1

Mobility Management for All-IP Mobile Networks

Quan Le-Trung

Department of Informatics, University of Oslo, Norway

Paal E. Engelstad

Department of Informatics, University of Oslo, Norway
Simula Research Laboratory, 1325 Lysaker, Norway
Telenor Research and Innovation (R&I), Fornebu, Norway

Tor Skeie
Department of Informatics, University of Oslo, Norway
Simula Research Laboratory, 1325 Lysaker, Norway

Frank Eliassen
Department of Informatics, University of Oslo, Norway
Simula Research Laboratory, 1325 Lysaker, Norway

Amirhosein Taherkordi

Department of Informatics, University of Oslo, Norway

CONTENTS

One of the recent trends in the networking has been concentrated on realizing all-IP mobile networks, together with new Internet applications and services. For the next-generation of all-IP mobile networks, one of the challenging issues is the mobility management. The contribution of this chapter is multi-fold. Firstly, the state of the art in mobility management for all-IP mobile networks is presented, mainly at the network and link layers. Qualitative analysis on positive and negative points among alternative approaches are also discussed and shown. However, the cross network/link-layer relations to control mobility management in all-IP mobile networks are normally specified in a high-level of abstraction. Thus, the second contribution of this chapter is to consider mobility management for all-IP mobile networks spanning into a specific one, the wireless LAN (WLAN) IEEE 802.11 networks, in both network and link layers. Thirdly, different issues in mobility management in all-IP mobile networks spanning into the mobile ad-hoc networks (MANETs) are considered. The reason is that WLANs are 1-hop access networks, while MANETs are multihop access networks. Due to multihop connectivity in MANETs, different problems appear due to the mobility of MANET nodes among different domains of all-IP mobile networks. The discovered problems and the proposed solutions to either reduce partly, or remove completely these discovered problems, are the forth contribution of this chapter. Finally, this chapter ends with conclusions and lay out some research directions in the next-generation of all-IP mobile networks.

1.1 Introduction

All-IP mobile networks are defined as the networks in which IP is employed from a mobile subscriber to the access points (APs) that connect the wireless networks to the Internet [40]. In the fourth-generation (4G) of wireless networks, the goal is to allow users to communicate reliably and cost effectively via any media, at any time, and anywhere [18, 4, 45]. Thus, 4G [1] systems will integrate existing and new wireless networks seamlessly, allowing mobile users to roam globally with no limit to underlying access technologies, such as wireless LANs, wireless PANs, wireless MANs, and cellular systems [18]. The mobility management, therefore, is one of the most challenging issues in 4G all-IP mobile networks.

The mobility management for all-IP mobile networks is originally considered *host-based mobility*, where individual or a small number of mobile users roam across heterogeneous wireless networks. This sub-class is further classified as either *macro-mobility* or *micro-mobility*. The *macro-mobility* is used to manage the mobility of mobile users among wireless networks under the different authorities, and mobile IPv4 (MIPv4) as well as mobile IPv6 (MIPv6)[1] is probably the most well-known IP mobility support protocol. However, handoff[2] latency and signaling overhead in the macro-mobility are the main causes of packet loss and resulting in performance degradation. Thus, numerous methods of minimizing handoff latency and/or signaling overhead have been proposed. In the scope of this book Chapter, only IP-layer (L3) and link-layer (L2) solutions are considered, mostly in the scope of micro-mobility, i.e., the mobility of mobile users among sub-networks managed by the same authority. Security is out of scope of this chapter.

Micro-mobility solutions aim at achieving low latency handoff mechanisms for delay-sensitive or real-time applications, with low to zero packet loss, and to ensure that the signaling overhead is kept to a minimum. Micro-mobility protocols are designed for environments where mobile nodes change their point of attachment to the network so frequently, that the basic MIPv4/v6 protocol tunneling mechanism introduces the unacceptable network overhead, in terms of the increased delay, packet loss, and signaling. Micro-mobility protocols support the handling of local movement, e.g., within a domain, of mobile nodes without interaction with the mobile IP enabled Internet. This has the benefit of reducing the delay and the packet loss during the handoff and eliminating

[1]Though IPv6 is the de facto network protocol in the next-generation all-IP mobile networks, IPv4 and network address translation (NAT) are likely to co-exist with IPv6 for a long time. Thus, mobility management protocols for both IPv4 and IPv6 mobile networks are considered in this book Chapter.

[2]Handoff and handover are the interchangeable terminologies used in this book Chapter. These two terms refer to the process of transferring an ongoing connection if either the source or the destination mobile node of this connection moves from one domain (e.g., home network in MIPv6) to another (e.g., foreign network in MIPv6).

registration between mobile nodes and possibly distant home agents when mobile nodes remain inside their local coverage areas. Micro-mobility solutions can be further classified according to: i) the techniques used to reduce the signaling overhead, e.g., Hierarchical MIPv4/v6 (HMIPv4/v6) [29, 85], cellular IP (CIP) [86], handoff-aware wireless access Internet infrastructure (HAWAII) [78]; ii) the techniques to reduce the handoff latency, e.g., Fast Handover in MIPv4/v6 (FMIPv4/v6) [41, 42], or iii) the hybrid techniques, e.g., Fast Handover for HMIPv6 (FHMIPv6) [39].

However, MIPv4/v6 and its various enhancements basically require modifications in the protocol stack of mobile nodes. On the other hand, recently, in *network-based host mobility*, e.g., Proxy MIPv6 (PMIPv6) [28, 88], the serving network handles the mobility management on behalf of the mobile nodes. Thus, the mobile node is not required to participate in any mobility-related signaling. Therefore, the tunneling overhead and a significant amount of packet exchanges via wireless links can be reduced, resulting in higher performance and lower latency. In the Internet Service Provider (ISP) perspective, the deployment of the mobility management in all-IP mobile networks will be faster and more reliable.

While *host mobility* considers the mobility of individual mobile nodes, *network mobility* (NEMO) considers the mobility of a whole mobile network, e.g., mobile users within a train, an airplane, or a ship. There are recent research extending beyond host mobility to network mobility, e.g., IETF NEMO. However, the NEMO basic support protocol [17] does not address important issues such as route optimization and handoff, causing unacceptable delay and low performance. Further improvements of route optimization, which are border gateway protocol (BGP)-based, have been implemented in Connexion by Boeing [20], and wide-area IP network mobility (WINMO) [33].

Currently, the IEEE is working on a standard, i.e., 802.21 [3], for media-independent handover (MIH) services. IEEE 802.21 enables handover and interoperability between heterogeneous network types, including both 802 and non-802 networks. The key ideas in this standard are: i) providing access network service to take care of link-layer triggers, and allow upper layers use this information timely; ii) providing command service to allow upper layers to send various instructions to the link-layer; and iii) providing an information service as a basis for making more effective handover decisions.

While there are many approaches to mobility management for all-IP mobile networks, there has still been the lack of a complete, up-to-date comparison on the state of the art among these approaches, and this is the first contribution of this Chapter (Section 1.2). In this Section, qualitative analysis on positive and negative points among alternative approaches will be discussed and presented. Additionally, a major direction in reducing the handoff latency in the mobility management of all-IP mobile networks is to use link-layer triggers. However, the cross L3/L2 relations to control the mobility management are normally specified in a high-level abstraction. To map this high-level abstraction into a specific one, in Section 1.3, the mobility management for all-IP

mobile networks spanning into a specific and most popular one, the WLAN IEEE 802.11 domain, is considered in both L3 and L2 layers. This is the second contribution of this book Chapter.

Though WLAN 802.11 is one of the most popular wireless access networks, it only provides 1-hop coverage. In the very near future, mobile nodes will roam across multiple heterogeneous platforms while continuously maintaining connection connectivity. A mobile node may connect to a WLAN, and then move into an area where the coverage from the WLAN does not exist. In this situation, a mobile ad-hoc network is required to extend the 1-hop to n-hop coverage. Thus, one of the wireless network types in 4G systems will be the MANET. The third contribution of this Chapter (cf. Section 1.4) is the mobility management for all-IP mobile networks spanning the MANET domain.

To ensure the self-configured, infrastructureless, and mobility-controlled characteristics of MANETs when accessing to the Internet, in Section 1.5, different required functions need to be provided [46]. These functions include: i) MANET node location determination, ii) Internet gateway discovery, selection, and forwarding strategy, iii) auto-configuration addressing scheme, and iv) handoff control. This Section will also review the existing approaches of each required functions on providing Internet connectivity for MANETs and the mobility management. Additionally, the connection between a MANET node and an Internet gateway (IGW) is multihop. Therefore, there is normally no direct wireless link from this MANET node to the IGW, as in the WLAN. Instead, they are connected via other intermediate nodes. Thus, different problems, e.g., inconsistent context, cascading effect, can happen during the mobility of ad-hoc nodes within a MANET domain if multiple IGWs exist [47, 48] [23]. The required functions, the discovered problems, and the proposed solutions in Section 1.5 is the fourth contribution of this chapter. Section 1.6 presents open issues for IP-based mobility management and some research directions in the mobility management of next-generation all-IP mobile networks Finally, in Section 1.7, this chapter ends with conclusion.

1.2 IP Mobility Management for Mobile Networks

1.2.1 Host Mobility Management

1.2.1.1 Host-based Mobility Management

In the IP environment, when a mobile node moves and attaches itself to another network, it needs to obtain a new IP address. This changing of IP address means that all existing IP connections to the mobile node need to be terminated, then re-established. Thus, there is a request on how to keep on-going IP connections upon the mobility of a mobile node. This Section presents pro-

tocols to ensure the on-going connections to a mobile node when the mobile node changes its points of attachment, and does not consider further characteristics such as security and quality of service (QoS) support.

Host-based Macro Mobility Management:
Mobile IP (MIP) describes a global solution (macro-mobility) that overcomes this problem by using a set of network agents. It does not require any modifications to existing routers or end correspondent nodes [45](Le et al. 2006). With MIP, each mobile node is identified by an address from its home network, regardless of the point of attachment. While a mobile node is away from its home network, it obtains an IP address from the visiting network, which is called foreign agent care-of address (FA CoA) or co-located care-of address (CCoA) in MIPv4, or CCoA in MIPv6, and registers this IP address with a home agent (HA) within its home network. The home agent intercepts any packets destined to the mobile node, and tunnels or explicitly routes (source routing) them to the current location of mobile node. Thus, initiating this indirection requires a timely address reconfiguration procedure and a home network registration. The time taken for a mobile node to configure a new CoA in the visiting network, and the time taken to register with the HA, together constitute the handoff latency. Figures 1.1 and 1.2 show the architectures and operations of MIPv4 and MIPv6, respectively. Figures 1.3 and 1.4 describe the packet sequences in the registration process of MIPv4, using FA CoA and CCoA, respectively. Figure 1.5 shows the packet sequences in the handoff procedure of MIPv6 with the standard IPv6 neighbor discovery [63].

Handoff latency in macro-mobility is the primary cause of packet loss and resulting in performance degradation, especially in the case of reliable end-to-end communication. As a result, numerous methods of minimizing the handoff latency have been proposed in the literature. The proposed schemes can be broadly classified into: i) those that operate above the IP layer such as TCP-Migrate [84] and pTCP [30]; and ii) those that operate at the IP layer. In general, the solutions that operate at the IP layer are regarded as being more suitable as they do not violate any of the basic Internet design principles, and more importantly because they do not require any changes to the protocols at the correspondent nodes [12]. The IP layer solutions are presented and discussed in the next sub-Section, mostly in the micro-mobility scope.

Host-based Micro Mobility Management:
Micro-mobility protocols support local movement, e.g., within a domain, of mobile nodes without the interaction with the MIP-enabled Internet, and thus, reducing delay and packet loss during handoff. Several micro-mobility protocols have been proposed, which can be classified as in Figure 1.6. Hierarchical mobility management reduces the performance impact of mobility by handling local migrations locally and hiding them from home agents. In this case, the Internet address known by a home agent no longer reflects the exact attachment point of mobile node. Rather, it represents the address of

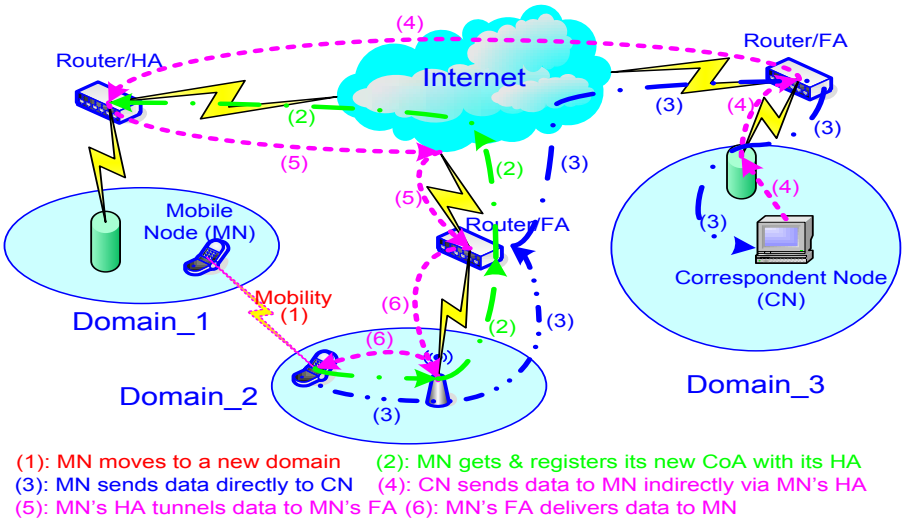

FIGURE 1.1
Architecture and operations of Mobile IPv4 (MIPv4).

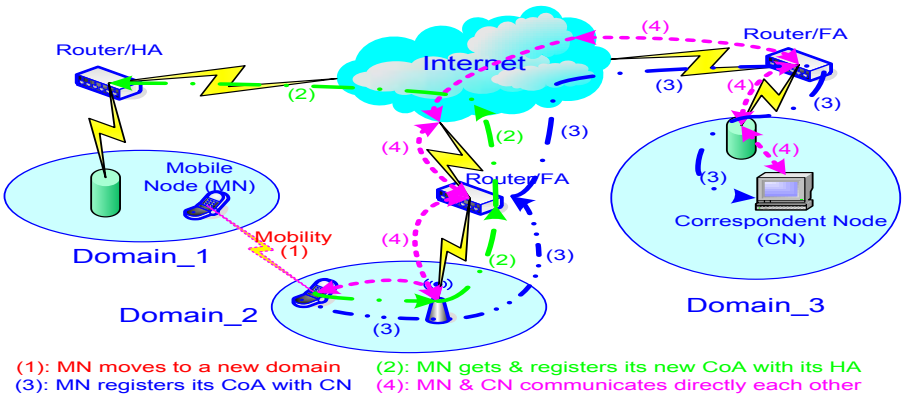

FIGURE 1.2
Architecture and operations of Mobile IPv6 (MIPv6).

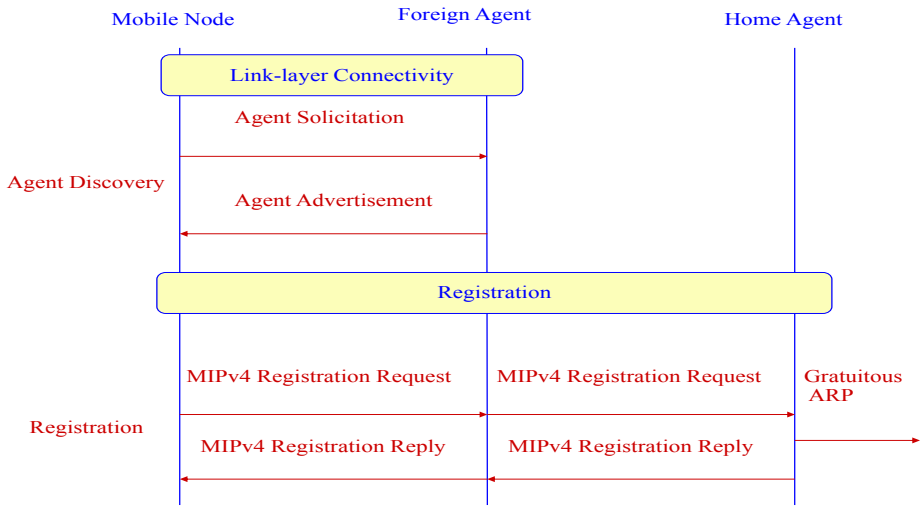

FIGURE 1.3
MIPv4 registration via foreign agent care-of address.

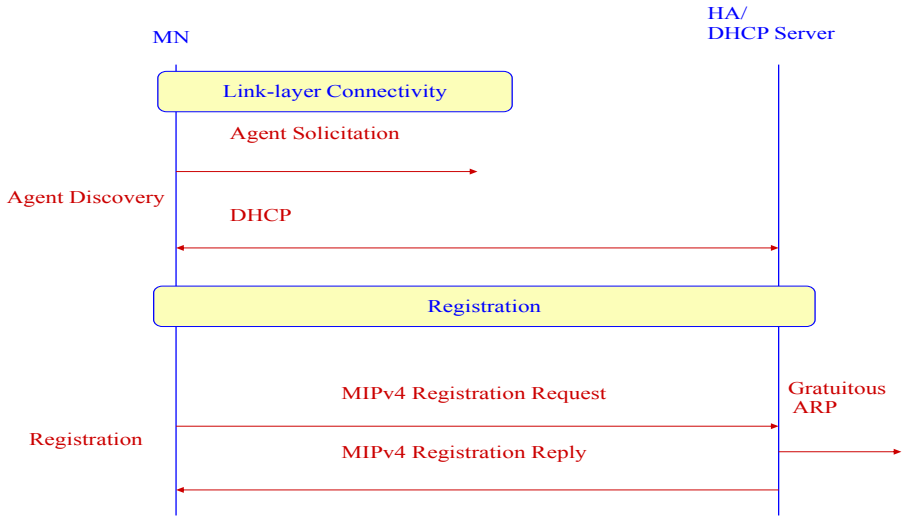

FIGURE 1.4
MIPv4 registration via co-located care-of address.

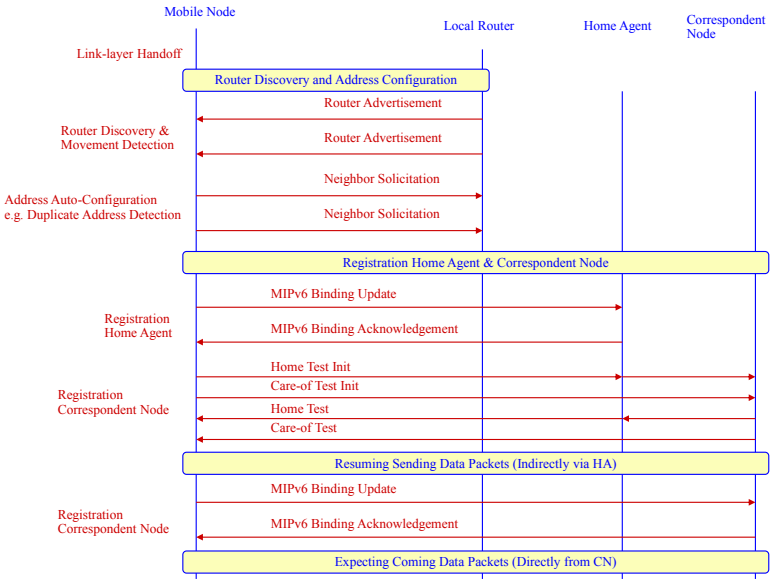

FIGURE 1.5
Handoff procedure with Mobile IPv6 (MIPv6).

a gateway that is common to a potentially large numbers of network access points. When a mobile node moves from one access point to another one, which is reachable through the same gateway, there is no need to inform its home agent. The role of micro-mobility protocols is to ensure that packets arriving at the gateway are forwarded to the appropriate access point. Two styles of hierarchical mobility management are supported by micro-mobility: i) hierarchical tunneling, and ii) mobile-specific routing.

