Fast Handover Support in a WLAN Environment: Challenges and Perspectives

Lila Dimopoulou, Georgios Leoleis, and Iakovos S. Venieris
National Technical University of Athens, Greece

Abstract

While handover management has traditionally used radio-technology-specific mechanisms, the need for integration of this diverse network environment has obviated the "push" of the handover functionality to the generic IP layer that serves the rendezvous point of underlying technologies. In this context, we study and analyze the implications of the link-layer agnostic operation of IP handover control on handover performance, having as a reference the Fast Mobile IPv6 protocol. We show that the behavior of the protocol (i.e., whether a reactive or proactive operation will be executed) is highly dependent on the timely availability of link layer information. A non-exhaustive list of generic link-layer triggers used for this purpose, as identified by the IEEE 802.21 WG, is also presented. Last, we apply this generic framework to a WLAN environment running Fast Mobile IPv6 and study the improvements in fast handoff support.

The explosion of mobile data communications, the emergence of multitechnology environments with diverse capabilities, the integration of such environments at both terminal and network sides, and the great variety of offered end-user services have completely changed the role of handover management, which nowadays faces the challenge of adaptation to such heterogeneous and multiparametric environments. Traditionally, handover management in cellular networks is carried out by technology-specific mechanisms since it has only involved intratechnology handovers, given the single-tier network landscape and the capabilities of single-mode terminals. By intratechnology handovers we mean handovers between equipment of the same technology [1], such as Universal Mobile Telecommunications System (UMTS) to UMTS or Global System for Mobile Communications (GSM) to GSM. Another important aspect is that such handovers, even after the integration of the IP and telecom worlds that brought the IP layer to the terminal side, are transparent to the IP layer and still only involve layer 2 mechanisms (i.e., radio-technology-specific mechanisms). Therefore, intratechnology handover in such networks is in some way synonymous with the notion of layer 2 handover.

In all cases, session continuity and minimal handover disruption time has always been the primary goal of handover management. This concept of *handover seamlessness* is, however, very much dependent on the service being provided. For example, in pure voice networks such as GSM, seamlessness is perceived as delivering the voice service with bounded handover latency in order not to disturb voice conversation. In General Packet Radio Service (GPRS)/UMTS networks offering data services as well, handover seamlessness of a Web session is interpreted as minimizing packet loss without further concerns on additional delays. This target is more easily reached by certain technologies, such as UMTS, that support macrodiversity, that is, the capability of a terminal to send/receive radio frames to/from more than one base stations (BSs) at the same time (a layer 2 capability). Therefore, a mobile terminal can be serviced in parallel by more than one BSs and thus perform soft handover, where no break in radio communication occurs. For other technologies, such as wireless LANs (WLANs), no such capability is supported. The mobile terminal cannot be serviced in parallel by more than one access point (AP) — the BS equivalent in WLAN terminology — and therefore has to break its communication with its current AP before establishing a connection with a new one. This break in communication is from a layer 2 perspective.

The need for integration of this heterogeneous network environment and the emergence of multimode terminals have placed extra requirements on handover management. It is common sense that this functionality needs to be "pushed" to the IP layer, which is generic enough and serves as the rendezvous point for all underlying technologies. Therefore, handovers between APs of different types (intertechnology handovers) are better treated at the IP layer, the lingua franca of communication protocols. In addition to this well accepted decision of the research community, the need for migration of technology-specific core infrastructures toward all-IP networks has been identified, and great effort is also put in this direction [2, 3]. This makes more evident the fact that in the near future, intertechnology handoffs will be handled at the IP layer since the IP gateway of such evolved networks will not be a distant network entity but rather collocated with the radio-specific gateway, for example, the serving GPRS support node (SGSN) or radio network controller (RNC) in the UMTS network.

In this context, handoff techniques in conjunction with mobility management mechanisms *at the IP layer* will be the main research topic within the broader handover management field. Certainly, such techniques are not coming to replace the well-performing soft handover capabilities of certain radio technologies but have a different application area. For exam-

0890-8044/05/$20.00 © 2005 IEEE

ple, there is no gain in removing macrodiversity functionality from interconnected UMTS radio access networks and using IP techniques instead. IP handoff mechanisms will, however, be triggered more often since users' movement in this evolved environment will result in more frequent changes of their IP path. In other words, inter-access router (AR) handovers will be common in the mobility of users. The support of *seamless IP handovers* becomes an even more challenging task in cases where radio communication is lost when switching between APs (e.g., switch from UMTS to WLAN radio communication or handover between WLAN APs belonging to different IP subnets). Several techniques can be employed for the IP handover mechanism not to simply react to the restoration of radio communication, but to proactively take actions and establish state information in the involved ARs. The difficulty of this task lies in the inherent nature of the IP layer, which assumes no cooperation with underlying technologies and consequently is not informed of impending handovers.

