The FA relays the HA reply to the MN using the interface over which the original request was received (IF2).*
6. *MN updates its interface record.*
7. *When MN needs to use IF2, e.g., when MIH detects that IF1 is unplugged and notifies MIP, the MIP sends a registration through IF2 with the "A" bit set and "B" bit unset to inform FA to use IF2 to forward packets. Meanwhile, the MN performs the necessary local update, e.g., Routing Table update. Note that if "A" bit and "B" bit are both set, the FA enters bi-casting state and both IFs are active.*
8. *Because IF2 is marked as "auth", the FA sets IF2 as active and deactivates IF1, and sends an intermediate reply to MN. After this, the MAC of IF2 is used to send packets and the transmission is moved to IF2, eliminating most of the latency from T3~T15. Additionally, the FA forwards the registration (without the extension) to the HA for normal processing (this include forwarding any HA registration reply to the MN when it is received from the HA).*

```
0                   1                   2
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
+-+-+-+-+-+-+-+-+-+-+-+-++-+-+-+-+-+-+-+-+-
|      Type     |      Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Lifetime                 |A|B|Reserved
+-+-+-+-+-+-+-+-+-+-++-+-+-+-+-+-+-+-+-++-+-+
```

Type: Identifies this simultaneous FA binding extension.

Length: Indicates the length (in bytes) of the data fields in this extension, not including the Type and Length bytes.

Lifetime: The time in seconds before the binding with the FA is considered expired. A zero value indicates a MN's request for deregistration of the specific interface over which this extension is sent. Note that it is different from the lifetime of MN registered at HA.

A: Active bit. If the "A" bit is set, this link is active and should be used to forward traffic. If it is unset, this link is not ready for transmission.

B: Bi-casting. If this bit is set, the FA uses all active interfaces to forward traffic. If unset, only the interface over which the latest registration is sent will be the active interface.

Fig. 4. Simultaneous FA binding extension

The basic rules for processing of the "A" and "B" bits are:

(1) *"A" bit set, "B" bit set: add this interface to the set of active interfaces*
(2) *"A" bit set, "B" bit unset: activate this interface and deactivate all other interfaces, just as the step 7 in the basic flow*
(3) *"A" bit unset: either pre-registering an interface that you do not want activated or deactivating an active interface, in all cases the "B" bit is ignored.*

After two interfaces from the same MN are bound to a FA, we say that this FA is in the FA-backup-binding state for the given MN.

When a MN activates an interface, it sends to the FA a MIP registration with this extension and the "A" bit set. The FA forwards the MIP registration (with this extension removed) to the HA and sends an intermediate reply to the MN. Since the MN has already been authenticated, the reply is used to notify the MN that it can begin using the new interface immediately (i.e., it needs not wait for the MIP registration reply). The MN will retry if no response is received within a short period (depending on the specific link technology).

For simplicity, the basic flow described above is based on the case with only two interfaces present. It can be easily extended for MNs with > 2 interfaces.

Before registering the second interface, MN needs to determine if both interfaces are connected to the same or different FAs. This can be easily determined by checking the agent advertisement or the reply to the MN solicitation.

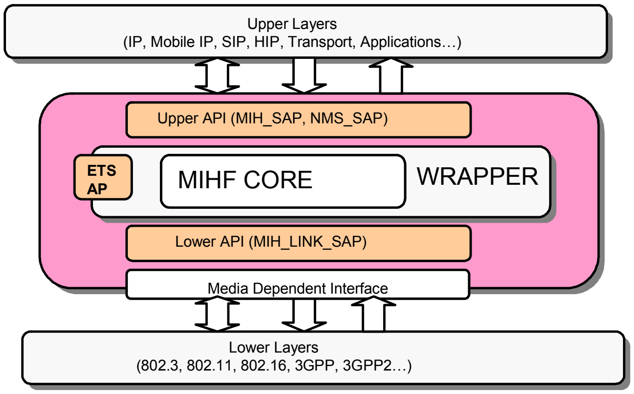

Fig. 5. MIH software architecture

Interface deregistration is achieved when the MN sends a registration request that includes this new extension with a zero-lifetime over the interface that needs to be deregistered. It is also possible that the MN will not send any re-registration messages through the interface. In this case, the FA removes the corresponding interface when the timer expires. Another option arises when the MN sends a de-registration message without the extension attached. In this case, all interfaces are deregistered at the FA and the entire MN-FA context is torn down. Note that this deregistration needs to be authenticated by the HA, as required by [2].