In hierarchical tunneling approach, the location database is maintained in a distributed form by a set of foreign agents in the access network. Each foreign agent reads the original destination address of incoming packet, searching its visitor list for a corresponding entry. If the entry exists, then it contains the address of next lower level foreign agent. The sequence of visitor list entries corresponding to a particular mobile node constitutes the location information of mobile node and determines the route taken by its downlink packets. Entries are created and maintained by registration packets, which are transmitted by mobile nodes. These proposals rely on a tree-like structure of foreign agents. Encapsulated traffic from the home agent is delivered to the root foreign agent. Each foreign agent on the tree decapsulates, then re-encapsulates data packets as they are forwarded down the tree of foreign agents toward the attachment point of mobile node. As a mobile node moves between different access points, location updates are made at the optimal point on the tree, tun-

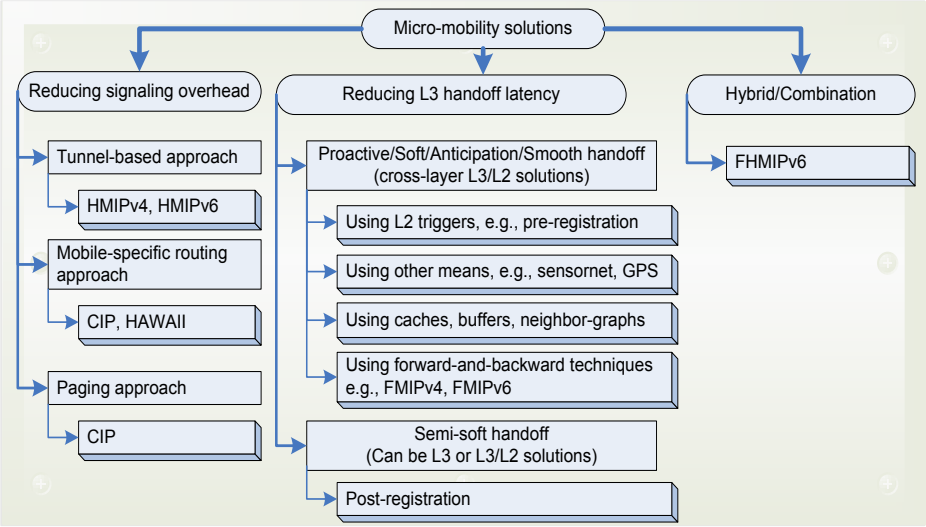

FIGURE 1.6
Classification of host-based micro-mobility management.

neling traffic to the new access point. Examples of micro-mobility protocols, that use hierarchical tunneling, include HMIPv4 [29] and HMIPv6 [85].

Mobile-specific routing approaches avoid the overhead introduced by decapsulation and re-encapsulation schemes, as is the case with hierarchical tunneling approaches. These proposals use routing to forward packets toward an attachment point of mobile node using mobile specific routes. These schemes typically introduce implicit, e.g., based on snooping data, or explicit, signaling to update mobile-specific routes or they are aware that a routing protocol is in use. In the case of cellular IP (CIP), mobile nodes attached to an access network use the IP address of the gateway as their mobile IP care-of address. The gateway decapsulates packets and forwards them toward a base station. Inside the access network, each mobile node is identified by its home address and data packets are routed using mobile-specific routing without tunneling. The routing protocol ensures that packets are delivered to the actual location of mobile node. Examples of micro-mobility protocols that use mobile-specific routing include CIP [86] and HAWAII [78].

Typically, fixed hosts connected to the Internet remain on-line even though most of the time they do not communicate. Mobile nodes connected to the wireless Internet will expect the same service. In order to maintain the location information for routing support, each mobile node requires frequent location updates. Thus, the maintenance of location information consumes the bandwidth and battery power. This signaling overhead can be reduced through the introduction of paging. Mobile nodes are typically powered by the batteries with limited lifetime. This makes it important to save idle mobile nodes

from having to transmit frequent location update packets. This requires explicit support from networking protocols, such as the ability to track location approximately and the ability to page idle mobile nodes. Idle mobile nodes do not have to register if they move within the same paging area. Rather, they only register if they change paging area. A number of micro-mobility protocols, such as CIP and HAWAII, have implemented paging.

Supporting fast and/or low-latency handover, which reduces delay and packet loss during the handoff, is another important attribute of micro-mobility protocols. A number of design choices influence handoff performance including handoff control, buffering and forwarding techniques, radio behavior, movement detection and prediction, coupling and synchronization between IP and radio layer [12]. The low latency handoff proposal for MIPv4 (LLMIPv4) [53] describes two methods of achieving this, namely pre-registration and post-registration. With pre-registration handoff, the mobile node is assisted by the network to perform L3 handoff before it completes the L2 handoff. It uses L2 triggers, which arises as a result of beaconing signals from the network that the mobile node is about to move to, to initiate an IP layer (L3) handoff. Its design however, diverges from the clean separation of L2 and L3 of the base MIPv4 scheme. With post-registration handoff, L2 triggers are used to setup a temporary bi-directional tunnel between the old foreign agent (oFA) and the new foreign agent (nFA). This allows the mobile node to continue using its oFA while performing the registration at the same or later time. A combined method is also possible where, if the pre-registration does not complete in time, the oFA forwards traffic to the nFA using the post-registration method in parallel.

Fast-handover for MIPv4 (FMIPv4) [42], and for MIPv6 (FMIPv6) [41] are similar in concept to the combined method described above, and consists of three phases: i) handover initiation (HI), ii) tunnel establishment, and iii) packet forwarding. Figures 1.7 and 1.8 show the architecture and packet sequences of the handoff procedure in FMIPv6, which is similar as those of FMIPv4, except for the name of agents, e.g., previous access router (pAR) instead of oFA, new access router (nAR) instead of nFA, and signaling packets, e.g., binding update (BU) instead of registration request (RegReq).

The handover initiation (HI) is started by the L2 trigger based on certain policy rule (unspecified by IETF at the time of writing). This is done by the mobile node, which sends a router solicitation proxy (RtSolPr) packet to pAR, indicating that it wishes to perform a fast handoff to a new attachment point. The RtSolPr contains the link-layer address of the new attachment point, which is derived from the beacon packets of nAR. The mobile node will receive, in response, a proxy router advertisement (PrRtAdv) packet from the pAR, with a set of possible responses indicating that the point of attachment is either unknown, or known but connected through the same access router, or is known and specifies the network prefix that the mobile node should use in forming the new CoA. Subsequently, the mobile node sends a fast binding update (F-BU) to the pAR using its newly formed CoA based on the prior

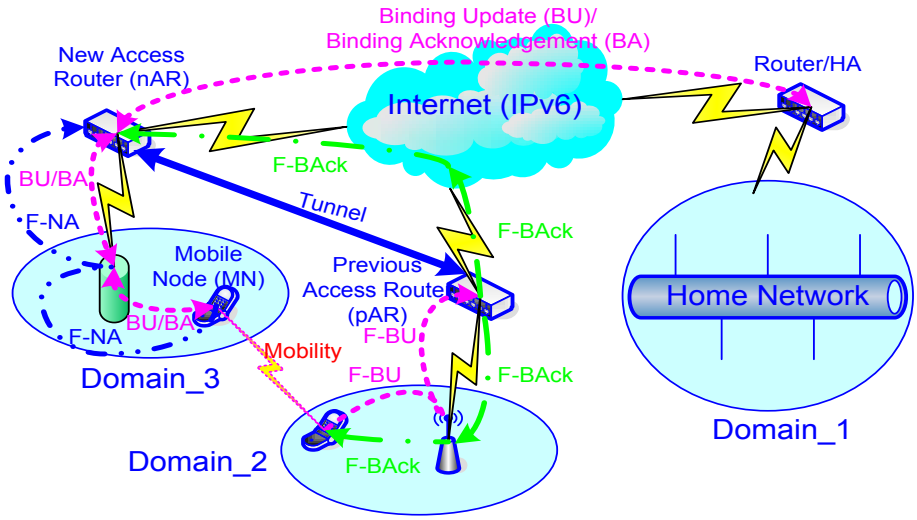

FIGURE 1.7
Architecture and operations of FMIPv6.

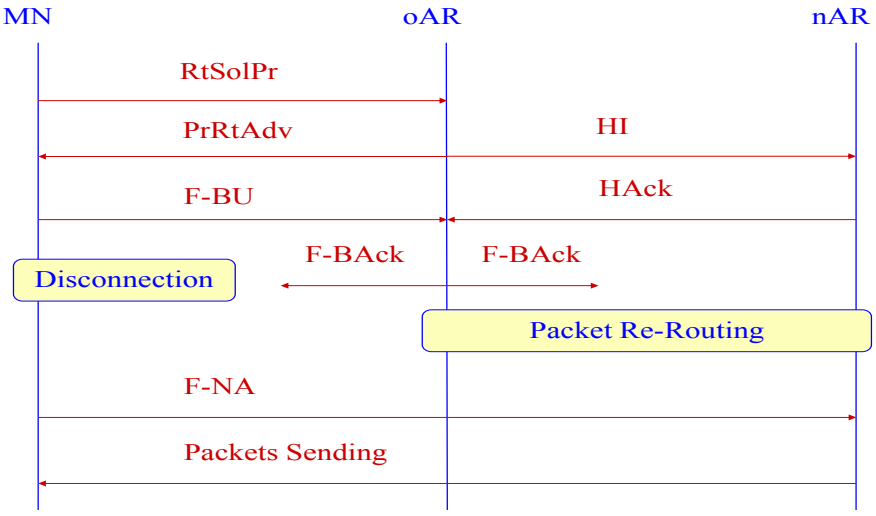

FIGURE 1.8
Predictive mode of FMIPv6 protocol.

PrRtAdv response, as the last packet before the handover is executed. The mobile node receives a fast binding acknowledgement (F-BAck) via either pAR or nAR indicating a successful binding [31].

The tunnel establishment phase creates a tunnel between the nAR and the pAR. To establish a tunnel, the pAR sends a handover initiation (HI) packet containing the requesting CoA and the current CoA of mobile node to the nAR. In response, the pAR receives a handover acknowledgement (HAck) from the nAR. If the new CoA is accepted by the nAR, the pAR sets up a temporary tunnel to the new CoA. Otherwise, the pAR tunnels packets destined for the mobile node to the nAR, which will take care of forwarding packets to the mobile node temporarily. Finally, the forwarding phase of packet is performed to smoothen the handoff until subsequent registration by the mobile node to the home agent is completed. The pAR interacts with the nAR to facilitate the forwarding of packets between them, through the previously established tunnel. The initiation of the forwarding is based on an anticipation timing interval heuristic, i.e., the network anticipates as to when a mobile node is likely to handoff and therefore infers the appropriate packet forwarding moment based on the anticipation timing interval. Such an interval is, however, extremely difficult to generalize, and forwarding too early, or too late will result in packet losses, negating the purpose of packet forwarding. Once arriving at the new access network, the mobile node sends the fast neighbor advertisement (F-NA) packet to initiate the flow of packets (to itself) from the nAR.

Host-based Hybrid Mobility Management:
Hierarchical mobile IPv6 with fast-handover (FHMIPv6) [39] is another attempt to further reduce the overall handoff latency from what fast-handover can offer alone. By combining HMIPv6 with fast-handover, the latency due to the address configuration, and the subsequent home network/agent registration, can be reduced. The mobility anchor point (MAP) can be viewed as the local home agent, and in most cases, it is located closer to the mobile node than the home agent. Therefore, the signaling cost saved is the difference between the round trip time of the mobile node to the MAP and the round trip time between the mobile node to the home agent, assuming that the packet processing time within a network node is insignificant in comparison. This combination requires the minor modification to the standard HMIPv6 protocol and the fast-handover protocol, i.e., relocating the forwarding anchor point from the pAR to the MAP.

An alternative to the scheme of packet forwarding has also been proposed, namely, the simultaneous bindings framework (SBF) [54]. It proposes to reduce packet losses at the mobile node by n-casting packets for a short period to the current location of mobile node and to n-other locations where the mobile node is expected to move to. The n-casting can be carried out by the pAR, the MAP or the HA. The simultaneous binding scheme recognizes the problem of not knowing when the mobile node is likely to move, the timing ambiguity,

and attempts to remove it by careful packet duplication to multiple access networks. It also claims to be able to address the problem associated with rapid back-and-forth movement of mobile nodes between two access routers, called ping-pong movement, by this packet duplication process, as it is not necessary to re-configure the mobile node CoA during ping-pong movement.

The seamless handoff architecture for mobile IP (SMIP) is described in [31]. SMIP is a proactive advanced configuration, where a movement pattern of mobile node (linear/ping-pong) is augmented with a specialized forwarding algorithm when performing a handoff, provides the best result. Unlike FHMIPv6, which uses the L2-trigger, or SBF, which uses the L2-trigger and n-casts packets to multiple destinations, in SMIP, the network uses mobile node location and movement patterns to instruct the mobile node when and how handoff should be carried out. It is a structured approach combining the mobile node location tracking, the movement patterning and the handoff algorithms into an integral unit, providing smarter handoffs. The mobile node initiates a handoff while the network determines the handoff decision. SMIP uses signal strength together with triangulation to track the mobile node location, and determine its movement pattern. The entity which stores the history of the locations and determines the movement pattern, is referred to as the decision engine (DE), is similar to a MAP, and covers a group of access routers and the associated mobile nodes. The DE makes the handoff decision on behalf of the mobile node. Once a decision is made, it is relayed to the access routers (pAR and nAR) and eventually to the mobile node as part of the fast-handover PrRtAdv packet. Note that SMIP is built on the structure of FHMIPv6 and operates similar to that of the mobile node initiated fast handover.

1.2.1.2 Network-based Mobility Management

One of the challenges of host-based host mobility management in Section 1.2.1.1 is the requirement of changes in the IP stack of mobile nodes. These changes limit the broad usage even if the modifications are small. This sub-Section describes existing and on-going work on network-based host mobility management, which enables IP mobility for a mobile node without requiring its participation in any mobility-related signaling. The network is responsible for managing IP mobility on behalf of the mobile node. The mobility entities in the network are responsible for tracking the movements of the mobile node and initiating the required mobility signaling on the behalf of mobile node.

The scope of this Section is local mobility[3], which is restricted to providing IP mobility management for mobile nodes within an access network. While some mobile node involvement is necessary and expected for generic mobility functions such as movement detection and to inform the mobile access gateway (MAG) about mobile node movement, no specific mobile-node-to-network

[3]The local mobility has the same meaning as the micro host-based host mobility in Section 1.2.1.1. We use this terminology as it is defined in the IETF NETLMM charter.

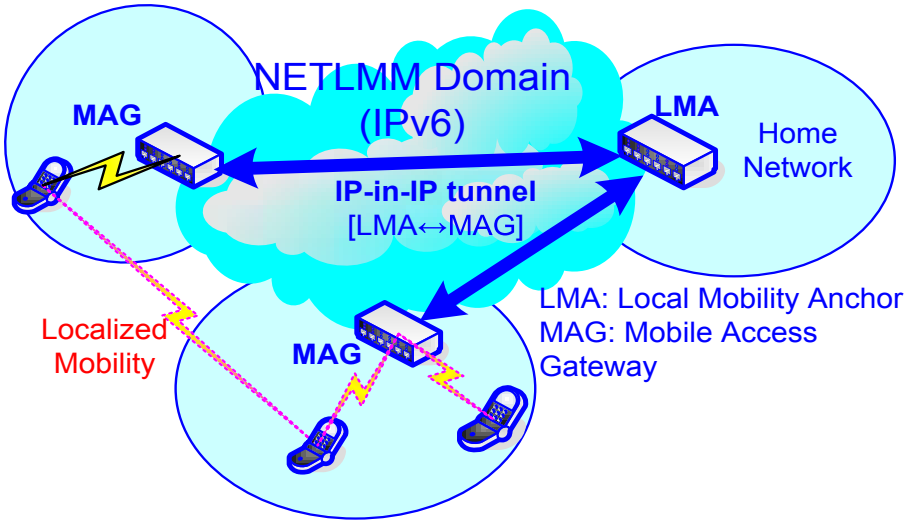

FIGURE 1.9
Proxy Mobile IPv6 (PMIPv6) architecture.

protocol will be required for localized mobility management itself. The mobile node stack involvement in mobility management is thereby limited to generic mobility functions at the IP layer, and no specialized localized mobility management software is required.

Proxy mobile IPv6 (PMIPv6) [28] supports mobility for IPv6 mobile nodes without the involvement of IPv6 mobile nodes, by extending MIPv6 signaling messages between a network node, called MAG in PMIPv6, and a home agent, called local mobility anchor (LMA) in PMIPv6. The MAG in the network performs the signaling with the LMA and does the mobility management on behalf of the mobile node attached to the network. Figures 10-11 show an operation overview and handoff procedure of PMIPv6, respectively. Once a mobile node (MN) enters its PMIPv6 domain, see Figure 1.9, the serving network ensures that the MN is always on its home network and can obtain its home-of-address (HoA) on any access network. That is, the serving network assigns a unique home network prefix to each MN, and conceptually this prefix always follows the MN wherever it moves within a PMIPv6 domain. From the perspective of the MN, the entire PMIPv6 domain appears as its home network. Accordingly, it is needless to configure the care-of-address (CoA) at the MN.

The new principal functional entities of PMIPv6 are the MAG and LMA. The MAG typically runs on the access router (AR). The main role of the MAG is to detect the MN movements and initiate mobility-related signaling with the MN LMA on behalf of the MN. In addition, the MAG establishes a tunnel with the LMA for enabling the MN to use an address from its home network

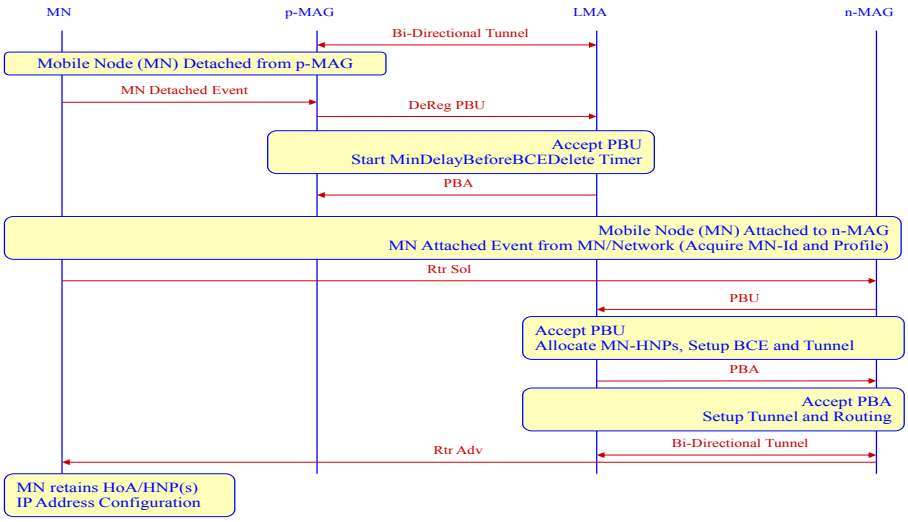

FIGURE 1.10
Handoff procedure with PMIPv6.

prefix and emulates the MN home network on the access network for each MN. The main role of the LMA is to maintain reachability to the MN address while it moves around within a PMIPv6 domain, and the LMA includes a binding cache entry for each currently registered MN. Figure 1.10 shows the signaling call flow for the mobile node handoff from the previously attached mobile access gateway (p-MAG) to the newly attached mobile access gateway (n-MAG). After obtaining the initial address configuration in the PMIPv6 domain, if the mobile node changes its point of attachment, the p-MAG will detect the mobile node detachment from the link. It will signal the LMA and will remove the binding and routing state for that mobile node. The LMA, upon receiving this request, will identify the corresponding mobility session for which the request was received, and accepts the request after which it waits for a certain amount of time to allow the mobile access gateway on the new link to update the binding.