In the following, we attempt to obviate the need for such cooperation between layers. More specifically, we provide a thorough analysis of the Mobile IPv6 (MIPv6) [4] and Fast MIPv6 [5] protocol operation, focusing on their contributing factors to handover delay. We further examine to what degree the enhancements offered by Fast MIPv6 operation toward seamless handover support are dependent on the timely availability of handoff-related information. A generic framework is presented where link layer triggers assist in the IP handover preparation and execution phases targeted at optimal synchronization of layer 2 and 3 handovers. Then the application of such a framework in a WLAN network environment running Fast MIPv6 is further studied. Lastly, we conclude the article.

Enabling Mobility Support in IPv6 Networks

Mobile IP has been widely accepted as the most appropriate protocol for addressing the needs of IP mobility management in future wireless mobile networks. It suffers, though, from several well-known weaknesses. As further elaborated below, the main weakness is the introduced latency in restoring the communication path to the host's new point of attachment. It can be said that MIP is a *path update* protocol rather than a *handover management* protocol. Note that the handover functionality is not to be confused with the path update functionality. The former involves a time-critical operation that "locally" redirects packets to the host's new location for preserving transparency to running services, whereas the latter reestablishes the path after the handoff has been performed and IP connectivity regained [6].

Handover management is responsible for maintaining the active sessions of the mobile host (MH) as the latter moves across the coverage area of various APs. Here, we are concerned with handovers that result in a change of the network-layer (IP) connectivity of the host. Note, however, that this is not always the case since a change in link-layer connectivity does not necessarily result in an IP handover. A handover control protocol should ensure that handoffs are fast and smooth; that is, they should be performed without significant delays and without loss of packets; a requirement that is also dependent on the provided service. As handover delay we define the time between the delivery of the last packet to the host from the old AP and the delivery of the first packet from the new AP. Among the most adopted handover schemes are the establishment of temporary tunnels between the old and new APs, and the bicasting of packets to both APs. Both of these schemes might also employ buffering techniques.

Regardless of the handover mechanism used within an IP access network, it should be stressed that the handover performance is also highly dependent on the underlying radio technology and the information the latter provides to the IP layer. For example, a radio layer that provides indications to the IP layer of an impending handover enables the preparation and possibly completion of the IP handover before the MH loses its layer 2 (L2) connectivity. Alternatively, a handover decision can be solely based on layer 3 (L3) indications and completely independent of the L2 technology, resulting in greater handover delay and a greater possibility of service disruption. It is evident that each approach has its pros and cons; thus, it remains to be decided within the standardization bodies to what degree these two layers should be coupled and synchronized. All these issues are discussed in detail below.

Mobile IPv6: Does It Suffice for Seamless Handover Support?

Mobile IPv6 [4] comprises the Internet Engineering Task Force (IETF) solution to handling mobility of hosts in IPv6 networks. It extends the basic IPv6 functionality by means of header extensions rather than being built on top of it, as is the case with MIPv4. Its fundamental principle is that an MH should use two IP addresses: a permanent address, the home address, assigned to the host and acting as its global identifier, and a temporary addres, the care-of address (CoA), providing the host's actual location. An MH, while attached to its home network, is able to receive packets destined to its home address, forwarded by means of conventional IP routing mechanisms. When the host crosses the boundaries of its current serving network, movement detection is performed in order to identify its new point of attachment and further acquire a new CoA (nCoA). In its simplest form, movement detection consists of the receipt of a router advertisement from an AR other than the one currently serving the host. The new CoA is obtained through stateful or stateless IPv6 address autoconfiguration mechanisms; however, the time needed for the autoconfiguration procedure and binding management to complete make MIP operation inadequate for fast handoff support.

Once configured with a CoA, the MH needs to send a binding update (BU) message to its home agent (HA) to register this "temporary" address, which is referred to as the primary CoA. Only one CoA should be registered with the HA at a time, although the MH may have formed more than one CoA. According to typical MIP operation, the correspondent host (CH) addresses the MH at the latter's home address, and consequently does not need to implement the specific IPv6 extensions that actually form MIPv6.

In the opposite case — when the CHs are augmented with MIPv6 functionality — route optimization can be used for direct delivery of packets to the MH without intervention of the HA. Keep in mind that route optimization is integrated into MIPv6 and does not constitute a set of optional extensions. The MH transmits BUs to its communicating CHs for the latter to associate the MH's home address with a CoA. Data packets are not encapsulated for delivery to the MH, as is the case in MIPv4, but instead an IPv6 routing header is used for this purpose. These packets have as their destination address the MH's CoA. The home address information, required to preserve transparency to upper layers and ensure session continuity, is included in the routing header. In the reverse direction, packets have as source address the host's CoA, while the home address is included in the newly defined home address destination option [4].