Sometimes, it is possible to achieve a better transmission performance using FA bi-casting in the presence of high data error rate in interfaces. Our method can provide FA bi-casting if multiple interfaces are active simultaneously. Then, packets are duplicated, properly framed for each active L2 connection and sent via a corresponding interface. We call this FA bi-casting, which is unlike the conventional HA bi-casting in that the packets are duplicated in the FA, rather than in HA.

D. Security and Other Issues

In the client MIP, the MN-FA authentication is optional while the MN-HA authentication is mandatory. However, generally, to achieve fast handover when multiple interfaces are connecting to a single FA, security association between MN and FA is required, because the messages to switch an active interface have to be authenticated by the FA to achieve desired latency reduction.

The interface Simultaneous/backup FA binding messages are transparent to the HA. To achieve this, the FA removes the extension after it processes the registration request. The extension order should be as defined in [2]. Hence, the authentication between MN and HA will not get out of sync.

III. Test-bed Setup

For our research, we have developed an implementation of the IEEE 802.21 (MIH) draft. Due to lack of space, we only show the high-level software architecture in Fig. 5 without giving detailed explanation. The MIH detects the changes in the link layer and sends a link event to the MIP, which operates accordingly. Fig. 6 illustrates the basic topology of our test-bed. The interface drivers are modified to provide MIH capabilities and faster link status detection. The simulators can introduce impairments, such as traffic delay, packet loss, etc.

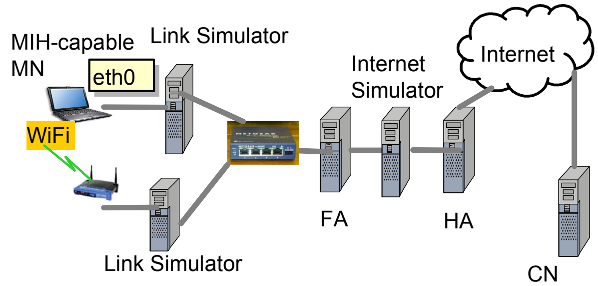

Fig. 6. Test bed topology

IV. Performance Evaluation

To our knowledge, our algorithm is the first to address the fast handover problem for multiple-interface devices connecting to a single FA. Therefore, we only compare the SFB-MIP with the MIP. To measure the latency accurately, we implement a packet generator in the CN and a data sink in the MN. The packet generator keeps sending back-to-back packets (100 Bytes each) with time-stamps, to allow for latency calculation.

First, we show that MIH support is important for SFB-MIP. In MIP, two basic methods are defined in [2] for move detection. The first one is to setup a timer based on the agent-advertised lifetime. Failure to receive a new advertisement when the timer expires implies a disconnection. The second method uses new network prefixes. When a MN receives an advertisement with a new network prefix, it concludes that it has moved. However, when multiple interfaces of a MN connect to the same FA, this method is unsuitable for move detection. Table II shows a typical result for the case when the lifetime in agent advertisement is 90 seconds and the interval between advertisements is 30 seconds. Obviously, the move detection time is fairly large which is impractical in many scenarios. To have a quicker detection, we can reduce the interval time between the advertisements. However, the advertisement message will consume more channel capacity, which is undesirable.

TABLE II
Handover (HO) latency without MIH support

HO latency (sec) Without MIH	1	2	3	4	Average
	83	71	46	72	68

Second, we compare SFB-MIP with MIP under different FA-HA latencies (Fig. 7) with MIH support. The data from 10 rounds of experiments are averaged. The results show a great advantage of SFB-MIP. As seen in Fig. 7, with the increase of latency between the FA and HA, for MIP, the handover latency keeps rising, whereas for SFB-MIP, the handover latency is not affected. The SFB-MIP shows increasing benefit when the FA-HA latency grows. This is

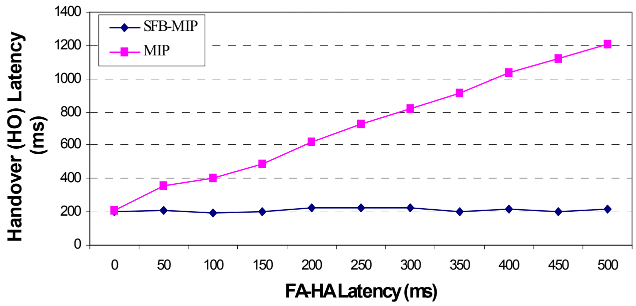

Fig. 7. Latency comparison with MIH support

just as expected and validates our algorithm. To give a clearer idea about the latency, we show the disconnection-detection latency and handover latency of one set of typical experiments in Fig. 8. As seen, the handover latency mainly depends on the detection latency. (The variation is because the operating system schedules different tasks at discrete time intervals, so a disconnection will not be detected immediately; the variation appears great only due to the vertical axis zoom-in.) The SFB-MIP latency, which is the difference between the handover latency and the detection latency, is only around 20ms on average. The average handover latency is around 200ms which implies that the disconnection is not perceivable to a VoIP user when she unplugs the cable. The detection latency can be greatly reduced if the operation system is modified accordingly.