However, if it does not receive any proxy binding update (PBU) message within the given amount of time, it will delete the binding cache entry. The n-MAG, upon detecting the mobile node on its access link, will signal the LMA to update the binding state. Upon accepting this PBU message, the LMA sends a proxy binding acknowledgement (PBA) message including the MN home network prefix(es) (MN-HNPs). It also creates the Binding Cache Entry (BCE) and sets up its endpoint of the bi-directional tunnel to the n-MAG. After completion of the signaling, the serving n-MAG will send the router advertisement (RtrAdv) containing the MN-HNPs, and this will ensure the mobile node will not detect any change with respect to the layer-3

attachment of its interface. Note that the router solicitation (RtrSol) message from the mobile node may arrive at any time after the MN attachment and has no strict ordering relation with the other messages in the call flow. While PMIPv6 requires only IPv6 mobile nodes and IPv6 transport network between the mobility entities, it is also reasonable to expect the same mobility infrastructure in the PMIPv6 domain to provide mobility to the mobile nodes operating in IPv4, or in dual IPv4/IPv6 mode, and whether the transport network is IPv4 or IPv6 network. This is due to the transition from IPv4 to IPv6 is a long process and during this period of transition, both the IPv4/IPv6 protocols will be enabled over the same network infrastructure. [88] describes the IPv4 support in PMIPv6 (IPv4/PMIPv6) for two scenarios: i) IPv4 home address mobility support, and ii) IPv4 transport network support.

The IPv4 home address mobility support essentially enables an IPv4 mobile node in a PMIPv6 domain to obtain IPv4 home address configuration for its attached interfaces and be able to retain that address configuration even after performing an handoff anywhere within that PMIPv6 domain. The mobile node on the access link using any of the standard IPv4 address configuration mechanisms supported on that access link, such as dynamic host configuration protocol (DHCP), will be able to obtain an IPv4 home address (IPv4-MN-HoA) for its attached interface. Although the address configuration mechanisms for delivering the address configuration to the mobile node is independent of the PMIPv6 protocol operation, however, there needs to be some interactions between these two protocol flows. Figures 1.11 and 1.12 illustrate, respectively, how DHCP-based address configuration support can be enabled for a mobile node in a PMIPv6 domain, in two scenarios: i) DHCP server co-located with the MAG, and ii) DHCP relay agent co-located with the MAG. The DHCP server (DHCP-S) or the DHCP relay agent (DHCP-R) configured on the MAG is required to have an IPv4 address for exchanging the DHCP messages with the mobile node. This address is the default router address of mobile node provided by the LMA. Optionally, all the DHCP servers co-located with the MAGs in the PMIPv6 domain can be configured with a fixed IPv4 address. This fixed address can be potentially an IPv4 private address that can be used for the DHCP protocol communication on any of the access links. This address will be used as the server identifier in the DHCP messages.

The MAG may choose to ignore the DHCPDISCOVER messages till the PMIPv6 signaling is successfully completed, or it may choose to send a delayed response for reducing the additional delay waiting for a new DHCPDISCOVER message from the mobile node. For acquiring the mobile node's IPv4 home address from the LMA, the MAG will initiate PMIPv6 signaling with the LMA. After the successful completion of the PMIPv6 signaling and upon acquiring the mobile node's IPv4 home address from the LMA, the DHCP server on the MAG will send a DHCPOFFER message to the mobile node. The offered address will be the mobile node's IPv4 home address, assigned by the LMA. The DHCPOFFER message will also have the subnet mask option

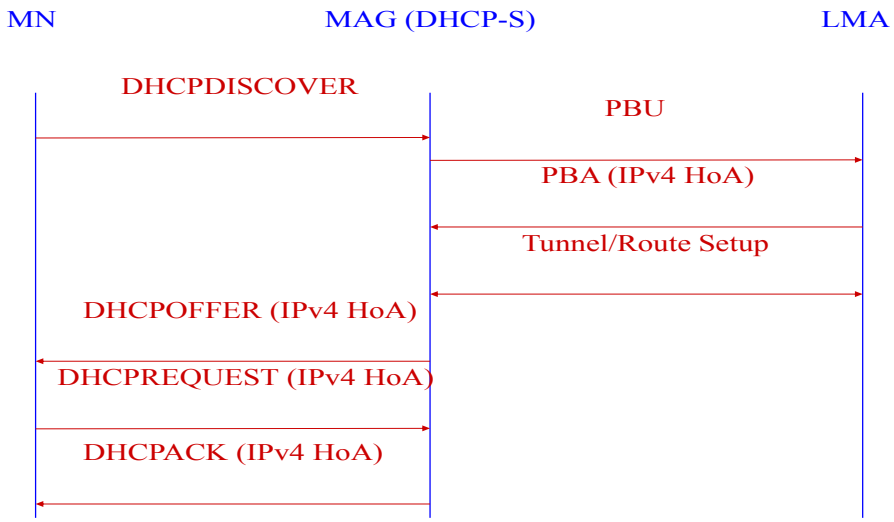

FIGURE 1.11
DHCP server co-located with MAG.

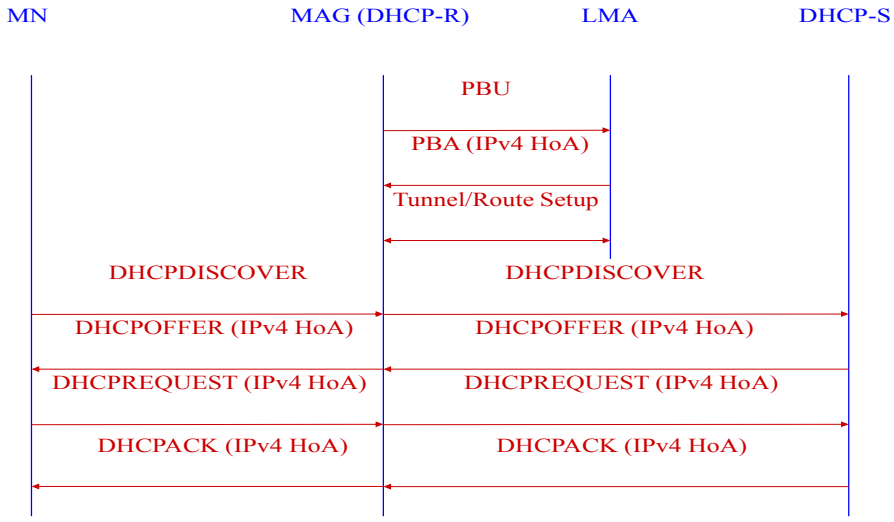

FIGURE 1.12
DHCP relay agent co-located with MAG.

and router option, with the values in those options set to the mobile node's IPv4 home subnet mask and default router address respectively. Additionally, the Server Identifier option will be included and with the value in the option set to the default router address.

In the IPv4 transport network support, the LMA and the MAG are configured and reachable using only IPv4 addresses, the MAG serving a mobile node can potentially send the PMIPv6 signaling messages over IPv4 transport and register its IPv4 address as the CoA in the mobile node's Binding Cache entry. An IPv4 tunnel with any of the supported encapsulation modes, e.g., IPv4 with UDP header, can be used for tunneling the data traffic of mobile node. The MAG can be potentially in a private IPv4 network behind a network address translation (NAT) device, with a private IPv4 address configured on its egress interface. But, the LMA must not be behind a NAT and must be using a globally routable IPv4 address. However, both the LMA and the MAG can be in the same private IPv4 routing domain, i.e., when both are configured with private IPv4 addresses and with no need for NAT translation between them. Note that the IPv6 address configuration requirement on the MAG does not imply that IPv6 routing enabled between the LMA and the MAG, is needed. It just requires each of the MAGs and LMAs in a PMIPv6 domain to be configured with a globally unique IPv6 address.

1.2.2 Network Mobility Management

While host mobility management deals with the mobility of individual mobile nodes, network mobility (NEMO) is concerned with managing the mobility of an entire network, which changes, as a unit, its point of attachment to the Internet. The mobile network includes one or more mobile routers (MRs), which connect the mobile network to the global Internet.

In NEMO basic support (NEMO-BS) [17], the IETF NEMO charter has adopted the methods for host mobility support used in MIPv6, without routing optimization, and has extended these methods in the simplest way possible to achieve its goals, i.e., allowing the session continuity for every node in the mobile network as the network moves. The basic support solution is for each MR to have a HA, and use bi-directional tunneling between the MR and HA to preserve the session continuity while the MR moves. The MR acquires a CoA at its attachment point much like what is done for mobile nodes using MIPv6. This approach allows nested mobile networks, since each MR will appear to its attachment point as a single node. Additionally, the NEMO basic support is transparent to mobile nodes behind the MR, and as such does not require: i) mobile nodes taking any actions in the mobility management, and ii) modifications to any mobile nodes other than MRs and HAs.

When a MR moves away from the home link and attaches to a new access router (AR), it acquires a CoA from the visited link. Then the MR immediately sends a binding update (BU) to its HA. When the HA receives this BU, it creates a cache entry binding the MR HoA to its CoA at the current point

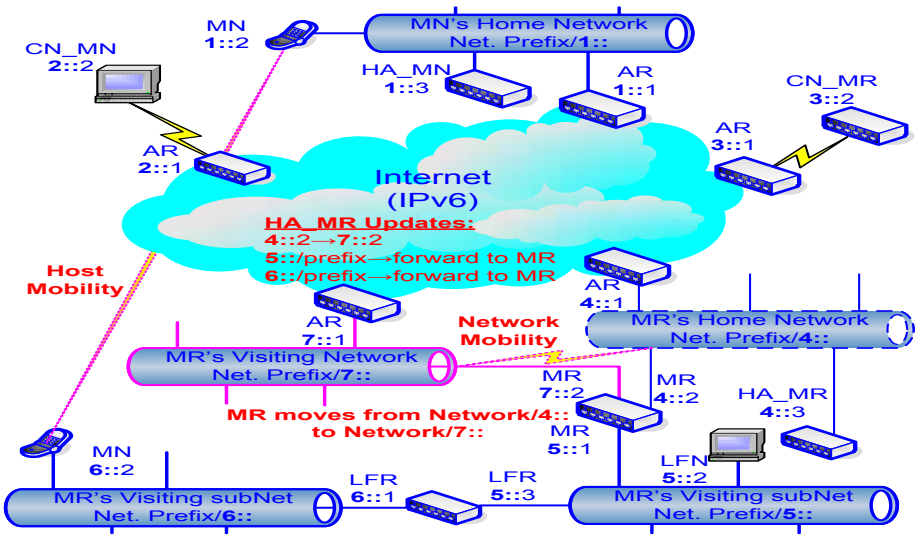

FIGURE 1.13
NEMO basic scenario operation.

of attachment. The HA acknowledges the BU by sending a binding acknowledgement (BA) to the MR. Once the binding process finishes, a bi-directional tunnel is established between the HA and the MR. The tunnel end points are MR CoA and HA address. If a packet with a source address belonging to the mobile network prefix (MNP) is received from the mobile network, the MR reverse-tunnels the packet to the HA through this tunnel. This reverse-tunneling is done by using IP-in-IP encapsulation. The HA decapsulates this packet and forwards it to the correspondent node (CN). When a CN sends a data packet to a node in the mobile network, the packet is routed to the HA that currently has the binding for the MR. The MR network prefix would be aggregated at the HA, which would advertise the resulting aggregation.

Alternatively, the HA may receive the data packets destined to the mobile network by advertising routes to the MNP. When the HA receives a data packet meant for a node in the mobile network, it tunnels the packet to the MR current CoA. The MR decapsulates the packet and forwards it onto the interface where the mobile network is connected. Before decapsulating the tunneled packet, the MR has to check whether the source address on the outer IPv6 header is the HA address. The MR also has to make sure that the destination address on the inner IPv6 header belongs to a prefix used in the mobile network before forwarding the packet to the mobile network. If it does not, the MR should drop the packet.

To illustrate the operation of NEMO basic support, Figure 1.13 depicts a scenario where both the mobile node (MN) and the MR move from their own home links and attach to visited links. The mobile network consists of

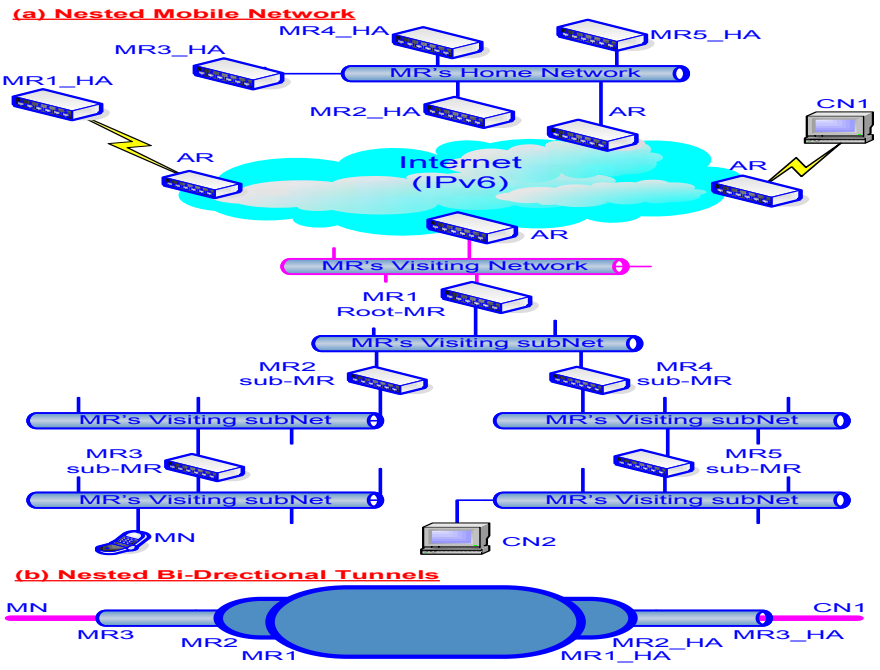

FIGURE 1.14
NEMO routing optimization problem.

a local fixed node (LFN) and a local fixed router (LFR). The MR sends a BU to its HA (HA_MR) when it attaches to a visited link and configures a CoA. HA_MR creates a binding cache entry for the MR HA and also sets up forwarding for the prefixes on the mobile network. The MN also configures a CoA from the prefix advertised on the mobile network and sends a BU to its HA (HA_MN) and to its correspondent node (CN_MN). Both HA_MN and CN_MN create binding cache entries for the MN home address. Note that an AR is used to connect the correspondent network into the Internet, and IPv6 is used for addressing.

However, in NEMO basic support, IETF NEMO charter does not consider: i) routing issues inside the mobile network, and existing routing protocols, including MANET protocols, can be used to solve these problems; ii) routing optimization; iii) managing multiple bi-directional tunnels between the MR(s) and the corresponding HA in NEMO multihomed configurations. Figure 1.14 shows a scenario of a nested mobile network using NEMO basic support, and thus, the flow of packets between MN and CN1 would need to go through three separate tunnels. This results in the increase of: i) length of packet route, ii) signaling overhead, iii) handoff delay, iv) packet transmission delay, v) protocol complexity and processing load.

To overcome above limitations, in [65] authors has analyzed various scenarios of route optimization in NEMO, and explores the benefits and tradeoffs in different aspects of NEMO route optimization. These scenarios of route optimization include: i) non-nested route optimization, ii) nested mobility optimization, iii) infrastructure based optimization, and iv) intra-NEMO optimization. However, the proposed solutions are not completed, not evaluated, and require further improvements. The border gateway protocol (BGP) announcements has been applied as a mobility method in the Connexion by Boeing (CbB) network [20]. The proposed solution provides a global IP mobility architecture which does not require the use of special IP stacks on Internet hosts or mobile nodes. Virtual private networks (VPNs) and other long-lived TCP sessions can be maintained across satellite transponder handoffs. The benefit of Boeing Connexion is that the protocol does not rely on the use of tunneling, compared to NEMO basic support, and thus, substantial latency incurred due to tunneling, and route optimization problem, are improved. However, the extra signaling overhead, i.e., BGP routes, corresponding BGP updates and announcements due to the aircraft handoffs, also need further study and examination. Further improvements of Boeing Connexion to reduce BGP updates and resolve the issue of route optimization, are presented in wide-area IP network mobility (WINMO) [33]. Through scoped BGP updates, route aggregation, tunneling, and mobility packet state, WINMO has achieved low stretch global Internet routing for mobile networks roaming across wide areas with minimal inter-domain routing overhead.

1.2.3 Hybrid Mobility Management

While host mobility management (Section 1.2.1) and network mobility management (Section 1.2.2) propose different protocols and network architectures to support the mobility of individual mobile nodes and mobile networks either at a macro scope or a micro scope, there is still a loose coupling between these existing work towards a solution for providing global IP mobility. The Boeing Connexion can be considered as an exception, which introduces a total new BGP-based, network-based NEMO solution. Next, this Section continues with a new hybrid protocol, called *Identifiers Separating and Mapping Scheme* (ISMS) [19], for a network-based host mobility and NEMO management in future Internet.

ISMS is designed to satisfy the requirements of future Internet: i) faster handover, ii) route optimism, iii) advanced management, and iv) location privacy and security. To achieve these requirements, ISMS separates customer networks from transit providers, and decomposes Internet addresses into accessing identifiers (AIDs) and switching routing identifiers (SRIDs) to decouple the locator/identifier overload of IP address. AIDs are network layer identifiers to be used by the transport layer. On the other hand, SRIDs serve as routing tags in the core network. For flexibility, AID to SRID mapping is used to bridge the customer networks and the core network. As a result, a

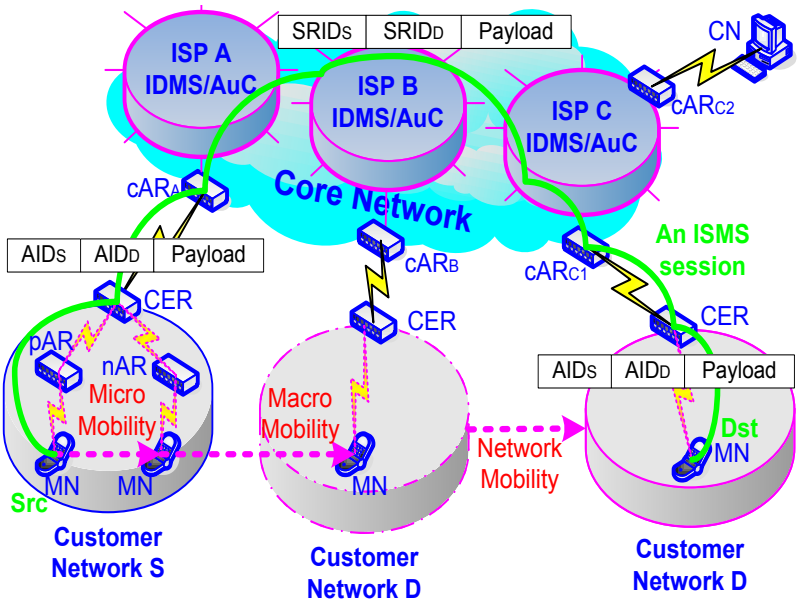

FIGURE 1.15
ISMS architecture, session, and mobility scenarios.

network-based mobility management scheme is obtained, which provides lower handoff latency and highly efficient mobility management. Figure 1.15 shows the ISMS architecture, a simple ISMS session from a source mobile node (MN) Src to a destination mobile node Dst, under different scenarios of mobility: i) micro host mobility, ii) macro host mobility, and iii) network mobility.