Before proceeding, we should further explain why stateless address autoconfiguration adds delay to the handoff procedure. As defined in [7], the stateless address autoconfigura-

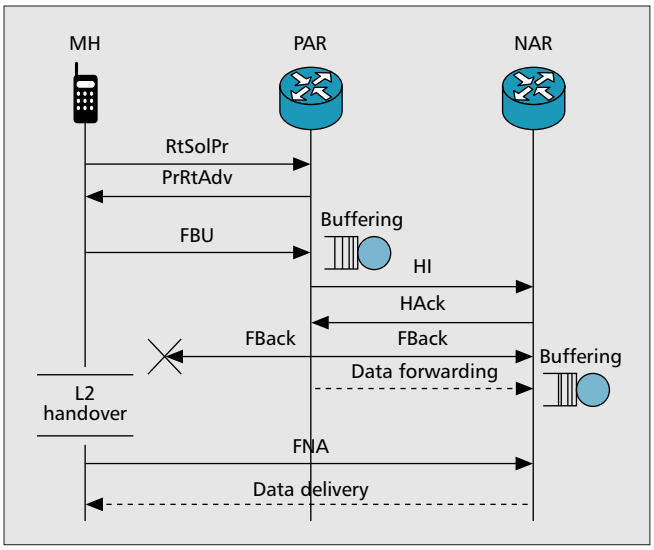


Figure 1. *FMIPv6 protocol operation: predictive behavior.*

tion mechanism allows a host to generate its own addresses in the following way. ARs advertise prefixes that identify the subnet(s) associated with a link, while hosts generate an *interface identifier*, used as a suffix that uniquely identifies an interface on each subnet. A global address is formed by combining the two. The formation of an address must be followed by the duplicate address detection (DAD) procedure in order to avoid address duplication on links. First, the link-local address of the host is generated and tested for its uniqueness. This address, as indicated by its name, serves the communication of hosts on the same link, and is formed by appending the interface identifier to the link-local prefix (FE80::). After the assignment of the link-local address to the interface, global addresses are formed and should also be tested for uniqueness. This is needed since all of an interface's unicast addresses are not necessarily generated from the same interface identifier. In brief, the address autoconfiguration is composed of the following steps:

1 The host generates a link-local address for its interface on a link. When in handoff, it can use the same interface identifier used in the previous link.
2 It then performs DAD to verify the uniqueness of this address (i.e., the interface identifier on the new link).
3 It uses the prefix(es) advertised by routers for forming a global address to be able to communicate with hosts other than the neighboring ones. DAD is also needed here for each generated global address.

The most time-consuming procedure is DAD execution, not to mention the resource consumption it causes on wireless links. In particular, during DAD the host transmits a neighbor solicitation for the tentative link-local address and waits for RetransTimer ms [8] till it considers the address unique. DAD only fails if in the meantime the host receives a neighbor advertisement for the same address, meaning that another host is using the questioned address, or if another host is in the process of performing DAD for the same address and has also transmitted a neighbor solicitation. It is therefore deduced that at least a link-wide round-trip is needed for performing DAD while 2.5–3 round-trips are required in total for the whole autoconfiguration procedure if router discovery (step 3) is performed in the sequence.

In addition to the autoconfiguration delay, the handoff delay is further increased by the time needed for the BU procedure to complete; until then, all packets addressed to the host's previous CoA are lost. IETF has proposed the Fast MIPv6 protocol for eliminating the aforementioned autoconfiguration delay while also decorrelating the binding management delay from handoff delay. The protocol, however, performs satisfactory with respect to the achieved handoff disruption time in the presence of link-layer triggers, as elaborated on in the following section.

Fast Handovers in Mobile IPv6

Fast MIPv6 (FMIPv6) [5] comes to address the following problem: how to allow an MH to send packets as soon as it detects a new subnet link, and how to deliver packets to an MH as soon as its attachment is detected by the new AR. In other words, FMIP's primary aim is to eliminate the factors of delay introduced by the address autoconfiguration procedure. It achieves this by informing the MH of the new AR's advertised prefix, and validating the prospective nCoA from any duplication on the new link prior to the host's movement.

The MH is already configured with the new address at the time it attaches to the new link. Typically, it can start sending packets in the uplink direction, setting the new address as the source address of these packets. In the downlink direction, a factor of delay is introduced before the new AR (nAR) can start delivering packets to the host. The nAR needs to perform neighbor discovery as soon as it receives packets for a host, in order to detect its presence and resolve its link layer address. This operation results in considerable delay that may last multiple seconds. In order to circumvent this delay, the FMIPv6 procedure requires that an MH announce its attachment through a fast neighbor advertisement (FNA) message that allows nAR to consider it reachable.

Although the host quickly regains IP connectivity as described above, it cannot maintain its active sessions with communicating CHs due to the following:

- The MH cannot start sending packets to CHs setting as the source address the new CoA prior to sending a BU to them; the CHs will drop these packets (CHs drop received packets that reveal a binding *home address–CoA* not present in their caches).
- The MH will not be able to receive packets from CHs at its new address until the CHs update their caches. Therefore, for this period of time downstream packets sent to the old address will be lost.

These two problems are basically addressed by setting up a bidirectional tunnel between the old AR and the MH at its nCoA, and requiring both oAR and nAR to buffer traffic during handover execution. The tunnel remains active until the MH completes the BU with its communicating hosts. To CHs, the MH is located at the old subnet; the old path is temporarily extended with the branch *old AR–nCoA of host* to allow communication to continue during the IP handoff transition period. The full path is reestablished when the BU procedure completes.