Data loss of traffic stream during handover is directly dependent on the handover latency and the data rate. Table III shows that SFB-MIP achieves much better performance compared to MIP with a common FA-HA delay.

V. Discussion and Future Work

Simultaneous binding (or registration) as defined in client MIP [2] has quite a different meaning from that in SFB-MIP. In MIP, simultaneous binding means that a MN registers at a HA using two or more FAs. In SFB-MIP, it means that multiple interfaces simultaneously bind to the same FA. We also call it "backup binding" because one or more interfaces are ready for use in case of an unexpected disconnection. In cases where handover occurs unexpectedly, our method yields a great improvement in handover latency. The benefits might be reduced when the handover can be predicted in advance, giving the secondary interface enough time to perform full registration. Even then, one can still take advantage of the bi-casting feature.

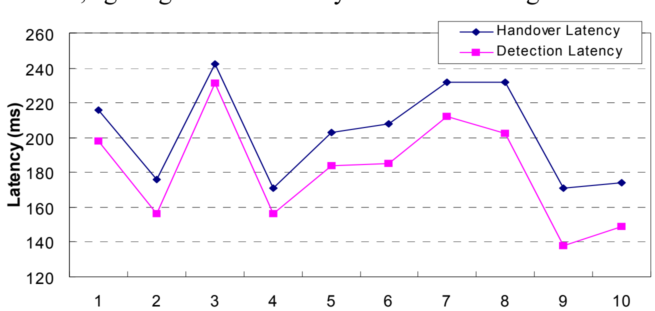

Fig. 8. Detection and Handover latency of SFB-MIP, FA-HA latency=400ms, with MIH support. (Notice the scale difference of the vertical axis compared with Fig.7)

TABLE III
Data-Loss for voice data, stream rate=16Kbps, packet Size = 64 bytes, FA-HA latency=100 ms

	Observed Data loss	Observed Packet loss
SFB-MIP	256~448 bytes	4~7 packets
MIP	576~960 bytes	9~15 packets

SFB-MIP achieves a much faster handover when an unexpected L2 handover occurs. In the example given in Section II.A above, when the user unplugs the Ethernet cable (IF1), the MN detects this event with the help from MIH support. The MN sends a message to FA to activate the WiFi (IF2) without the latency of *T2* and *T5~T14* in Fig. 2, but the MN still needs to send a simultaneous FA binding extension.

The re-registration with FA and HA can be combined in one message and therefore, the number of re-registration requests to HA will not necessarily increase because they can be sent over the bound interfaces in turn, combining with the re-registration. The identification field defined in [2] suffices to distinguish the reply for each registration request.

Binding multiple interfaces to an FA does not imply that all registered interfaces need to stay powered all the time. The bound-but-inactive interfaces can go into a power-saving mode, and they just need to wake up periodically to refresh the binding timer at the FA.

Our current evaluation is based on WiFi and Ethernet. In the near future, we are planning to extend our evaluation to include EvDO and WiMax. Our preliminary results show that the EvDO L2 handover latency is significantly longer than that of WiFi and we expect SFB-MIP to provide greater performance improvement in the case when EvDO interfaces are involved.

VI. Conclusion

In this paper, we identified a previously unknown weakness of client MIP to handle multiple-interface mobile node. We proposed Interface Simultaneous/Backup FA Binding and a new Mobile IP (MIP) extension to achieve fast handover with MIH (IEEE 802.21) support, particularly when the current link goes down unexpectedly. Our proposal also makes FA bi-casting possible, which in some scenarios can help improve transmission reliability. Our experiments validated superior performance of our scheme. We also showed that MIH (IEEE 802.21) could help to reduce handover latency significantly.

VII. References

[1] K. El Malki, Ed. Athonet, "Low-Latency Handoffs in Mobile IPv4", RFC 4881, 2007.
[2] E. Perkins, "Mobility Support for IPv4", RFC 3344, August 2002.
[3] E. Fogelstroem, A. Jonsson, C. Perkins, "Mobile IPv4 Regional Registration", RFC 4857. June 2007
[4] Dynamics Mobile IP System, developed at Helsinki University of Technology, http://dynamics.sourceforge.net/
[5] IEEE 802.21/D8.0, Dec 2007.