When a source mobile node Src in customer network S wants to send a packet to a destination mobile node Dst in customer network D, it forwards the packet with traffic engineering (TE) requirements to one of S providers, say ISP A. The core access router (cAR_A) contacts the identifiers mapping server (IDMS) to find the SRIDs of Dst. According to the TE requirements of Src, cARA chooses the SRID with higher preference, swaps the Src/Dst AID to Src/Dst SRID, respectively, and forwards it to cAR_{C1}. Upon receiving the packet, cAR_{C1} swaps the Src/Dst SRID to Src/Dst AID, respectively, and forwards it to Dst. In ISMS, IP address is overloaded with the semantics of both who, i.e., end-point identifier as used by transport layer, and where, i.e., locators for the routing system. The overloading is one of the main causes of the mobility problem. ISMS decomposes IP addresses into AIDs and SRIDs, and uses a separating and mapping scheme to support host mobility and network mobility.

For the micro host mobility of a MN within a single customer network S, from its previous access router (pAR) to a new AR (nAR), this MN will

receives AR advertisement from the nAR and triggers the route update mechanism from nAR up to customer edge router (CER), and back to pAR. Thus, the packets to this MN will be redirected at CER and sent to nAR. Note that micro host mobility will not cause the remapping of AID and SRID. Only the customer network S where the MN Src locates is aware of the movement.

For the macro host mobility of a MN across different cARs, e.g., from customer network S to customer network D via cAR_B in Figure 1.15, ISMS uses separating and mapping scheme to support for the mobility. Figure 1.16 shows the packet exchange sequences of the handoff procedure when a MN moves from customer network S to customer network D, while communicating with the correspondent node (CN). Note that when the macro host mobility handoff process finishes, the packet from CN to MN will follow the route $CN \rightarrow cAR_{C2} \rightarrow cAR_A \rightarrow cAR_B \rightarrow MN$, which is not the shortest path. Route optimization is then achieved by using traffic-driven method. After MN has moved to cAR_B, when cAR_A receives the first packet from CN to MN, it detects that MN has moved away and sends AR mapping update message to cAR_{C2} which includes the new mapping of MN. Alternatively, if MN sends packets to CN firstly, when cARB receives the first packet from CN to MN, it will inform cAR_{C2} the new mapping of MN by sending AR mapping update message. Once cAR_{C2} gets the new mapping of MN, the packet sent by CN to MN will be routed by the optimal route $CN \rightarrow cAR_{C2} \rightarrow cAR_B \rightarrow MN$.

For the network mobility of a mobile network, e.g., a customer network D in Figure 1.15, ISMS exploits the network-based mobility scheme, i.e., only the mobile router (MR) is aware of the moving when a mobile network moves to a new cAR. The MR works likes an MN and sets the new core AR as its default route. Since the MR and the core AR can communicate with each other using their AIDs, the MR needs not require a CoA. The MR will require the new core AR to reassign new mappings for the nodes in the mobile network. The nodes in the mobile network are no longer aware of the network which they attach to. All the connections based on AIDs will not interrupt. Benefit from the mapping service and unchanging AIDs, the problems of nested tunnels and triangle routing will not occur.

1.2.4 Discussion

The operations of many IP mobility management protocols have been described in previous Sections, in various mobility scopes of mobile nodes and mobile networks. This Section concludes with a qualitative comparison among these mobility protocols, see Table 1.2.4, which are based on the following metrics: i) low handoff latency (LHL), ii) low signaling overhead (LSO), iii) routing optimization (ROP), iv) location privacy (LoP), and v) tunneling (TUN). While the LHL metric is normally considered as a QoS metric for real-time applications, LSO, in conjunction with ROP and TUN metrics, are used to evaluate the successful ratio of packet delivery, and the efficient bandwidth

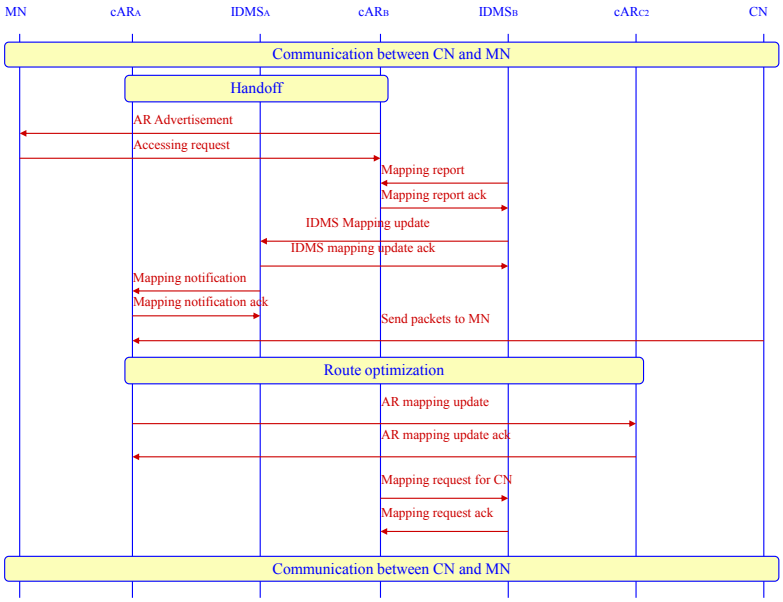

FIGURE 1.16
Macro-mobility management with ISMS.

utilization. Finally, the LoP metric ensures the location privacy upon the user request.

The localized or micro mobility of mobile nodes, in host-based host mobility (HBHM), between IP subnets under control of the same authority achieves LHL, LSO, and ROP, and thus, these micro HBHM mobility protocols are subjected to guarantee the required QoS for time-sensitive and real-time applications, but only in the local scope, e.g., within the enterprise networks. Tunneling or mobile-specific routing within IP subnets should also be considered in the forwarding packet strategies, which will affect much on the performance of wireless multihop access technologies, e.g., MANET, with more details in Section 1.4. For global or macro HBHM, MIPv4/v6 is the de facto protocol of mobility management due to its operational independence to underlying and above layers, though signaling overhead and latency handoff are two critical weaknesses. Thus, hybrid HBHM comes to balance for a global mobility solutions. However, the existing work in hybrid HBHM exploits L3/L2 cross-layer interaction without any standardized interface, and thus, the protocols are dependent on the underlying wireless access technologies.

While existing protocols for HBHM can satisfy different QoS requirements on latency and packet delivery, their success is limited to only theory work, or in the network testbeds. The reason is due to the modification of IP stacks in mobile nodes to support for the mobility, and thus, the operations of these

TABLE 1.1
Comparison of IP mobility management protocols.

Metrics	Host-based Host Mobility									Network-based Host Mobility		Network Mobility (NEMO)			Hybrid Mobility
	Macro	Micro					Hybrid								
	MIPv4/v6	HMIPv4/v6	CIP	LLMIPv4	HAWAII	FMIPv4/v6	FHMIPv6	SBF	SMIP	PMIPv6	IPv4/PMIPv6	NEMO-BS	CbB-BGP	WINMO	ISMS
LHL	X	O	O	O	O	O	O	O	O	X	X	X	X	O	O
LSO	X	O	O	O	O	O	O	O	O	X	X	X	X	O	O
ROP	O	O	O	O	O	O	O	O	O	O	O	X	O	O	O
LoP	X	X	X	X	X	X	X	X	X	X	X	X	X	X	O
TUN	O	O	X	O	X	O	O	O	O	O	O	O	X	X	X

LHL	Low handoff latency	ROP	Routing optimization	TUN	Tunneling	O	Yes	
LSO	Low signaling overhead	LoP	Location privacy			X	No	

HBHM protocols are not safe and reliable in the ISPs perspective. Therefore, the real deployment of HBHM protocols are not popular in 2G/3G mobile networks, and this problem leverages the development of: i) network-based host mobility (NBHM) protocols, in which network devices will control the signalling on behalf of nobile nodes; and ii) NEMO protocols to deal with the mobility of the whole mobile network. However, existing work in NBHM and NEMO protocols provide only basic support for mobility, i.e., lack of LHL, LSO, with or without ROP, and thus, further optimization needs to be developed. In NEMO, such an optimization exploits the BGP announcements to update the mobility of mobile networks, e.g., CbB-BGP, WINMO. This new concept is totally different from MIPv4/v6 for macro HBHM, avoiding the tunneling problem, and improving much the handoff latency in a global scope. The limitation is only in the use of BGP, which requires access routers on visited domains and network routers within Internet infrastructure uses BGP as inter-domain routing protocol.

Location privacy is a new metric, which is required to be support in recent work in hybrid mobility protocols (e.g., ISMS). Since the current IP addresses of mobile nodes in visited networks also show their locations, an additional sub-layer is needed for mapping between IP addresses and mobile node identifiers, and hiding locations of mobile nodes from upper application layers. This overlayed sub-layer can increase the signaling overhead for mapping between IP addresses and identifiers of mobile nodes.

1.3 IP Mobility Management in 1-Hop WLAN

The fast handoff and low latency schemes proposed in IP mobility management in Section 1.2 usually assume that mobile nodes can anticipate link-layer handoffs and maintain connectivity with the old as well as the new agents, e.g., access routers, access points. That is, these optimizations are applicable mainly to soft and backward handoffs. While such assumptions are valid for WLANs operating in the ad hoc or peer-to-peer mode, they do not hold for WLANs running in the infrastructure mode [81].

In infrastructure mode, a mobile node is associated with only one access point at a time. Although the IEEE 802.11 network interface card (NIC) on a mobile node may be able to access the signal strength information for all neighboring access points, such information is not available to mobility management software. As a result, previous proposals on fast handoff in IP mobility management in Section 1.2 that rely on the ability to anticipate an imminent link-layer handoff through signal strength comparison cannot be applied to IEEE 802.11 networks operating in infrastructure mode. Moreover, since the mobile nodes cannot receive packets transmitted from other access points, it is not possible to receive multiple foreign agent advertisements and maintain a list of neighboring foreign agents a priori for the future use. Thus, the link-layer handoffs in the WLAN infrastructure mode are hard and forward.

A handoff is hard if a mobile node can communicate with exactly one access point before and after a handoff, and it is forward if the mobile node cannot communicate with the old agent during the handoff and has to carry out the handoff by reestablishing a connection with the new foreign agent. Both properties are detrimental to the fast mobile IP handoff. Because the link-layer handoff is hard, a mobile node cannot obtain the address of new foreign agent before the handoff. Because the link-layer handoff is forward, a mobile node cannot contact the old agent during the handoff and request for additional buffering to minimize data loss during the handoff period [81].

Numerous schemes have been proposed to reduce the WLAN 802.11 handoff latency, which can be classified as either native (only L2) handoffs between *Basic Service Sets* (BSSs) within an *Extended Service Set* (ESS), or cross-layer (L3/L2) handoffs between ESSs (usually with different IP subnets). In this Section, WLAN L2 handoff procedure is first described in Section 1.3.1, including discovery phase and re-authentication/re-association phase. Then a classification on different L2 techniques to reduce the WLAN L2 handoff latency, including probe delay, authentication delay, and reassociation delay, is presented in Section 1.3.2. Finally, Section 1.3.3 shows cross-layer L3/L2 techniques to reduce WLAN 802.11 handoff latency, and inter-working with IP mobility management, e.g., MIP. For IP mobility management in 1-hop WLAN, existing work only focuses on the host mobility management, and

inter-working with IP host mobility management is the scope of Section 1.3.3 in this Chapter.

1.3.1 Handoff Procedure in 802.11/WLAN Networks

The handoff process in IEEE 802.11 networks has several phases each with its own costs. First, a client must determine that it is nearing the periphery of its coverage, and thus, must find an alternative access point to continue. This involves detecting that packets are no longer received successfully. However, typical commercial implementations also monitor the current signal-to-noise ratio (SNR) or current receiving signal strength indicator (RSSI), and initiate the scanning phase when this value passes a pre-defined minimum threshold. Setting this threshold is something of a black art: if the mobile node waits too long to look for new access points then it may incur additional disruption, yet if the mobile node is too eager then it may ping-pong between access points needlessly [77].

Pure Layer-2 Handoff Algorithm:
Current WLAN products support handoff at layer 2 (L2) with the aid of information from physical layer (L1). The handoff is mobile-initiated. Each mobile node is equipped with a WLAN adapter, and a serving channel will be selected for carrying all the data packets between the mobile node and the corresponding access point (AP). The mobile node, while staying in the radio coverage of the access point, will periodically check the current RSSI and calculate the current frame error rate (FER) on the serving channel. Besides, the mobile node actively scans all other channels for their receiving signal strength, also periodically. At any time, if the quality of the serving channel falls down under some pre-defined thresholds, the mobile node will decide to start a handoff process and look for a new serving channel. The pure layer-2 handoff procedure is illustrated in [92].

Figure 1.17 shows the sequence of steps that are designed to occur during a WLAN handoff. The first step (not indicated in Figure 1.17) is the termination of a mobile node association to the current AP. Either entity can initiate a disassociation for various reasons. Due to mobility or degradation of physical connectivity (signal strength), it might not be possible for the mobile node or the access point to send an 802.11 disassociate packet. In such cases, a timeout on inactivity or communication between APs or the receipt of an inter-access point protocol (IAPP) Move-Notify packet terminates the association [60].

During the second step, the mobile node scans for access points by either sending probe request packets (in active scanning), or by listening for beacon packets (in passive scanning) broadcasted by APs on the channels of interest. Packets A, B, C and D in Figure 1.17 show the active scanning. For each channel, a probe request (packets A and C), is sent by the mobile node and probe responses (packets B and D) are received from the APs in the vicinity of the mobile node. After scanning all intended channels, the mobile node selects

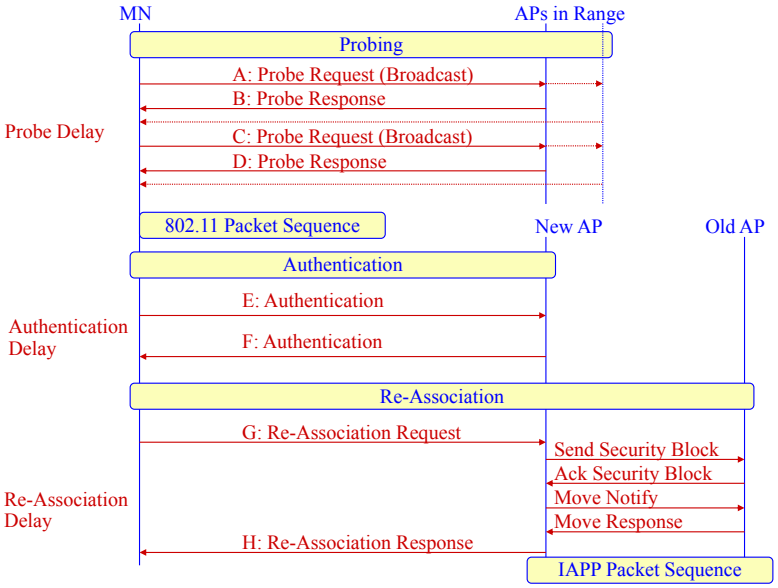

FIGURE 1.17
Handoff procedure in 802.11 networks.

the new-AP based on the data rates and signal strength, see the pure layer-2 handoff procedure above as an example. Probe delay is the time spent by the mobile node in scanning and selecting the next AP.

In passive scanning, mobile nodes listen to each channel for the beacon frames (typically every 100 ms). The main inconvenience of this method is how to calculate the time to listen to each channel. This time must be longer than the beacon period, but the beacon period is unknown to the mobile node until the first beacon is received. Incidentally, the mobile node cannot switch to another channel when the first beacon arrives and has to wait for the whole beacon period because several access points of different WLANs can operate in the same channel. Since the standard mandates that the whole set of allowed channels must be scanned, mobile nodes need over a second to discover the access points in range with the default 100 ms beacon interval, e.g., there are 11 allowed channels in USA, thus it would take 1.1 seconds.

When faster scanning is needed, mobile nodes must perform active scanning. Active scanning means that mobile nodes will broadcast a probe-request management frame in each channel and wait for probe responses generated by access points. Each mobile node should wait for the responses in each channel within an interval, which is controlled by the two timers: MinChannelTime and MaxChannelTime. The first is the time to wait for the first response in an idle channel. If there is neither response nor traffic in the channel during

MinChannelTime, the channel is declared empty, i.e., no access point in range. The second timer, MaxChannelTime, indicates the time to wait in order to collect all responses in a used channel. This limit is used when there was activity in the channel during MinChannelTime. Both timers are measured in Time Units (TU), which the IEEE standard defines to be 1024 microseconds. Exact values for these timers are not given in the standard [87].

After the probe, the mobile node and new-AP exchange 802.11 authentication packets, and the latency incurred is the authentication delay (packets E and F). After authentication, the mobile node sends an 802.11 re-association request to the AP (packet G) and receives a re-association response from the AP (packet H) which completes the handoff process. The latency incurred during this exchange is the re-association delay and this process is called the re-association process. During re-association, the APs involved exchange mobile node context information. This is achieved via the use of the IAPP [2].

The mobile node attempts to re-authenticate to an AP according to the priority list. The re-authentication process typically involves an authentication and a re-association to the posterior AP. The re-authentication phase involves the transfer of credentials and other state information from the old-AP. As mentioned earlier, this can be achieved through a protocol such as IAPP [2]. IAPP plays a significant role during a handoff. The two main objectives achieved by inter-access point communication are:

- Maintaining a single association of a mobile node with the wireless network.

- Securing transfer of state and context information between APs involved in a re-association. The client context information can include but is not limited to IP flow context, security context, quality of service information, e.g., differentiated service or integrated service, header compression and authentication, authorization, and accounting (AAA) information [60].

Association and re-association events change a mobile node point of access to the network. When a mobile node first associates to an AP, the AP broadcasts an Add-Notify packet notifying all APs of the mobile node association. Upon receiving an Add-Notify packet, the APs clear all stale associations and state for the mobile node. This enforces a unique association for the mobile node with respect to the network. When a mobile node re-associates to a new-AP, it informs the old-AP of the re-association using IAPP packets. At the beginning of a re-association, the new-AP can optionally send a Security Block packet to the old-AP, each of which acknowledges with an Ack-Security-Block packet. This packet contains security information to establish a secure communication channel between the APs. The new-AP sends a Move-Notify packet to the old-AP requesting mobile node context information and notifying the old-AP of the re-association. The old-AP responds by sending a Move-Response packet. Although IAPP communications serve to fulfill the

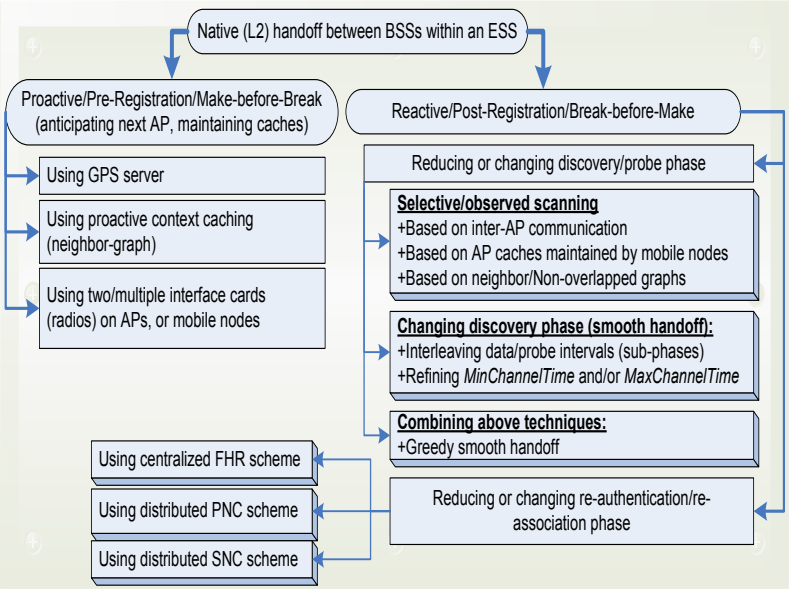

FIGURE 1.18
Classification of methods to reduce handoff latency in WLAN.

mandatory distribution system functions, they invariably increase the overall handoff latency because of their reactive nature [60].