In brief, the operation of the protocol (Fig. 1) is as follows: the host sends a router solicitation for proxy (RtSolPr) message to its default AR so as to obtain information (e.g., prefix, link layer address) related to available APs. The host has possibly discovered other APs by means of layer 2 methods (radio channel scanning). The AR serving the user responds with a proxy router advertisement (PrRtAdv) containing the requested information for another AR and thus allowing the MH to perform address autoconfiguration as if it had already migrated to a new link. The host, after formulating a prospective new CoA, sends a fast BU (FBU) to its AR instructing the tunneling of packets addressed to its old CoA (oCoA) toward its nCoA. The AR currently serving the host (referred to as old AR, oAR) starts buffering newly coming packets with oCoA as their destination and exchanges handover initiate (HI) and handover acknowledge (HAck) messages with the

Handover supporting mechanism	Figure reference	Time period when packets reaching oAR are lost	Handoff disruption time (no packet reception at MH)
MIPv6	Figure 2a	$D_{L2} + D_{DAD} + D_{RD} + D_{MH\text{-}CH} + D_{CH\text{-}oAR}$	$D_{L2} + D_{DAD} + D_{RD} + D_{MIPR}$
FMIPv6 (completely reactive case) *No information available*	Figure 2b	$D_{L2} + D_{RD} + D_{MH\text{-}oAR}$	$D_{L2} + D_{RD} + D_{FMIP}$
FMIPv6 (reactive case) *nAR information (L2 address, prefix) available*	Figure 2c	$D_{L2} + D_{MH\text{-}oAR}$	$D_{L2} + D_{FMIP}$
FMIPv6 (FBack receipt at new link) *nAR + impending handover information available*	Figure 2d	No loss, unless buffers overflow at oAR or nAR	Max (D_{FMIP} — $2D_{MH\text{-}oAR}$, $D_{diff} + D_{L2} + 2D_{MH\text{-}nAR}$)
FMIPv6 (FBack receipt at old link) *nAR + impending handover information available*	Figure 2e	No loss, unless buffers overflow at oAR or nAR	(DFMIP — $2D_{MH\text{-}oAR}$) + $D_{L3\text{-}L2} + D_{L2} + 2D_{MH\text{-}nAR}$

Table 1. *Effects of MIPv6 and FMIPv6 on packet loss and handoff disruption time.*

nAR to initiate the process of the MH's handover. This HI/HAck message exchange also serves the validation (DAD) of the nCoA already formed by the host. The oAR responds to the MH with a fast binding acknowledge (FBack) message on both links (old and new) and starts the tunneling of buffered and arriving data toward the MH's nCoA. The MH, as soon it attaches on the new link, transmits an FNA to inform the nAR of its presence. Packets from this point on are delivered to the MH with FBack most probably the first packet on the new link.

It should be stressed here that the sequence of messages described above, corresponding to a predictive behavior of the protocol, poses some requirements for the *information* made known to the MH and the *timing* of its availability. This information includes:

- The MH becoming aware of the impending handover, prior to the L2 handover execution, to have enough time to send the FBU message
- The MH becoming aware of the nAR's L2 address to be able to send the FNA immediately after attaching to the new link without the need for router discovery, and of the new subnet's prefix to form a prospective CoA and request from the oAR the redirection of packets to this new address

If part of this information is not made available to the IP layer prior to the layer 2 handover, the fast handover protocol reverts to reactive behavior where the FBU is sent from the new link (encapsulated in the FNA message). Moreover, additional delay may occur due to the possible need for router discovery [9].

A qualitative analysis showing the effects of MIPv6 and FMIPv6, with or without expedited information, on the handoff disruption time (fourth column) and the time period when packets reaching oAR are lost (third column) is presented in Table 1. For the analysis, the following assumptions were made:

- The one-way delays between two nodes in both directions are the same (e.g., the propagation delays from MH to oAR and from oAR to MH are equal).
- When FBU is sent from the old link, all downstream packets sent to the MH from oAR prior to the receipt of the FBU are delivered to the MH (i.e., the MH does not execute L2 handover immediately after the transmission of the FBU).
- Duplicate address detection and router discovery are performed in sequence. Router discovery also includes the DAD delay for the global address.

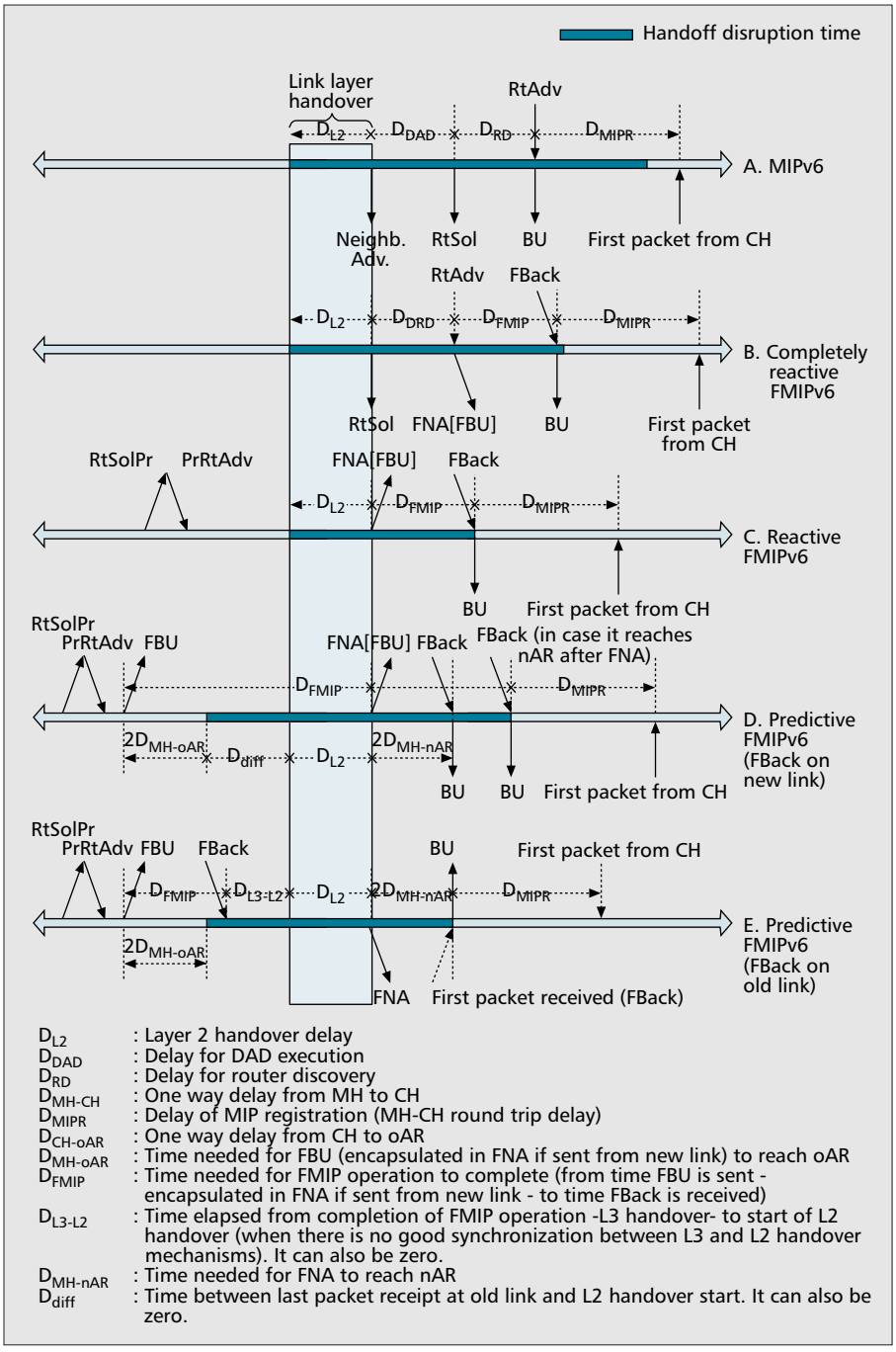
Link Layer Assistance for Fast MIP Handovers

In the previous sections, we identified the need for link-layer triggers (i.e., events fired at the link layer module and communicated to the network layer modules) to aid the IP handover preparation and execution, and illustrated how cross-layer information exchange speeds up the handover procedure. In this direction, great effort on defining a generic interface to facilitate such event delivery is allocated nowadays by standards organizations such as the IETF and IEEE. Link layer triggers are delivered to network layer modules as events for reporting changes in respect to link and physical layer conditions. In addition, they provide indications regarding the status of the radio channel. These indications take the form of (threshold crossing) measurement reports for particular parameters (e.g., signal strength, interference status, error rate) that in general characterize the quality of the wireless medium. In this way, the network layer is informed that certain events (e.g., link establishment or disconnection) have taken place at particular moments, and consequently can execute the entire handover procedure in a more timely fashion and in synchronization with the layer 2 handover. More accurate estimations can be performed in respect to the anticipation of a handover, since the report to the IP layer of events such as radio conditions (e.g., the progressive deterioration in quality of the signal received) may be utilized for intelligent handover prediction. Table 2 presents a preliminary list, extracted from [10], of link layer triggers based on the IEEE 802 suite.

The MIP specification, dealing with a simple network layer mobility supporting mechanism, has made no assumptions as to the technology used for the underlying layer, resulting in both link and network layers operating independent of one another. This constitutes one of the main shortcomings of MIPv6 that render it inappropriate for fast handover support. On the other hand, the FMIPv6 specification and proposed optimizations over simple MIPv6 operation are clearly based on reliable prediction of handover that enables proactive con-

figuration of the involved nodes. The availability of triggers and furthermore the exact time they are fired — dependent on protocol intelligence — influences handover takeoff and actually determines whether proactive, reactive, or no fast handover optimizations will eventually take place. We stress here that the absence of accurate prediction (e.g., very early or erroneous handover detection) may significantly undermine the seamlessness promised by the protocol. The decision of when the MH sends the FBU message is completely an issue for the IP module to handle and should be based on a kind of *link quality crosses threshold* or *link going down* event, with reference to Table 2. However, the handover process will not make the correct decision unless adequate and timely delivered link layer information becomes available to it.