1.3.2 Native Layer-2 Handoff in WLAN

Based on the description of 802.11 operations and its handoff procedure in Section 1.3.1, different methods have been proposed to reduce the L2 handoff latency in WLAN 802.11, which can be classified as either proactive or reactive, see Figure 1.18. In [62], authors use the global positioning system (GPS) system in the handoff management. The idea is to predict the next mobile node point of attachment and the associated sub-network using the position of the mobile node. The selection of the next mobile node AP prior to the handoff allows the configuration of all the required parameters on the mobile node to reduce L2 discovery phase and L3 new link detection.

Proactive context caching and forwarding technique [60] introduces a novel and efficient data structure, the neighbor graph, which dynamically captures the mobility topology of a wireless network through real-time examination of the handoffs occurring in the network. The neighbor graph is use to provide the new-AP with the mobile node context prior to the handoff, or pro-actively. Thus, this technique reduces the re-association latency.

A novel AP with two transceivers to improve network efficiency in terms

of supporting seamless handoff and balancing the traffic load is shown in [55]. In this scheme, the novel AP will use the second transceiver to scan and find neighboring mobile nodes in the transmission range and later send the results to its associate AP. This information will be useful for the AP to control its associated mobile nodes initiate the handoff process when a neighbor AP can provide higher quality and/or sharing the traffic load with neighbor APs. Ramachandran *et al.* [76] introduces the novel concept of applying make-before-break mechanism to 802.11 MAC layer handoffs. It optimizes the L2 handoff by periodically probing in the background, which uses a second radio card, to gather neighborhood information even when it is already connected to an AP. An empirical analysis of the 802.11 MAC (L2) handoff process carried out in [59] has shown that the probe phase is the significant contributor to the handoff latency. Thus, most reactive schemes concentrate on reducing the probe delay, while some focus on reducing the authentication and/or re-association delay.

Shin *et al.* [83] concentrates on reducing the probe delay. First, the probe delay is reduced by improving the scanning procedure, using a selective scanning algorithm. Second, the number of times in the previous scanning procedure is minimized using the AP caching mechanism. Specifically, full-scan is triggered at first and the channel mask is then constructed by the information obtained in the first full-scan. In 802.11b, only three channels do not overlap among all 11 channels. Hence, in a well-configured wireless network, almost all APs operate on channels 1, 6, and 11. Thus, the channel mask is formed by combining these three frequency channels. By using this channel mask, a mobile node can reduce the amount of unnecessary time that it spends on probing non-existent channels among neighboring APs. To reduce further the handoff delay, a cache mechanism is also introduced. The basic idea of the caching mechanism is for each mobile node to store its handoff history. When a mobile node associates with an AP, the AP is inserted into the cache maintained at the mobile node. When a handoff is needed, the mobile node first checks whether there is an entry corresponding to the current AP MAC address in the cache. If there is a matched cache entry, the mobile node can associate with the AP without any further probing procedures [67].

In [82], authors describe the use of a novel and efficient discovery method using neighbor graphs and non-overlap graphs to reduce the total number of probed channels and the total time spent waiting on each channel. Using neighbor graph, the set of channels on which neighboring APs are currently operating and the set of neighbor APs on each channel can be learned. Based on this information, a mobile node can determine whether a channel needs to be probed or not. On the other hand, the non-overlap graph abstracts the non-overlapping relation among APs. Two APs are considered non-overlapped if and only if the mobile node cannot communicate with both of them simultaneously with acceptable link quality. Therefore, if a mobile node has received a probe response frame from an AP_i, it implies that this mobile node cannot receive a response frame from an another AP_j if AP_i and AP_j are non-

overlapping each other. By means of the non-overlap graph, a mobile node can prune some of the APs, which are non-overlapped with the current APs that have already responded [67].

[51] presents a smooth MAC layer handoff scheme and a greedy smooth MAC layer handoff scheme. In the former, the scan channel phase is divided into multiple sub-phases. The mobile node can use the interval between two consecutive sub-phases to send and receive data frames. Obviously, this can reduce the packet delay and jitter during the scanning channel phase, which is important for time critical applications such as VoIP. The latter scheme not only scans channel smoothly but also reduces the number of channels being scanned. [87] shows a tuning scheme, using the packet loss distribution caused by collisions, in order to determine the optimal handoff trigger timing to a new AP. To reduce the handoff detection time, the mobile node starts the channel probe procedure as soon as it deems that collision can be excluded as a reason for the packet transmission failure. In other words, based on the probability distribution, if a packet and its next two consecutive retransmissions fail, the mobile node concludes that the packet failure is caused by the mobile node's movement instead of collision and therefore a further handoff process is required. In addition, they leverage the active scan mode and derive new values for MinChannelTime and MaxChannelTime from their measurement results and analytical models [67].

Even though the probe delay takes a large portion of the total handoff delay, the authentication/re-association delay should be also reduced to achieve seamless mobile services. Actually, in public WLAN services, the authentication scheme based on the centralized authentication server is widely adopted for the sake of secure service and efficient accounting. In such an environment, the authentication/re-association delay may be higher than that observed in the case where the open authentication procedure is employed. Different fast authentication methods in IEEE 802.11 networks are overviewed and analyzed in [67] in terms of network architectures and trust models. [68] proposes a fast handoff scheme that minimizes the re-authentication latency in public wireless LANs. When a mobile node sends an authentication request, the AAA server authenticates not only the currently used AP, but also multiple other APs, and sends multiple WEP keys to the mobile node. A centralized algorithm is used to select these APs. This selection algorithm is based on the frequent handoff region (FHR), traffic pattern, and user mobility characteristics within the public WLAN.

Instead of using the centralized system, a proactive scheme based on a distributed cache structure, called Proactive Neighbor Caching (PNC) [60], is introduced. This scheme uses a neighbor graph, which dynamically captures the mobility topology of the wireless network for pre-positioning a mobile node context. The scheme ensures that the mobile node's context is always dispatched one-hop ahead, and therefore the handoff delay can be substantially reduced. The neighbor graph is constructed using the information exchanged during the mobile node's handoff, and it is maintained at each AP in a dis-

tributed manner. The propagated mobile node context is stored in the cache. Recently, this scheme is included in the IAPP specification [2]. [69] enhances the PNC scheme by adding a new concept of neighbor weight. The neighbor weight represents the handoff probability for each neighboring AP. Based on the neighbor weight, the mobile node context is propagated only to the selected neighboring APs. The neighbor graph and its neighbor weights can be constructed by monitoring the handoff patterns among the APs.

1.3.3 Cross-Layer Handoff in WLAN

The problems with mobile wireless Internet arise even at the availability of MIP and WLAN. One of them is that MIP and WLAN solve their problems independently at different layers, i.e., MIP works at IP layer (L3) and does not talk to L2, and vice versa. Another drawback is that there is no standard or guidelines on cell planning for the Internet in the wireless environment. The cell planning is essential on cellular networks such as GSM. With cell planning, each base station is given a neighbor lists consisting of the network candidates to which a mobile nodes can be handed over. Without such planning, a mobile node does not know where the neighbor is. It also does not know which radio channel to tune into at the lower layers [92].

When an IEEE 802.11-based wireless network is configured in the infrastructure mode, each mobile node is associated with the access point of the WLAN segment in which it currently resides. The link layer (L2) handoff implementation in infrastructure-mode wireless LANs poses two problems in reducing MIP handoff latency:

- Because L2 handoff is transparent to software, it is not possible for MIP code to detect its occurrence immediately.
- Because a mobile node can only receive packets from one access point at a time in infrastructure mode, MIP code can only receive advertisements from the current foreign agent, but not those from neighboring foreign agents. As a result, even if a mobile node can detect the L2 handoff immediately, without advertisements from the new foreign agent it still does not know which foreign agent to contact to facilitate the L3 handoff process.

To reduce MIP handoff latency, let us first identify the individual delay components involved. A MIP handoff period for an infrastructure-mode WLAN can be divided into four distinct sub-periods [81]:

- Time between when a link-layer handoff takes place and when it is detected by the mobile IP software.
- Time from link-layer handoff detection to the reception of first mobile IP advertisement from the new foreign agent.

- Time required for a mobile node to register with the new foreign agent after receiving the first advertisement.
- Time between request sent to and response returned from the new foreign agent.

To minimize the overall handoff latency, each of these delay components needs to be reduced as much as possible. A low-latency mobile IP handoff scheme is proposed in [81], that can reduce the handoff latency of infrastructure-mode wireless LANs to less than 100ms. This scheme expedites link-layer handoff detection based on the access point ID probing supported by Wi-Fi cards, and speeds up network-layer handoff by replaying cached foreign agent advertisements. The proposed scheme strictly adheres to the MIP standard specification, and does not require any modifications to existing MIP implementations. A new architecture is described in [92], which is compatible with MIP and most L2 wireless networks. The solution consists of three major extensions: i) packet buffering, ii) neighbor list update, and iii) L2 handoff notification. The effect of this enhancement provides a linkage between different layers for preventing packet loss and reducing handoff latency.

[94] proposes a new low latency handoff method, where access points used in a wireless LAN environment and a dedicated MAC bridge, are jointly used to alleviate packet loss without altering the MIP specifications. [25] introduces the fast hinted cell switching movement detection method for MIP. This method assumes that the L2 is capable of delivering information, i.e., IEEE 802.11 SSID field was utilized to contain information important to MIP such as the identity of the local mobility agent, to MIP regarding the identity of the local mobility agent. This tends to negate the need for MIP movement detection as well as agent discovery, and leads to accelerated MIP hand-offs. Moreover, the existence of the L2 information renders the presence of periodic mobility agent broadcast advertisements as unnecessary, which enables a more efficient utilization of the network capacity.

Finally, last but not least, [56] describes how a FMIPv6 could be implemented on link layers conforming to the 802.11 suite of specifications. This work is among little effort, which attempts to give a set of examples for FMIPv6 over 802.11 networks, and examine how and when handover information might become available to the IP layers that implement fast handover, both in the network infrastructure and on the mobile node.

1.4 IP Mobility Management in MANET

Attempts are in progress to connect mobile ad hoc networks (MANETs) to the Internet infrastructure to fill in the coverage gaps in the areas where the first-hop coverage, e.g, WLANs or cellular networks, is not available. In

the very near future, mobile nodes will roam across multiple heterogeneous platforms while continuously maintaining session connectivity. A mobile node may connect to a WLAN, and then move into an area where the coverage from the WLAN does not exist. There, it may reconfigure itself into ad hoc mode and connect to a MANET. Essential to such seamless mobility is efficient mobility management. This Section focuses on IP host mobility management in n-hop MANET, in conjunction with the integration of MANET with IP networks, and how MANET adapt to the network functionalities within IP networks.

1.4.1 Comparison of 1-Hop and N-Hop Access Networks

The location of a node in the Internet is identified by the network part of its IP address. When a packet arrives at the router connecting to the 1-hop WLAN hosting the destination, the destination IP address is converted to a MAC address. This conversion uses either the address resolution protocol (ARP) in IPv4, or neighbor discovery protocol (NDP) in IPv6, before the packet is sent in a frame in the last hop using the MAC address as the identifier of the destination. The MAC address represents a flat address space without information about the host locality within the WLAN [5].

Considering the n-hop MANET has a flat address space, it can be seen as a either major network or sub-network within the Internet. This will identify a MANET connected to the Internet by its own network number. There is, however, a major difference between a WLAN and a MANET. Mobile nodes connected to a WLAN are within the same broadcast domain, and are managed as 1-hop connections by the IP protocol. A packet broadcasted from a mobile node in the WLAN will reach all other mobile nodes connected to the network.

In the MANET, however, a broadcast sent by one MANET node may not reach all other MANET nodes. A broadcast needs to be retransmitted by immediate MANET nodes in the network, so that it will reach all MANET nodes. A broadcast in the MANET running the IP protocol uses the time to live (TTL) value to limit the spreading of a packet. A packet TTL when arriving at a router connected to a WLAN requires a value of "1" to reach a node connected to the network. In the MANET, a TTL of "1" when forwarded on the MANET will be discarded after the first hop. This behavior needs to be managed by gateways connecting MANETs with the Internet. Instead of using the ARP/NDP, the MANET routing protocol need to be used in what is defined as the last hop in the Internet. The functionality in reactive ad hoc routing protocol maps well to the functionality in the ARP/NDP protocol, with a request for the mobile node in the last hop and a soft state table. For example, the route request (RREQ) in the reactive MANET routing protocol can be compared to the ARP request, and the soft state MANET routing table to the ARP table [5].

A node performs a handoff if it changes its Internet Gateway (IGW) while

communicating with a correspondent node (CN) in the Internet. In conventional mobile networks, e.g., WLAN in Section 1.3, the quality of the wireless link between mobile node and base stations (i.e., APs) determines when to handoff from one AP to another. The performance of these types of handoffs depends on the mobility management protocol in the access network. In MANETs, the situation is more complicated. In general, some nodes do not have a direct wireless link to an AP, but they are connected via other immediate nodes. Thus, they cannot initiate handoffs that are based on the link quality to the AP. Rather, the complete multihop path to the AP, which serves as the current IGW, must be taken into the consideration. A handoff can occur if the mobile node itself or any of the intermediate relay nodes moves and breaks the active path. In general, if the path between a mobile node and the IGW breaks, and there is no other path to the same IGW, the mobile node has to perform the gateway discovery to establish a new path to another IGW [61].

1.4.2 Functionalities of IP Mobility Management in MANET

In this section, the required functions of providing Internet access and IP host mobility management for MANET nodes are described first. They include: i) MANET node location determination, ii) IGW discovery, iii) IGW selection, iv) IGW forwarding strategy, v) address auto-configuration, and vi) handoff-style. Next, the related work is discussed following the descriptions of above functions.

Figure 1.19 illustrates a typical mobility scenarios of a mobile node (i.e., node (1)) in MANET while connecting to the Internet. This is a scenario where above functionalities are needed. In Figure 1.19, MIPv4 is used for the macro-mobility management, i.e., between MANET domains, while ad-hoc routing is used for the micro-mobility management, i.e., within each MANET domain. This is a popular solution for IP mobility management in MANETs [47].

1.4.2.1 Node Location Determination

MANET node location determination is the function that allows a source MANET node to determine whether a destination node is located within the same MANET domain as the source node, or outside the MANET domain (e.g., on an external network, such as an Internet host). The function can be implemented by one of the following methods:

- **Network prefix**. All MANET nodes share the same network prefix. With this method, each MANET node must be assigned a global unicast IP address, both home of address (HoA) and care-of address (CoA), i.e., the MANET node address is topologically correct [73].

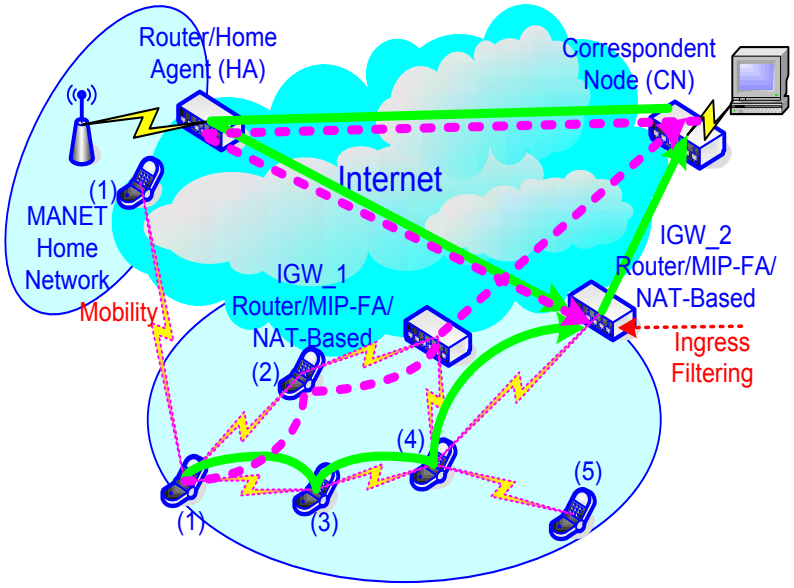

FIGURE 1.19
Overview of mobility management in MANET.

- **Routing table** in MANET proactive ad-hoc routing protocols [16, 70]. If an entry for the destination is in the routing table of the source MANET node, the destination is either in the same MANET domain, or an Internet host reachable via the IGW. Otherwise, the destination is unreachable.

- **Flooding** the route request (RREQ) and waiting for the route reply (RREP) in reactive ad-hoc routing protocols [72, 71]. If a host route is returned, the destination is in the same MANET domain. If a default route is returned, the destination is either unreachable or an Internet host reachable via the IGW.

- **Internet gateway**. The IGW responds to a RREQ, sending a proxy RREP to signal that it can route to the requested destination, i.e., analogous to functionality of a proxy ARP, but over the multihops. To do this, an IGW must determine that the destination is not in the same MANET by keeping a list of currently known active nodes, called visitor list, or by pinging the destination on the IGW network interface attached to Internet, or by flooding the whole MANET with a new RREQ [4, 73].

1.4.2.2 Internet Gateway Discovery

Internet gateway discovery is the function that allows a MANET node to discover an IGW to which traffic bound for the Internet can be forwarded, and from which traffic returned from the Internet can be received. The different discovery mechanisms can be classified into three sub-classes: proactive, reactive, or hybrid [35]. In the proactive approach, each IGW broadcasts periodically an advertisement, while in the reactive approach a MANET node sends a solicitation and waits for a reply from the IGW. The former requires much overhead traffic on the MANET, while the latter entails the longer discovery delay. The hybrid approach compromises with the balance, in which each IGW periodically broadcasts the advertisement within the radius of n-hops. MANET nodes that are located further than n-hops away from the IGW, must use the reactive approach to discover the IGW [80].

1.4.2.3 Metrics for Internet Gateway Selection

Internet gateway selection is a function used when a MANET node discovers multiple IGWs for accessing the Internet. A metric is normally needed in order to select the right one. Different metrics can be used:

- **Shortest hop-count**. To the nearest IGW [73].
- **Load-balancing**. For intra-MANET traffic, choose different immediate relays node to destination MANET nodes within the same MANET domain [79, 14], while for inter-MANET traffic, choosing different IGWs for forwarding traffic from MANET to Internet and vice verse [32].
- **Service class**. Depending on the service classes and data caches provided and managed by each IGW, respectively [13], as well as the wireless link quality among MANET nodes and IGWs [64].
- **Euclidean distance**. Spatial distance between the MANET node and the IGW [7].
- **Hybrid**. A combination of some of the above metrics [7].

1.4.2.4 Internet Gateway Forwarding Strategies

Internet gateway forwarding strategies is a function that takes the responsibility to forward traffic within the MANET, out of the MANET to the Internet, or from the Internet into the MANET. Typically, it can be classified into inter-MANET and intra-MANET forwarding strategies. The inter-MANET forwarding strategies uses different approaches as follows:

- **Default routes**. Representing the default next-hop to send packets to that do not match any other explicit entry in a MANET node's routing table. Usually, the default route is used to forwards packets towards an

IGW, where packets are further forwarded towards the destination in the Internet [73, 9].

- **Tunneling (or encapsulation)**. Usually, the IP-in-IP encapsulation technique is used to get traffic into and out of the MANET. The outer IP header is for the tunneling connection between the source MANET node and the IGW, while the inner IP header is for the connection between the source MANET node and the destination [38].

- **Half-tunneling**. Traffic to the Internet from the MANET domain uses tunneling, while traffic from the Internet to the MANET domain uses ad-hoc forwarding without tunneling [38].

- **Source routing**. A list of all intermediate nodes between the source MANET node and the IGW are added into the IP header. At the IGW, the source routing header is removed and the packet is forwarded further to the Internet as a normal packet [11].

- **Spanning tree rooted at the IGW**. A tree rooted at the IGW is built and maintained using the agent advertisements broadcasted periodically by the corresponding IGW [24].