Last but not least, it is noted that an even more optimal case than simple reporting of primitive link-layer information to the network layer would be for the IP module itself to have execution control of certain L2 handover-related actions. As suggested in [11], this means an L2 interface becoming available to the IP module for ordering the execution of such actions. In the context of the IEEE 802.11 WLAN technology, the IP module could have the option, for example, to request radio channel scanning. It is further envisioned that the IP module will be enabled to set up or tear down a link layer connection with a particular peer (i.e., an AP) on demand. This issue becomes very important when the handover decision is based, apart from the radio signal characteristics, on diverse factors ranging from QoS support to security (authorization, roaming agreements) and accounting issues, the MH's velocity, or even the end user's personal preferences.



D_{L2} : Layer 2 handover delay
D_{DAD} : Delay for DAD execution
D_{RD} : Delay for router discovery
D_{MH-CH} : One way delay from MH to CH
D_{MIPR} : Delay of MIP registration (MH-CH round trip delay)
D_{CH-oAR} : One way delay from CH to oAR
D_{MH-oAR} : Time needed for FBU (encapsulated in FNA if sent from new link) to reach oAR
D_{FMIP} : Time needed for FMIP operation to complete (from time FBU is sent - encapsulated in FNA if sent from new link - to time FBack is received)
D_{L3-L2} : Time elapsed from completion of FMIP operation -L3 handover- to start of L2 handover (when there is no good synchronization between L3 and L2 handover mechanisms). It can also be zero.
D_{MH-nAR} : Time needed for FNA to reach nAR
D_{diff} : Time between last packet receipt at old link and L2 handover start. It can also be zero.

Figure 2. *Handoff disruption time in MIPv6 and FMIPv6 depending on availability and timing of handover-related information at the IP layer.*

A WLAN Case Study

In this section we investigate the application of the above analysis in a more realistic scenario involving a MH's fast handover when IEEE 802.11 WLAN standard [12] is utilized for the radio medium access. The standard foresees that an MH is obliged to perform particular tasks in order to hand over from one AP to another located in its vicinity. Typically the WLAN link layer handover comprises three distinct phases; discovery, authentication, and association, briefly described below.

Discovery: The MH is scanning the wireless medium in order to find a potential AP with which it may establish a radio link. The scanning procedure is performed with the MH locking on a radio channel where it passively waits for AP beacons. Alternatively, instead of waiting for beacons to be transmitted, the MH may actively request that each AP reply to probe messages sent for this purpose. In either case, the evaluation of each of the received radio signals at the mobile side finalizes the procedure or instructs its continuation to a next available channel until an AP is found with which communication may be established. The trigger for the initialization of the scanning procedure and the criteria indicating its successful outcome are implementation dependent.

Authentication: The station and AP selected by the scanning phase exchange appropriate management messages (in a two-handshake manner) for the station to authenticate itself, if required, according to an authentication algorithm. Note that

Link layer trigger	Description
Link up	The L3 process may start sending packets as a link has been established.
Link down	This indicates that the link cannot be used for data transmission any longer.
Link quality crosses threshold	The link quality has remained under or over a preconfigured threshold for a certain period of time so that the network layer may start preparing for a handover. It is not implied that the handover should start immediately.
Link going down	A link down event will be fired in the near future (or at a certain time), so the network layer must initiate the handover procedure.
Link going up	This trigger may be used for cases where the establishment of radio communication lasts long enough to influence network layer decisions such as network detection and selection (e.g., for avoiding the selection of a network).
Better signal quality AP available	The trigger specifies that the link layer receives radio signals with better link quality from a different AP than the one currently connected.

Table 2. *Link layer triggers [10].*

no authentication is performed if the AP supports *open system* authentication.

Association: The station requests from the AP an association identifier to be used for typical data delivery to the MH. For the case where the APs involved in handover belong to the same extended service set (ESS), that is, they are part of the same data distribution system resulting in the link layer handover to be transparent to the IP layer, the standard foresees that both APs may communicate via a specific-for-this-purpose Inter Access Point Protocol (IAAP [13]) in order to enable the delivery to the new AP, over the common distribution system, of already buffered data at the old AP. However, in case of interdomain movement this is not feasible.

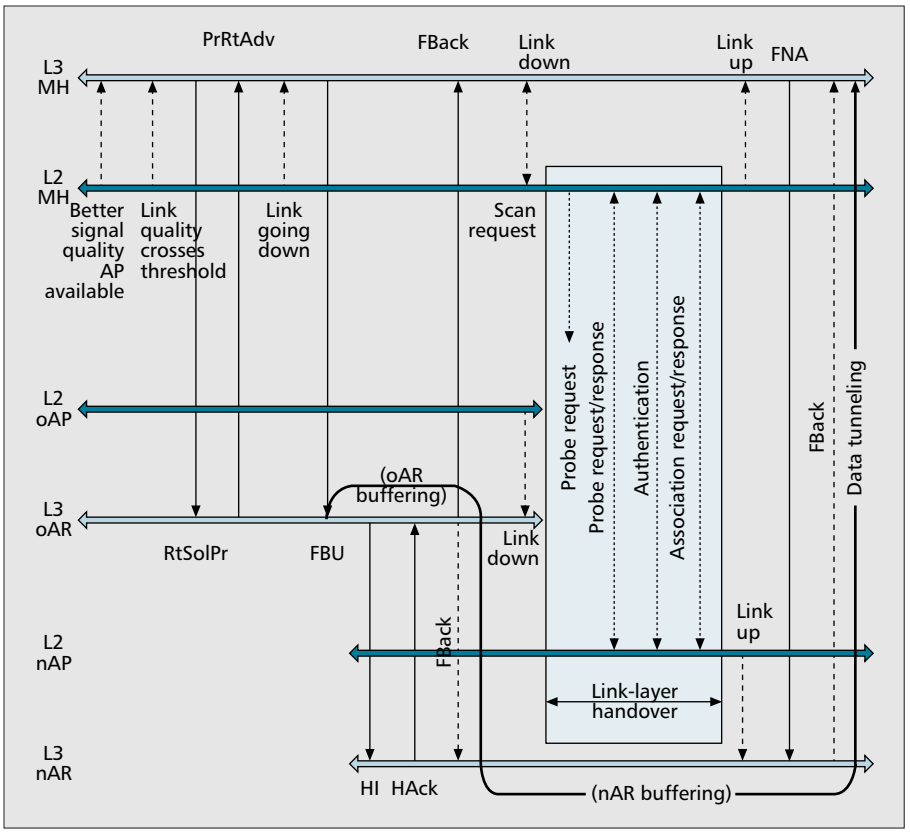
Figure 3 depicts a best case scenario where the complete set of optimizations has been used for both network and link layers. Thus, a predictive FMIPv6 handover is presented, utilizing the IEEE 802.21 specified "link layer trigger" model for the coordination of FMIPv6 handoff message generation. Moreover, *active channel scanning* has been assumed to minimize the delay of waiting for AP beacons. The procedure is initiated (not shown in Fig. 3) by the mobility management module of the host requesting (in terms of registering) link layer reports for certain radio channel parameters. It is considered that radio channel scanning is performed frequently enough (when there is no requirement for real-time data communication, for instance, or during power conservation) so that instead of that the MH is informed about available APs located in its vicinity. This results in frequent deliveries of *Better_Signal_Quality_AP_Available* triggers, containing appropriate information for the APs' identification. The delivery of a *Link_Quality_Crosses_Threshold* trigger to the MH's L3 module, given the parameter selected and the criticality associated with it, is a first-level indication of an anticipated handover. Hence, the handover decision module, based on valid information contained on both *Better_Signal_Quality_AP_Available* and *Link_Quality_Crosses_Threshold* triggers, requests that its default AR (i.e., oAR) proxy a router advertisement on behalf of another AR. These triggers provide an indication that a handover may be executed in the near future. The degree of certainty of an impending handover is high enough to justify triggering of the preparatory procedure required for network layer information collection but not adequate to guarantee handover execution. On the other hand, the delivery of the *Link_Going_Down* trigger provides the required credentials to the network layer in respect to imminent execution of a handover. In other words, it is interpreted as an instruction to start the handover support mechanism.

From this point on, the FMIPv6 procedure is executed, enabling the MH's proactive address auto-configuration, the buffering of arriving packets at oAR, and the establishment of a bidirectional tunnel between the PAR and the MH (PAR <-> nCoA). Special notice should be made with respect to the moment just before the beginning of link layer handover. As shown in Fig. 3, the *Link_Down* trigger is fired by the MH L2 module. We argue that for an advanced mobility management mechanism, the network layer should be enabled to request immediate execution of the scanning phase from the link layer (as part of the primitives supported by it), that is, to instruct execution of the L2 handover. For predictive handover, this request should be made just after reception of the FBack message on the old link. After finalization of the link layer handover, the *Link_Up* trigger is fired, instructing the MH to send the FNA message to the nAR. When the nAR successfully processes the FNA, packets queued at the oAR and nAR during handover execution start flowing from nAR toward the MH.

As aforementioned, the execution of a predictive or reactive FMIPv6 handover depends on the time required to preconfigure the MH and network node. This period is more or less determined by the delivery of the *Link_Going_Down* and *Link_Down* triggers; more specifically, by the selection of the monitored parameters, the choices made for their thresholds (which actually cause each trigger to fire), and the implementation of the decision-making algorithm that results in the execution of the handover management procedure. Obviously, if one of these factors is carelessly dealt with, the MH's handover cannot be anticipated with a satisfactory degree of certainty.

Conclusions

We have presented a thorough analysis of the two representative and well-accepted protocols for IP mobility and fast handover support in future mobile networks, Mobile IPv6 and Fast Mobile IPv6, respectively. After obviating the inadequacy of MIPv6 in achieving seamless handovers, Fast MIPv6 is proposed as a solution to this problem. However, it is clearly shown that the enhancements offered by Fast MIPv6 operation toward seamless handover support are strongly dependent on the timely availability of handoff-related information.