The intra-MANET forwarding strategies, on the other hand, is based entirely on the operation of ad-hoc routing protocols. These can be classified as proactive, or reactive, or hybrid. In the proactive approach, each node continuously maintains up-to-date routing information to reach every other node in the network. Routing table updates are periodically transmitted throughout the network in order to maintain table consistency. Thus, the route is quickly established without any delay. However, for a highly dynamic network topology, the proactive schemes require a significant amount of resources to keep routing information up-to-date and reliable. In the reactive approach, a node initiates a route discovery throughout the network, only when it wants to send packets to its destination. Thus, nodes maintain the routes to only active destinations. A route search is needed for every new destination. Therefore, the communication overhead is reduced at the expense of delay due to the route discovery. Finally, in the hybrid approach, each node maintains both topology information within its zone via the proactive approach, and the information regarding neighbor zones via the reactive approach.

1.4.2.5 Address Auto-Configuration Scheme for MANET

In order to enable a MANET to support IP services and the internetworking with the Internet, a MANET address space based on IPv4/IPv6 is required. Moreover, the MANET addressing schemes must be auto-configured and distributed to support for the self-organized and dynamic characteristics of MANETs. Numerous addressing schemes for MANETs based on IP address auto-configuration have been proposed in the literature. They can be classified

into two approaches: conflict-detection allocation and conflict-free allocation [91].

Conflict-detection allocation mechanisms are based on picking an IP address from a pool of available addresses, configuring it as tentative address and asking the rest of the nodes of the network, checking the address uniqueness and requesting for approval from all the nodes of the network. In case of conflict, e.g., the address has been already configured by another node, the node should pick a new address and repeat the procedure (as a sort-of "trial and error" method). This process is called duplicate address detection (DAD). Conflict-free allocation mechanisms, on the other hand, assume that the addresses are delegated uniquely, and that they are therefore not being used by any other node in the network. This can be achieved by ensuring that the nodes, that delegate the addresses, have disjointed address pools. In this way, there is no need of performing the DAD procedure.

Research has also been in-progress to apply IP address auto-configuration scheme for the addressing of MANETs. However, only stateless mechanism is suitable for MANETs [10]. This is because the stateful mechanism requires a centralized server to maintain a common address pool, while the stateless mechanism allows the node to construct its own address and is suitable for self.organized MANETs. However, to use the IP address stateless auto-configuration scheme for MANET addressing, i.e., a conflict-detection allocation approach, a DAD mechanism is required to assure the uniqueness of the address with multi-hop distance, especially to support for MANET merging and partitioning.

Finally, the address allocation space is important. It must be large enough to cover the large-scale MANETs and reduce the probability of address conflicts. The following IPv4 and IPv6 addressing spaces have been proposed for MANETs [73]: 169.254.0.0/16 for IPv4, and FEC0:0:0:FFFF::/64 for IPv6 as MANET IP PREFIX.

1.4.2.6 Handoff Style

A node performs a handoff if it changes its IGW while communicating with a CN in the Internet. In MANETs, mobile nodes do not have a direct wireless link to an AP, and thus, they cannot initiate handoff that are based on the link quality to the AP. A handoff can occur if an ad-hoc node itself or any of the intermediate relay ad-hoc nodes moves and breaks the active path. In general, if the path between an ad-hoc node and IGW breaks, and there is no other path to the same IGW, the ad-hoc node has to perform an IGW discovery to establish a new path to another IGW.

The IGW discovery scheme and the ad hoc routing protocol both have huge influence on the multi-hop handoff performance. Multi-hop handoff schemes can be classified into forced handoff and route optimization-based handoff. The former occurs whenever the path between the source/destination mobile node and the IGW is disrupted during data transmission due to, e.g., the

TABLE 1.2 Protocols Acronyms.

Scheme	Description
TBBR	tree based bidirectional routing [24]
OLSR	optimized link-state routing [16]
AODV	ad-hoc on-demand distance-vector routing [72]
DSDV	destination-sequenced distance-vector routing [70]
DSR	dynamic source routing [36]
NAT	network address translation
DHCP	dynamic host control protocol
TD/RD	table-driven/root-driven routing [24]
MEWLANA	Mobile IP enriched wireless local area network architecture [24]

movement of the MANET node. Therefore, a new path to the Internet has to be set up. The following IGW discovery process may result in the detection of a new IGW, which will consequently result in a handoff. The latter is a handoff that results from route optimization. If the source/destination MANET node detects that a shorter path to the Internet becomes available while communicating with a corresponding node, the active path will be optimized. In case the shorter path goes via a different IGW, a route optimization-based handoff occurs.

1.4.3 Overview of IP Mobility Management in MANET

Table 1.4.3 summarizes the comparison of different approaches for providing Internet connectivity and mobility management for MANETs, based on the description of the required functions in Section 1.4.2. The acronyms in Table 1.2 are used for this comparison.

- **MIPv6+AODVv6**. Globalv4 [8] describes MIPv4 extensions for AODV routing protocol. Destinations are first searched for in the MANET. If none is found, a host route is setup to the IGW. This solution suffers from long route discovery delays and lack of the same route aggregation that half-tunneling provides. Similarly, Globalv6 [89] can also work with MIPv6 [37], but it is not mandatory. A node may acquire a network prefix from an IGW, and construct a globally routable IP address through IPv6 stateless address auto-configuration. Globalv6 employs a similar technique as Globalv4 to determine the locality of destinations. Routing towards the IGW is done on a hop-by-hop basis using a default route of AODVv6 [71]. Cascading effects [48], i.e., all immediate MANET nodes on the chain from the source MANET node to the IGW needs to flood RREQ to determine whether the destination is located in the same MANET, are avoided by requiring intermediate nodes to configure host route entries for Internet destinations, with

TABLE 1.3
IP mobility management mechanisms for MANET.

Index	Mechanism	Location determination	IGW discovery	IGW selection metrics	IGW forwarding	Addressing	Handoff style
1	MIPv6+ AODVv6	Network prefix	Proactive & reactive	Not specified, implicitly shortest hop-count	Default route & AODVv6	Deriving from IPv6 stateless auto-configuration	Both
2	MIPMANET	Flooding RREQ	Proactive	Shortest hop-count	Half-tunneling & AODV	Not specified, but Home Address must be IP global unicast	Route optimization-based
3	MIP+DSR	Using IGW	Reactive	Not specified, implicitly shortest hop-count	Source routing & DSR	Home Address must be IP global unicast	Not specified, implicitly route optimization-based
4	MIP+OLSR	Using routing table	Proactive	Not specified, implicitly shortest hop-count	Default route & OLSR	Not specified	Forced (when a prefix change)
5	MEWLANA TD/RD	Using routing table (DSDV) or TBBR Tree	Proactive	Shortest hop-count	Default route & DSDV or TBBR	Not specified	Forced (when a route change or node leave)
6	Two-tier MANET	Using routing table	Reactive	Load-balancing	Tunneling uses extra UDP/IP header & DSDV	Private address & NAT, allocating using DHCP	Not specified, implicitly route optimization-based
7	Hybrid MANET	Using routing table	Reactive	Hybrid: Euclidean distance & load-balancing	Default route & DSDV	Not specified	Forced (using automatic mode-detection and switching)
8	WLAN & MANET	Using routing table	Proactive	Not specified, implicitly shortest hop-count	Default route & OLSR	IPv6 stateless auto-configuration	Forced (using automatic mode-detection and switching)

the downside of losing route aggregation. A summary of Globalv4 and Globalv6 is also presented in [73].

- **MIPMANET** [38] studies the integration of MIPv4 in the MANET. Tunneling from ad hoc nodes to the foreign agent (FA) is proposed as a way to achieve default route like behavior. This is the half-tunneling approach, where the outbound traffic to the Internet from the MANET uses tunneling and the inbound traffic from the Internet to the MANET is delivered to the corresponding destination MANET node via the host route, using AODV routing protocol. However, this work does not explore the benefits of using tunneling, but studies different approaches to disseminate MIP information in the MANET instead.

- **MIP+DSR**. A technique is described in [11] to integrate MANETs with the Internet and to use MIP to support the migration of nodes between MANET and the Internet. Local delivery within a MANET subnet is accomplished using the DSR routing protocol, while standard IP routing mechanisms decide which packets should enter and leave the subnet. For sending packets from the MANET subnet to the Internet, a source MANET node uses the source routing header of DSR to forward the packet to the IGW, where the source routing header will be removed. The packet is then forwarded to the Internet. However, this technique requires that each MANET node selects a single IP address (its home address) from the ones assigned to it, and that it uses only that address when participating in the DSR protocol.

- **MIP+OLSR**. The management of universal mobility, including both large-scale macro-mobility and local scale micro-mobility, is the focus of [9]. A hierarchical architecture is proposed, including: i) extending micro-mobility management of a wireless access network to a MANET, ii) connecting this MANET to the Internet, and iii) integrating MIP and OLSR routing protocol to manage the universal mobility. In addition, it uses the optimal default route via the IGW to reach a host outside the MANET, i.e., the Internet host. The traffic to and from the Internet is distributed between base stations (APs) of the local network.

- **MEWLANA TD/RD**. In addition to on-demand and table driven routing protocols in MANETs, a novel ad hoc routing type called root driven routing is introduced [24]. Using this protocol type for networks where the intensity of inside traffic is negligible makes protocol efficient by eliminating the routing overhead. Main idea of this routing type is formation of a tree whose root is the foreign agent and branches are mobile nodes, and periodical initiation of this tree formation procedure by the root. To take into account these different cases, two protocols called MEWLANA-TD, which uses table driven routing type, and MEWLANA-RD, which uses root driven routing type, are designed. MEWLANA-TD uses the DSDV routing protocol, in which there is a trigger updating either periodically or when there is a change in the routing table. DSDV enables each node have an entry in their routing table for all other nodes. MEWLANA-RD uses tree based bidirectional routing (TBBR) as the routing protocol. TBBR is a special routing protocol designed only by using MIP entities, introducing low overhead at the expense of performance degradation. Two-tier MANET [32] shows a seamless roaming and load-balancing routing capability for providing Internet connectivity to the MANET. It modifies MIP to make traversing private networks, i.e., using NAT. It also proposes a load-balancing routing protocol to improve the Internet access quality by allowing mobile node dynamically changing their IGWs, thus relieving the bottleneck problem.

- **Hybrid MANET** [7] proposes an architecture to provide MANET nodes with Internet access using fixed IGWs and exploiting the mobility capability of additional mobile nodes (mobile IGWs). Since the Internet access for MANET nodes is provided via mobile IGWs, the quality of such service depends on the selection procedure used by MANET nodes to choose the most convenient mobile IGWs and register with. This work suggests using a hybrid criterion based on the weighted sum of the Euclidean distance between MANET nodes and mobile IGWs, and the load of mobile IGWs measured as the number of MANET nodes currently registered with them. MIP and DSDV are extended to integrate the suggested hybrid criterion.

- **WLAN & MANET** [43] presents a novel approach to integrate WLAN and MANET to the Internet (IPv6). The OLSR routing protocol is used within the MANET, and IGWs are used to connect the MANET to the Internet. In addition, the automatic mode-detection and the switching capability is also introduced in each MANET node to facilitate handoffs between WLAN and MANET. Mobility management across WLAN and MANET is achieved through MIPv6, which is integrated into the extended functionality of OLSR.

1.4.4 Discussion

A proactive approach to provide Internet connectivity to a MANET relies on ensuring that all nodes are registered with a foreign agent at all times. Mobile IP uses on link layer broadcasts to provide foreign agent information to interested nodes. However, these broadcasts can prove to be extremely expensive in a MANET where a broadcast means that packets are being flooded throughout the network.

In contrast, in a purely reactive approach, mobile nodes obtain foreign agent information by sending out agent solicitations only when data needs to be sent to a node outside the MANET. To limit the amount of flooding, these solicitations might be piggybacked on RREQ packets. An expanding ring search might also be used. In addition, intermediate nodes are allowed to reply with a route to the foreign agent, which reduces the overhead further.

The hybrid approach provides Internet access to MANETs while attempting to balance the proactive and reactive approaches, and it has many benefits. A proactive solution allows mobile nodes to find the foreign agent closest to them and enables better handoffs, which in turn leads to lower delay. Periodic registrations in such a proactive scheme help foreign agents track the mobility of the mobile node. However, if not all the nodes in the MANET require connectivity, the repeated broadcasting of agent advertisements and solicitations can have a negative impact on the MANET due to excessive flooding overhead. A hybrid approach combines the advantages of both approaches so that the required information is received in a timely fashion and the MANET's scarce resources are not further burdened with the MIP overhead.

Two strategies, default routes and tunneling, are usually used to integrate gateway forwarding with ad-hoc routing protocols. It is found that default routes that are adopted from traditional LAN settings need modifications to work in a multihop ad hoc environment. Despite these additions, default routes have problems with multiple gateways and inconsistent routing state. Establishing tunnels to the gateways, either half-tunneling or tunneling, on the other hand, provides an architecturally appealing solution and works well with multiple Internet gateways [48, 22, 23].

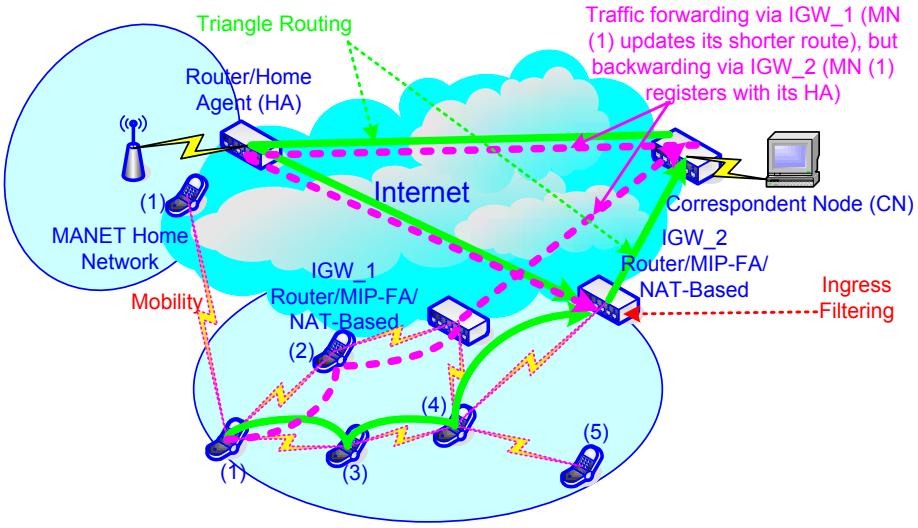

FIGURE 1.20
Issues on Internet Gateway Forwarding Strategies.

1.5 Issues and Solutions for Mobility Management

1.5.1 Mobility Management Issues in MANET

In this Section, we assume that MANET mobility management uses MIPv4 for the macro-mobility and ad hoc routing protocol for the micro-mobility. The data traffic sent from a MANET node to an Internet host is forwarded through IGWs using either the default route or the tunneling techniques [4, 23, 38]. There are also multiple IGWs, NAT devices, together with their ingress filtering policies between the MANET domain and the Internet. Figure 1.20 shows different scenarios on IGW forwarding strategies, which are used to illustrate for different discovered problems next.

1.5.1.1 Inconsistent Context due to Default Route Forwarding

1. **Type I**:

   In this forwarding mode, a MANET node (MN) sends packets to the Internet, i.e., communicate with the CN, using the default route. Whenever it associates with an IGW, it sets its default route pointing to this IGW. The default route is used to forward data packets to the unknown destination. In the scope of MANET, it usually means that the destination is not in the same MANET domain and can be located in the Internet, thus the data packet is forwarded to the IGW, where it is

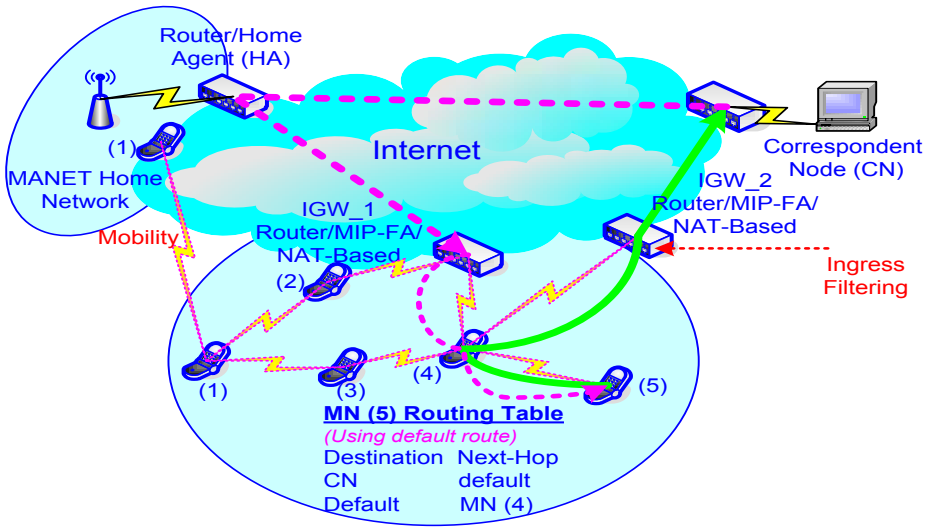

FIGURE 1.21
Inconsistent context due to default route forwarding: Type-I.

dropped if the destination is unreachable, or continued forwarding to the Internet if the destination is reachable via IGW.

MN (5) routing table using default route in Figure 1.21 shows an example, where MN (4) is used as the next-hop to the associated IGW_1. The problem with this default route setting is that MN (5) does not know its current associated IGW. Thus, the inconsistent context on sending packets from MN (5) to the Internet, e.g., via IGW_2, and receiving packets from CN to MN (5), e.g., via IGW_1, can terminate the 2-way connection like TCP. The next section shows an advanced setting to reduce this effect.

2. **Type II**:

   In this forwarding mode, a MANET node adds an additional entry into its routing table for the default route, indicating the current IGW that this MANET node associates. Figure 1.22 shows MN (5) routing table using extended default route. With this advanced setting, the ambiguous IGW association of MANET node is solved. However, an additional cost on storage and access time for the additional entry is introduced. Moreover, the inconsistent context problems are still existed in other scenarios, which are presented in the next section.

3. **Type III**:

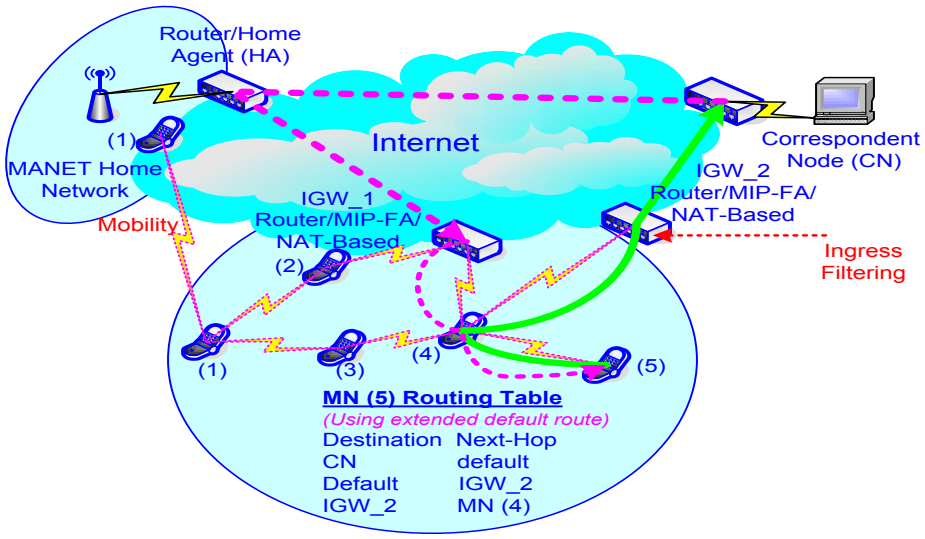

FIGURE 1.22
Inconsistent context due to default route forwarding: Type-II.

Different scenarios indicating the inconsistent context are analyzed in this section, which can be dependent on the operation of MANET routing protocols and IGW discovery methods (proactive vs. reactive vs. hybrid) [4, 35, 80].

- **Scenario I**: A MANET node updates a shorter route to another IGW without re-registering the new IGW with its home agent (HA), and/or its foreign agent (FA), as well as the NAT device located in the visiting MANET domain.

  Figure 1.23 shows an example, where MN (1) moves from its home network to the new MANET domain, registering to its home agent (HA) via IGW_2 through MN (3) as the next-hop to the IGW_2. The distance from MN (1) to IGW_2 is 3-hop. Later, MN (1) finds a shorter route to the Internet via IGW_1. MN (1) chooses MN (2) as the next-hop to the IGW_1 and the hop-count is 2. In this scenario, traffic from MN (1) to the CN is forwarded via IGW_1. However, traffic from the CN to MN (1) is still forwarded via IGW_2 since it is registered by MN (1) to MN (1) HA. This creates the inconsistent context.

- **Scenario II**: A MANET node associated with an IGW, e.g., IGW_1, forwards agent advertisement packet for another IGW, e.g., IGW_2. As a result, its downstream nodes can associate to the IGW_2 using it as the next-hop. However, the traffic is actually forwarded through IGW_1.

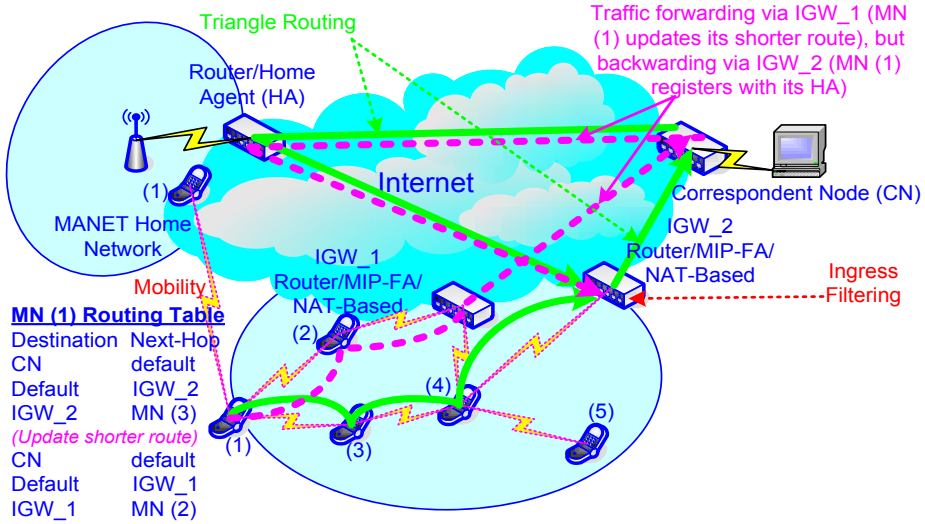

FIGURE 1.23
Inconsistent context due to default route forwarding: Type-III (scenario I).

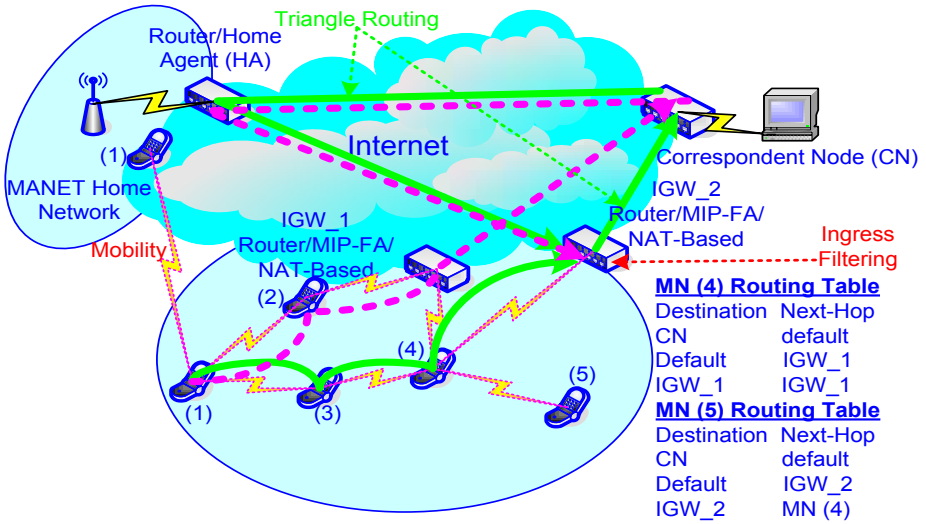

FIGURE 1.24
Inconsistent context due to default route forwarding: Type-III (scenarios II-III).

As an example, in Figure 1.24, suppose that MN (4) is currently associated with IGW_1 and MN (5) is not associated with any IGW. In the proactive IGW discovery, the IGW will broadcast periodically its agent advertisement packets. When MN (4) receives IGW_2 agent advertisement packets, MN (5) will set MN (4) as the next-hop to IGW_2 in its default route, if MN (4) continues forwarding IGW_2's agent advertisement packets. However, whenever the traffic to CN is generated by MN (5), and then forwarded to IGW_2 via MN (4), this traffic is actually forwarded via IGW_1 by MN (4). This creates the inconsistent context. Note that the decision of MN (4) whether or not to continue forwarding other IGWs packets, e.g., agent advertisement, is dependent on the corresponding operation and implementation of MANET routing protocol and IGW discovery method.

- **Scenario III**: A MANET node looses its association to the current IGW, e.g., a detection of broken link, and re-associates to another IGW. As a result, traffic to Internet from its downstream nodes choosing it as the next-hop to the current IGW will be forwarded via another IGW instead.

  As an example, in Figure 1.24, suppose MN (3) chooses MN (4) as its next-hop to the IGW_1 and MN (4) is currently associated to the IGW_1. Later, if the link between MN (4) and IGW_1 is broken, MN (4) re-associates to IGW_2. Thus, traffic to the Internet from MN (3) will be forwarded via IGW_2, though MN (3) thinks it is forwarded via IGW_1. This creates the inconsistent context.

1.5.1.2 Inconsistent Context in MIPv4-FA Triangle Routing

The use of MIPv4 can easily lead to triangle routing problem, see Figure 1.24, i.e., traffic from a MANET node to the correspondent node (CN) is sent directly, while return traffic is sent to the MANET node home agent (HA), then is tunneled to MANET node's foreign agent (FA) and delivered to the MANET node.

On providing Internet connectivity for MANET nodes, this also means that traffic to the Internet from a MANET node can be forwarded to one IGW, e.g., IGW_1 in Figure 1.24, for a shorter route, while return traffic is forwarded through registered IGW, e.g., IGW_2 in Figure 1.24. This creates the inconsistent context. In the case IGW_2 is behind a stateful firewall, i.e., the first outbound to Internet packet sets the soft state in the firewall for the return packets to enter the MANET domain, the connection will be terminated. Note that triangle routing problem can be considered as a consequence of Scenario I, thus solution for this problem can be referred in Scenario I, Section 1.5.2.

1.5.1.3 Inconsistent Context in MIPv4-FA Ingress Filtering

Ingress filtering means that a router/firewall will not accept on its ingress interface packets with a source IP address that is not topologically correct for that interface. The motivation is to prevent IP address spoofing. If the ingress filtering is integrated into IGWs, the traffic to the Internet from a MANET node ends up at another IGW than the IGW it is registered with, that IGW can drop its packet.

As an example, in Figure 1.24, MN (1) moves to a new MANET domain, getting a new topologically correct care-of address (CoA) and registering it with its HA via IGW_2. If MANET node updates later its shorter route through IGW_1, e.g., using another IP subnet, its outbound packets can be dropped at IGW_1 due to ingress filtering. Note that ingress filtering problem can be also considered as a consequence of Scenario I, hence solution for this problem can be referred in Scenario I, Section 1.5.2.

1.5.1.4 Inconsistent Context in MIPv4-FA NAT Traversing

If the IGWs connecting MANET domains with the Internet are either behind the NAT devices, or integrating the NAT function, and private IP subnets are used with MANET domains, a MANET node communicates with an Internet host through the private IP address of its associated IGW, which will be mapped into another public IP address. Thus, with a NAT-based IGW, the connection of a MANET node is bound to the NAT that the connection passes through. However, whenever the MANET node updates its new NAT-based IGW, the return traffic can be dropped at this one due to the stateful router/firewall. This creates the inconsistent context. Solution for this problem is shown in Section 1.5.2.3.

1.5.1.5 Cascading Effect for Node Location Determination

On reactive ad-hoc routing protocol, whenever a route to a destination is needed, source MANET node needs to perform a route discovery. Usually, it broadcasts a route request (RREQ) packet and waiting for a route reply (RREP), if the destination is located in the same MANET. On providing Internet connectivity for MANET nodes using the default routes, no RREP will be sent back to the source MANET node if the destination is an Internet host. Note that for some mechanisms, it is allowed for an IGW to send a RREP to the source MANET node to indicate that the destination is an Internet host and can be reached through this IGW, called proxy RREP.

If the destination is an Internet host, the source MANET node will send packets to the destination using its default route setting in its routing table. However, when its upstream MANET node receives its packets, its upstream MANET node again performs another route discovery for the destination. This route discovery is repeated at each upstream MANET node on the chain from the source MANET node to the corresponding IGW, creating a problem

called cascading effect. The advantage of cascading effect on MANET reactive routing protocols is that the shortest host route to the destination is found if the destination is located within the same MANET domain. However, this mechanism is redundant and tolerates much overhead if the destination is an Internet host.

Therefore, cascading effect problem should be removed on providing Internet connectivity for MANET nodes. One solution is to insert directly in the routing table of each MANET node an entry for each destination Internet host. When a relay MANET node receives data packets from its downstream nodes, it continues forwarding to the destination Internet host without any route discovery, if an entry for that destination Internet host is available in its routing table.

1.5.2 Mobility Management Solutions in MANET

Since the inconsistent context problems due to the MIPv4-FA triangle routing and ingress filtering are consequences of Scenario I, (cf. Section 1.5.1), and the solution for cascading effect problem has been already presented, this Section continues with only solutions for the inconsistent context due to default route forwarding. The solution for the inconsistent context due to MIPv4-FA NAT traversing is presented in Section 1.5.2.3.

1.5.2.1 Reducing Inconsistent Context of Using Default Route

- **Scenario I:**

  A MANET node is not allowed to update its shorter route to another IGW, unless its current transmissions on any 2-way connections to the Internet hosts are finished, and it has already re-registered this new IGW with its home agent. This re-registration can be prepared in advance, e.g., during the data transmissions on the current connections via the old IGW. Clearly, this rule removes completely the inconsistent context in Scenario I since the inbound/outbound traffic to/from Internet is always forwarded via the same IGW. However, its disadvantage is that intra-MANET route from the source MANET node to the registered IGW is not an optimal one, e.g., the shortest route.

  This rule also reduces the inconsistent context in Scenario II and Scenario III. The reason is that, the less changing on the IGW re-association of MANET nodes, the less inconsistent context problems appear on their downstream MANET nodes.

- **Scenario II:**

  On proactive IGW discovery, a MANET node that does not register to any IGW is allowed to re-broadcast the received agent advertisement if it decides to register with this IGW. Otherwise, the re-broadcasting

of agent advertisement is prohibited. On re-active IGW discovery, a MANET node is allowed to generate/forward an agent advertisement in one of the following three cases:

i) It does not register to any IGW, registering it to the IGW of which the agent advertisement it receives, then forwarding the agent advertisement to the source MANET node.

ii) It has already registered to an IGW, receiving the agent advertisement generating by the same IGW, forwarding this agent advertisement to the source MANET node.

iii) It has already registered to an IGW, receiving the agent solicitation from the source MANET node, generating itself an agent advertisement to the source MANET node.

iv) On hybrid IGW discovery, the above rules are applied whenever an agent advertisement or an agent solicitation is received.

Clearly, the above rules ensure that all the MANET nodes on the chain from any source MANET node to their registered IGWs uses the same IGW for inbound/outbound traffic to/from the Internet. However, on the proactive IGW discovery, the applied rule creates non-overlapped MANET domains, each domain associates to only one IGW. Thus, it does not take the advantage of using multiple IGWs for load balancing and fault-tolerance. Moreover, there can be the appearance of orphan MANET nodes due to the collision or high mobility.

On reactive or hybrid IGW discovery, the generation of agent advertisement of an IGW or an immediate MANET node, called gratuitous agent advertisement or proxy route reply (RREP), if this agent advertisement is piggybacked on RREP of MANET reactive routing protocol, to the source MANET introduces non-optimal route. This happens whenever the destination is located in the same MANET domain of the source MANET node, and the source MANET node receives a proxy RREP before the normal RREP sent by the destination, e.g., due to the collision or longer hop-count. This problem can be further reduced by the use of the destination sequence number, in which this field on the proxy RREP packet is always set to a small value, e.g., "0". Thus, when the source MANET node receives the RREP sent by the destination later, it will updates the host route to the destination instead of the default route via the IGW, since the destination sequence number sent by the destination is greater than that of proxy RREP sent by the IGW.

- **Scenario III:**

There should be a mechanism for the MANET node to detect the broken link, sending this information to its downstream MANET nodes so that

these MANET nodes can re-register their new IGWs with their home agents. This mechanism is MANET routing protocol dependent.

As an example, in AODV) routing protocol, a MANET node keeps a list of other neighbor MANET nodes using it as the next-hop to a set of destinations, called precursor list. Whenever this MANET node detects a broken link to any destination, it searches its precursor list for that destination and sends a route error (RERR) to all the nodes on this list. This process is propagated to the MANET nodes in its precursor list.

In the scope of this scenario, a MANET node detecting a broken link to its registered IGW will integrate this information to the RERR and send it to all nodes in its precursor list. However, to reduce also the inconsistent context in the Scenario II, the detected MANET node also sends the RERR to all the precursor lists associated to all IGWs as the destinations. This is because some MANET nodes can be associated with an IGW, but their traffic can be forwarded to another IGW instead.

1.5.2.2 Removing Inconsistent Context Using Tunneling

Clearly, on IGW forwarding strategies using default route between a MANET domain and the Internet connecting through multiple IGWs, the inconsistent context is always a problem. This is because the traffic from the MANET to the Internet and the returned traffic can be forwarded through different IGWs, taking to the termination of 2-way connections.

With the described solutions on the previous section, the problems on default route are reduced. However, depending on the implementation of MANET routing protocols, the IGW discovery, whether they are implemented independently or dependently, there can be the appearance of another scenarios of inconsistent context. Thus, it is the purpose of this part to introduce another approach to remove completely the effect of inconsistent context.

The idea is that a MANET node is always sent its outbound traffic to the Internet via its registered IGW, irrespective its immediate MANET nodes updating the shorter routes to other IGWs. This is achieved via the use of tunneling, e.g., IP-in-IP tunneling [48, 23]. In this scheme, the source MANET node encapsulates the original IP packet to another IP packet, in which its registered IGW is the destination IP address in the outer IP header. The encapsulated IP packet is then forwarded to its registered IGW using the MANET routing protocol. When the encapsulated IP packet arrives to the IGW, it is decapsulated and the inner original IP packet is forwarded to destination Internet host by IGW.

Figure 1.25 shows an example, where an IP packet to the Internet from the source MANET node MN (1) is first tunneled, i.e., encapsulated, to its registered IGW_2, then it is decapsulated and the original IP packet is forwarded to the CN. In the reversed direction using tunneling through MN (1) HA in this scenario, CN sends its packet to the MN (1) home address (HAddr) via

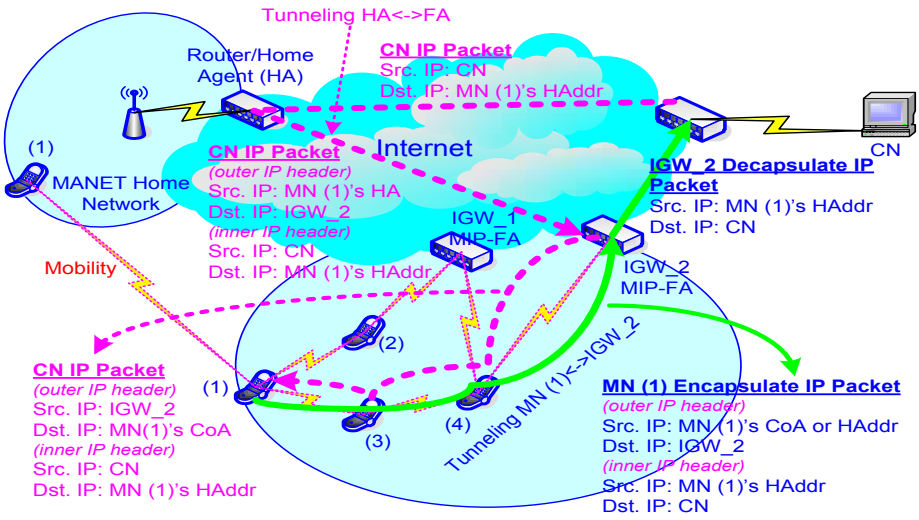

FIGURE 1.25
Packet header passing MIP-FA and IGW using tunneling.

MN (1) HA, where the packet is encapsulated and forwarded to MN (1) FA (IGW_2). At the IGW_2, this packet is decapsulated, then encapsulated and forwarded to the MN (1).

With the tunneling, the inconsistent context problems are completely removed. However, an additional cost of adding an outer IP header is introduced.

1.5.2.3 Solutions on Passing the NAT Device

Mobile IP with NAT traversal can be used to pass the NAT-based IGW. The NAT traversal is based on MIP UDP tunneling mechanism [49]. In MIP UDP tunneling, the mobile node may use an extension in its registration request to indicate that it is able to use MIP UDP tunneling instead of standard MIP tunneling if the home agent sees that the registration request seems to have passed through a NAT. The home agent may then send a registration reply with an extension indicating acceptance or denial. After assent from the home agent, MIP UDP tunneling will be available for use for both forward and reverse tunneling. UDP tunneled packets sent by the MANET node use the same ports as the registration request message. In particular, the source port may vary between new registrations, but remains the same for all tunneled data and re-registrations. The destination port is always 434. UDP tunneled packets sent by the home agent uses the same ports, but in reverse.

To reduce the inconsistent context problems, this solution requires an additional IP-in-IP tunneling from the source MANET node to its registered IGW,

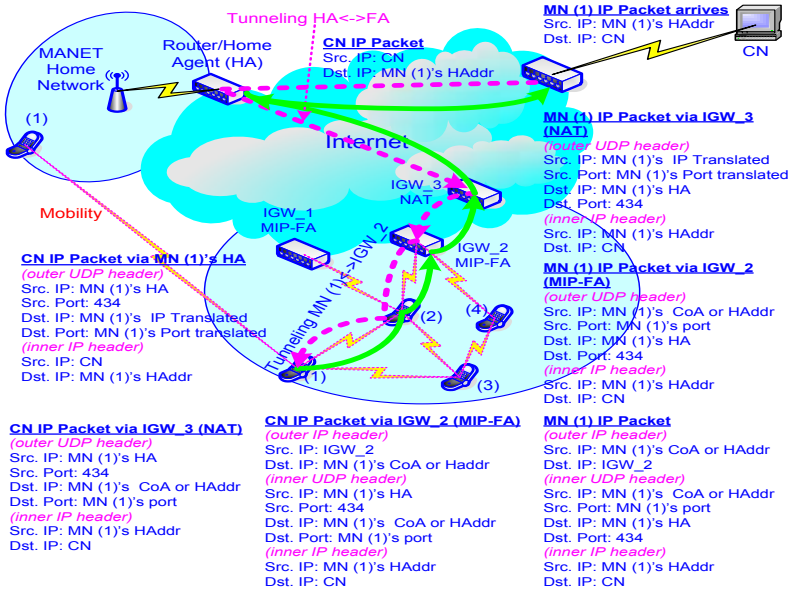

FIGURE 1.26
Packet header passing both MIP-FA and NAT-based IGWs.

to ensure that the source MANET IP address and source MANET UDP port number are translated consistently by the same NAT-based IGW.