■ Figure 3. *IEEE 802.11 trigger-assisted proactive fast MIPv6 handover.*

To this aim, it is essential that cooperation is established among the network and link layers in order for the latter to assist in IP handoff preparation and selection of the IP handoff execution time. An overview of generic link layer triggers, based on the IEEE 802 suite, is also given; such triggers can be used by IP handover modules to increase the degree of certainty for an anticipated handover and to gather all required information. An even more optimal case would be for the IP module to have execution control over certain L2 handover-related actions (e.g., the execution of the scanning phase in WLANs). These enhancements to the cross-layer communication have last been applied and studied in a WLAN environment running Fast MIPv6.

References

[1] J. Manner and M. Kojo, Eds., "Mobility Related Terminology," RFC 3753, June 2004.
[2] 3GPP TR 22.978 v0.4.0, "All-IP Network (AIPN) Feasibility Study," June 2004.
[3] I. Guardini, P. D'Urso, and P. Fasano, "The Role of Internet Technology in Future Mobile Data Systems," *IEEE Commun. Mag.*, vol. 38, no. 11, Nov. 2000.
[4] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," RFC 3775, June 2004.
[5] R. Koodli (Ed.), "Fast Handovers for Mobile IPv6," Internet draft, draft-ietf-mipshop-fast-mipv6-03.txt, Oct. 2004.
[6] A. K. Salkintzis, *Mobile Internet: Enabling Technologies and Services*, CRC Press, 2004.
[7] S. Thomson, T. Narten, and T. Jinmei, "IPv6 Stateless Address Autoconfiguration," Internet draft, draft-ietf-ipv6-rfc2462bis-07.txt, Dec. 2004.
[8] T. Narten *et al.*, "Neighbor Discovery for IP Version 6 (IPv6)," Internet draft, draft-ietf-ipv6-rfc2461bis-02.txt, Feb. 2005.
[9] P. McCann, "Mobile IPv6 Fast Handover for 802.11 Networks," Internet draft, draft-ietf-mipshop-80211fh-03.txt, Oct. 2004.
[10] V. Gupta and D. Johnston, "IEEE 802.21, A Generalized Model for Link Layer Triggers," IEEE 802.21 Media Independent Handoff Working Group, Mar. 2004 mtg. mins., http://www.ieee802.org/handoff/march04_meeting_docs/Generalized_triggers-02.pdf, Mar. 2004.
[11] K. Mitani *et al.*, "Unified L2 Abstraction for L3-Driven Fast Handover," Internet draft, draft-koki-mobopts-l2-abstractions-02.txt, Feb. 2005.
[12] ANSI/IEEE Std. 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.
[13] IEEE Draft 802.1f/D5, "Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation," Jan. 2003.

Biographies

LILA V. DIMOPOULOU [M] (lila@telecom.ntua.gr) received a Dipl.-Ing. degree from the National Technical University of Athens (NTUA), Greece, in 2000. Since 2000 she has been working at the Telecommunication Laboratory of NTUA as a Ph.D. student. Her research interests include IP mobility management, mobile Internet access, seamless handoffs in heterogeneous environments, 3G-WLAN interworking, and 4G networks. She is a member of the Technical Chamber of Greece.

GEORGIOS A. LEOLEIS [StM] (gleol@telecom. ntua.gr) He received a Dipl.-Ing. degree from NTUA in 2000. Since 2000 he has been a research associate in the Telecommunications Laboratory of the School of Electrical and Computer Engineering at NTUA and a Ph.D. student in the area of communication networks. His research interests include mobility support in IP and cellular networks, handover support in WLANs, IP multicasting, and MBMS support for UMTS. He is a member of the Technical Chamber of Greece.

IAKOVOS S. VENIERIS [M] (venieris@cs.ntua.gr) received a Dipl. -Ing. degree from the University of Patras,Greece, in 1988, and a Ph.D. degree from NTUA in 1990, all in electrical and computer engineering. During 1991–1992 he was with the National Defense Research Center, Athens, Greece, performing research in the area of telecommunication networks for military applications. From 1992–1994 he was a research associate in the Telecommunications Laboratory of NTUA. In 1994 he became an assistant professor in the Electrical and Computer Engineering Department of NTUA where he is now an associate professor. His research interests are in the fields of broadband communications, Internet, mobile networks, intelligent networks, internetworking, signaling, service creation and control, distributed processing, agents technology, and performance evaluation. He has over 100 publications in the above areas. He has received several national and international awards for academic achievement. He has been exposed to standardss body work and has contributed to ETSI and ITU-T. He has participated in several European Union and national projects dealing with B-ISDN protocols, mobile networks, intelligent networks, ATM switching and access techniques, Intelligent software, and Internet technologies. He is an Associate Editor of *IEEE Communication Letters*, a member of the editorial board of *Computer Communications* (Elsevier), and has been a guest editor for *IEEE Communications Magazine*. He is a reviewer for several journals and has been a member of the Technical Program Committee and session chairman of several international conferences. He is a member of the Technical Chamber of Greece.