Figure 1.26 shows an example, where the tunneling is established between the MN (1) and its registered IGW_2, and between the MN (1) MIP-FA (mapping to publish IP address of IGW_3) and MN (1) HA. MIP UDP tunneling is used to transmit packet through the NAT-based IGW_3, while the tunneling between MN (1) and IGW_2 is used to remove the inconsistent context problems. Note that in Figure 1.26, MN (1) IP packet is used to indicate the direction of sending data packets from source MANET node to the destination CN, while CN IP packet is used for indicating the packet transmissions in the reversed direction, respectively.

1.6 Mobility Management Open Issues

Major problems of existing work in IP mobility management for all-IP mobile networks are either the lack of, or little focus on:

i) a standardized interface for cross-layer, especially L3/L2, interaction to shorten the handoff latency and reduce the signaling overhead,

ii) the convergence towards an open architecture to support different mobility scenarios under both horizontal and vertical handoffs [4] in heterogeneous infrastructures and wireless access technologies,

iii) inter-operation between existing mobility protocols and new developed protocols to either improve further the performance (e.g., handoff latency, packet delivery) or reduce signaling overhead,

iv) location privacy,

v) implementation of protocols, network architectures, and testbeds for critical evaluations of different mobility metrics.

These problems are research challenges in the mobility management for next-generation all-IP mobile networks, and will be discussed next, together with the possibility of directions towards the solutions. Security is also a critical issue and should be included in the mobility management, but it is out of scope in this chapter.

1.6.1 Standardized Interface for Cross-Layer Interaction

To fulfil the required quality of service (QoS) in mobility, e.g., handoff latency, packet delivery ratio, signaling overhead, the cross-layer interaction, especially at L3/L2, is needed. However, existing work in general IP mobility management (Section 1.2) does not define any standardized interface or API for cross-layer interaction, and thus, functionalities for cross-layer interaction developed in WLAN for faster handoff (Section 3) are not compatible with those functionalities provided by other work. Therefore, the lack of a standardized interface for cross-layer interaction decreases the portability of existing mobility management protocols, especially for vertical handoff. Among ongoing work to fill in this gap, IEEE 802.21 or *Media-Independent Handover* (MIH) services [3, 74, 58, 44] is the best candidate.

The IEEE 802.21 MIH services aims at providing functionalities to assist with seamless handoffs between heterogeneous link-layer technologies, and across IP subnet boundaries. MIH services can be delivered through L2 specific solutions and/or through a L3 or above protocol. The IEEE 802.21/MIH advantages for mobility are multi-fold. Firstly, it enables seamless handovers across different access technologies. Secondly, it optimizes L3 handover, e.g., MIP or PMIP, by using L2 triggers associated with events to provide early warning of impending handovers and thereby decrease handover latency. Thirdly, it provides QoS continuity across different technologies and minimizes service interruption. Finally, it provides the ease of implementation through: i) thin software client on terminals, ii) no radio access network

[4] A horizontal handoff is a handoff between two network access points that use the same network technology and interface. A vertical handoff is a handoff between two network access points, which are usually using different network connection technologies.

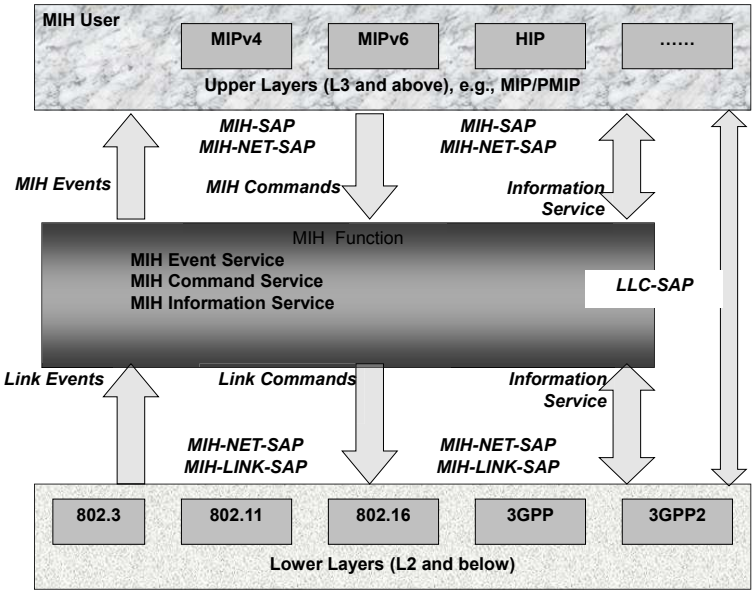

FIGURE 1.27
Media-Independent Handover (MIH) Reference Model.

modifications required, and iii) support either network or client-controlled handovers.

The IEEE 802.21 architecture consists of MIH users located above L4 that use MIH service access points (SAPs) to communicate with MIH services at lower layers. MIH users are the initiation and termination points for MIH signaling sessions. The MIH Function (MIHF), located between L2 and L3, provides handover services, including the event service (ES), information service (IS), and command service (CS), through SAPs that are defined by the IEEE 802.21 Working Group. Figure 1.27 present the standardized interfaces for cross-layer interaction in IEEE 802.21 MIH. The scope of IEEE 802.21 includes only the operation of MIHF and the primitives associated with the interfaces between MIHF and other entities. A single media-independent interface between MIHF and MIH user (MIH_SAP) is sufficient. On the other hand, there is a need for defining a separate technology dependent interface, which is specific to the corresponding media type supported, between the MIHF and the lower layers (MIH_LINK_SAP). The primitives associated with the MIH_LINK_SAP enable MIHF to receive timely and consistent link information and control link operation during handovers. Besides these, IEEE 802.21 specifies a media-independent SAP (MIH_NET_SAP), which provides transport services for L2 and L3 MIH message exchange with remote MIHFs. Functions over the LLC_SAP are not specified in IEEE 802.21.

However, there are still open problems in IEEE 802.21 MIH services. Firstly, reliable delivery is one of the key requirements to be imposed by MIH users, and thus, MIH signaling messages should be carried over a reliable transport protocol. Secondly, the scope of current specifications of IEEE 802.21 MIH is restricted to access technology-independent handovers. Intra-technology handovers, handover policies, security mechanisms, media-specific L2 enhancements to support IEEE 802.21, and L3 or upper-layer enhancements are outside the scope of IEEE 802.21.

Some ongoing work has recently focused on the reliable delivery of IEEE 802.21 MIH signaling messages. The IETF mobility for IP: performance, signaling and handoff optimization (MIPSHOP) has been working on developing a protocol for transport of MIH services information and mechanisms for discovering the MIH server. In particularly, a mobility services framework design (MSFD) is described in [58] for the IEEE 802.21 MIH protocol, which addresses mechanisms for mobility service discovery and transport layer mechanisms for the reliable delivery of MIH messages.

1.6.2 Convergence towards an Open Architecture

Up to now, the existing solutions of mobility management have mainly dealt with user handovers between wireless access points in an operator-controlled infrastructure. In the emerging wireless network scenarios with heterogeneous user devices and wireless technologies, the term mobility should have a wider sense and involve system responses to any changes in the user and network environments, including changes in radio and network resources as well as commercial conditions. Furthermore, mobility solutions need to support a simple-to-use, affordable, anytime-anywhere network access, so that mobile users will soon take for granted a rich set of services regardless of the underlying connectivity, in the forthcoming eSociety [21]. Therefore, a single mobility paradigm can not cover all this diverse set of requirements. Instead, a set of solutions, which can be flexibly combined and integrated on demand into an open architecture, is needed.

1.6.3 Ambient Networks

Ambient Networks project [6] is one of a few work, which addresses above challenges by developing innovative network solutions for increased competition and cooperation in an environment with a multitude of access technologies, operators, and business actors. The project aims at offering a complete, coherent open architecture solution based on dynamic composition, which provides access through the instant establishment of inter-network agreements, using common control functions to a wide range of different applications and access technologies, enabling integrated, scalable, and transparent control of network capabilities.

In particularly, three new concepts have been introduced in [66]. The mo-

bility toolbox is the first of these concepts, where different, both legacy and new, mobility components can be slotted together as needed. Components can be optimized to suit particular handover scenarios, such as the movement of applications between devices, mobility of networks, and handover in more complex, multi-domain environments between different administrative domains. The toolbox allows for new mobility services, which to a large extent, rely on the ambient network composition concept. The second concept is the separation of mobility-relevant triggers from the actual mobility management mechanism, which allows processing of a wide range of mobility-relevant triggers generated by virtually any part of the system. The third concept is the development of mobility functions for groups of nodes, which provide new services for network mobility and optimize existing solutions in terms of consumed resources and increased handover performance.

1.6.4 Inter-operation

Another research approach is to exploit the advantages of existing mobility management protocols to either improve or optimize further the protocol operations for different objectives: i) shorter handoff latency, ii) higher packet delivery ratio, iii) lower signaling overhead, to list a few. Such an inter-operation work include: i) PMIPv6/MIPv6 [26] for a global IP mobility management solution, ii) fast handovers for PMIPv6 [93] adapts FMIPv6 for PMIPv6 to shorten the handoff latency and reduce the packet loss, iii) NEMO/MANET [57] use NEMO to support the global reachability of MANET nodes, and iv) IEEE 802.21/MANET [50] is based on the IEEE 802.21 MIH services to develop an integrated framework for seamless soft handoff in ad hoc networks.

1.6.5 Location Privacy

While many IP mobility management protocols have been developed for wireless mobile networks, the location privacy has not been addressed much. The reason is that IP addresses of mobile nodes, and related entities, e.g., home agents or access routers or foreign agents, are often used as topological contact identifiers to establish and maintain the on-going connections. Since IP routing is based on the IP network prefix, which also shows the network location, an IP mobility management protocol exploited IP addresses as identifiers usually does not provide the location privacy. However, location privacy is one of the user requests in 4G mobile networks. A current approach to the location privacy is to enable a separation between the end-point identifier and locator role of IP address, and thus, overcomes one of paradigms of the current Internet architecture - the association of a (possible temporary) IP address with a host end-point identifier. Such examples in this approach include: identifiers separating and mapping scheme (ISMS) [19], and locator/ID separation protocol [15].

The IETF has also working on host identity protocol (HIP), in which all

occurrences of IP addresses in applications are eliminated and replaced with cryptographic host identifiers. However, the above approach to location privacy still has several unsolved aspects in the area of mobility management: i) a NAT traversal solution, ii) a description on the interactions between legacy location-unaware vs. location-aware applications, iii) how to build location-aware-based overlays, iv) how to match location privacy properties to the requirements of existing standards, and v) the support from existing infrastructure such as DNS. A possible solution to some of above problem can be exploited from Internet indirection infrastructure [34].

1.6.6 Implementation and Testbed

Many IP mobility management protocols are proposed without evaluation, and thus, there has been still the lack of wireless mobile network testbed, together with the qualitative/quatitative analysis and implementation of protocols for evaluation, testing, and further improvements in solutions for problems. Recently, a few work has been concentrated on these issues. These work include: i) quantitative analysis modeling for more comparisons and evaluations [40, 27, 52]; ii) GNU/Linux implementation of IEEE 802.21 [75]; and iii) on-going WEIRD project [90], which aims at implementing European national research network testbed using the WiMAX technology.

1.7 Conclusion

This book Chapter has presented and discussed on different scopes of mobility in all-IP mobile networks: i) from a general IP mobility management to the specific mobility management in 1-hop WLAN and n-hop MANET, ii) from micro or localized mobility to macro or global mobility, iii) from the host-based mobility with IP stack modification in mobile node to the network-based mobility without IP stack modification in mobile node, and iv) from individual host mobility to the whole network mobility.

Bibliography

[1] 3GPP. Setting Standards for Mobile Broadband, `http://www.3gpp.org/`.

[2] IEEE Std 802.1f. Recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting IEEE 802.11 operation. *IEEE Draft 802.1f/D3*, 2002.

[3] IEEE Std 802.21. IEEE standard for local and metropolitan area networks - Part 21: media independent handover services. *IEEE Std 802.21-2008*, 2009.

[4] F. M. Abduljalil and S. K. Bodhe. A survey of integrating IP mobility protocols and mobile ad hoc networks. *IEEE Communication Surveys and Tutorials*, pages 14–30, 2007.

[5] C. Ählund and A. B. Zaslavsky. Software solutions to Internet connectivity in mobile ad hoc networks. *Proc. of 4th International Conference on Product Focused Software Process Improvement*, pages 559–571, 2002.

[6] Ambient Network. `http://www.ambient-networks.org/index.html`.

[7] H. Ammari and H. E. Rewini. Using hybrid selection schemes to support QoS when providing multihop wireless Internet access to mobile ad hoc networks. *Proc. of 1st International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, pages 148–155, 2004.

[8] E.M. Belding-Royer, Y. Sun, and C.E. Perkins. Global connectivity for IPv4 mobile ad hoc networks. IETF Internet Draft draft-royer-manet-globalv4-00.txt, 2001.

[9] Minet P. Agha K. A. Adjih C. Benzaid, M. and G. Allard. Integration of mobile-IP and OLSR for a universal mobility. *Wireless Networks*, pages 377–388, 2004.

[10] C. Bernardos and M. Calderon. Survey of IP address autoconfiguration mechanisms for MANET. IETF Internet Draft draft-bernardos-manet-autoconf-survey-00.txt, 2005.

[11] Maltz D. A. Broch, J. and D. B. Johnson. Supporting hierarchy and heterogeneous interfaces in multi-hop wireless ad hoc networks. *Proc. of the International Symposium on Parallel Architectures, Algorithms and Networks*, pages 370–375, 1999.

[12] A. T. Campbell and J-G. Castellanos. IP micro-mobility protocols. *ACM Mobile Computing and Communications Review*, pages 45–53, 2000.

[13] Joshi R. C. Chand, N. and M. Misra. Cooperative caching in mobile ad-hoc networks based on data utility. *Mobile Information Systems*, pages 19–37, 2007.

[14] C. Y. Chiu and C. Gen-Huey. A stability aware cluster routing protocol for mobile ad hoc networks. *Wireless Communications and Mobile Computing*, pages 503–515, 2003.

[15] N. Choi, T. You, J. Park, T. Kwon, and Y. Choi. ID/LOC separation network architecture for mobility support in future Internet. *Proc. of the 11th International Conference on Advanced Communication Technology*, pages 78–82, 2009.

[16] T. Clausen, C. Dearlove, and P. Jacket. The optimized link-state routing protocol version 2. IETF Internet Draft draft-ietf-manet-olsrv2-02.txt, 2006.

[17] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert. Network mobility NEMO support protocol. IETF RFC 3963, 2005.

[18] S. Ding. A survey on integrating MANETs with the Internet: challenges and designs. *Computer Communications*, pages 3537–3551, 2008.

[19] P. Dong, H. Zhang, H. Luo, T-Y. Chi, and S-Y. Kuo. A network-based mobility management scheme for future Internet. *Computers and Electrical Engineering*, pages 291–302, 2009.

[20] A.L. Dul. Global IP network mobility using border gateway protocol. *Connexion by Boeing Company, http://www.quark.net/docs/Global_IP_Network_Mobility_using_BGP.pdf*, March 2006.

[21] e-Society. `http://www.york.ac.uk/res/e-society/`.

[22] P. E. Engelstad and G. Egeland. NAT-based Internet connectivity for on-demand ad hoc networks. *Proc. of the 1st IFIP TC6 Working Conference on Wireless On-Demand Network Systems*, pages 344–358, 2004.

[23] P. E. Engelstad, A. Tonnesen, A. Hafslund, and G. Egeland. Internet connectivity for multi-homed proactive ad hoc networks. *Proc. of International Conference on Communications (ICC)*, pages 4050–4056, 2004.

[24] M. Ergen and A. Puri. MEWLANA-mobile IP enriched wireless local area network architecture. *Proc. of the 56th IEEE Vehicular Technology Conference (VTC)*, pages 2449–2453, 2002.

[25] N. A. Fikouras, A. J. Könsgen, and C. Görg. Accelerating mobile IP handoffs through link-layer information, an experimental investigation with 802.11b and Internet audio. *Proc. of International Multiconference on Measurement, Modelling, and Evaluation of Computer-Communication Systems*, 2001.

[26] G. Giaretta. Interactions between PMIPv6 and MIPv6: scenarios and related issues. IETF Internet Draft draft-ietf-netlmm-mip-interactions-04.txt, 2009.

[27] D. Griffith, R. Rouil., and N. Golmie. Performance metrics for IEEE 802.21 media independent handover MIH signaling. *Wireless Personal Communications*, pages 537–567, 2008.

[28] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy Mobile IPv6. IETF RFC 5213, 2008.

[29] E. Gustafsson, A. Jonsson, and C. Perkins. An end-to-end approach for transparent mobility across heterogeneous wireless networks. IETF Internet Draft draft-ietf-mobileip-reg-tunnel-04.txt, 2001.

[30] R. Hsieh and A. Seneviratne. A comparison of mechanisms for improving mobile IP handoff latency for end-to-end TCP. *Proc. of 9th International Conference on Mobile Computing and Networking*, pages 29–41, 2003.

[31] R. Hsieh, Z.G. Zhou, and A. Seneviratne. S-MIP: a seamless handoff architecture for mobile IP. *Proc. of the 22nd IEEE INFOCOM*, pages 1774–1784, 2003.

[32] Y.Y. Hsu, Y.C. Tseng, C.C. Tseng, C.F. Huang, J.H. Fan, and H.L. Wu. Design and implementation of two-tier mobile ad hoc networks with seamless roaming and load-balancing routing capability. *Proc. of 1st International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, pages 52–58, 2004.

[33] X. Hu, L. Li, Z.M. Mao, and Y.R. Yang. Wide-area IP network mobility. *Proc. of the 27th IEEE INFOCOM*, pages 1624–1632, 2008.

[34] Internet Indirection Infrastructure. http://i3.cs.berkeley.edu/.

[35] X. Jin and B. Christian. Wireless multihop Internet access: gateway discovery, routing, and addressing. *Proc. of the 3rd Generation Wireless and Beyond*, 2002.

[36] D. Johnson, Y. Hu, and D. Maltz. The dynamic source routing (DSR) protocol for mobile ad hoc networks for IPv4. IETF RFC 4728, 2007.

[37] D. Johnson, C. Perkins, and J. Arkko. Mobility Support in IPv6. IETF RFC 3775, 2004.

[38] U. Jönsson, F. Alriksson, T. Larsson, P. Johansson, and G. Q. Maguire. MIPMANET mobile IP for mobile ad hoc networks. *Proc. of the 1st ACM Interational Symposium on Mobile Ad Hoc Networking and Computing*, pages 75–85, 2000.

[39] H.Y. Jung, S.J. Koh, and J.Y. Lee. Fast Handover for Hierarchical MIPv6 (FHMIPv6). IETF Internet Draft draft-jung-mobopts-fhmipv6-00.txt, 2005.

[40] K-S. Kong, Y-H. Han, M-K. Shin, and You H. Mobility management for all-IP mobile Networks: mobile IPv6 vs. proxy mobile IPv6. *IEEE Wireless Communications*, pages 36–54, 2008.

[41] R. Koodli. Fast Handovers for Mobile IPv6. IETF RFC 4068, 2005.

[42] R. Koodli and C. Perkins. Mobile IPv4 fast handovers. IETF Internet Draft draft-ietf-mip4-fmipv4-07.txt, 2007.

[43] L. Lamont, M. Wang, L. Villasenor, T. Randhawa, and S. Hardy. Integrating WLANs & MANETs to the IPv6 based Internet. *Proc. of the IEEE International Conference on Communications (ICC)*, pages 75–85, 2003.

[44] G. Lampropoulos, A. K. Salkintzis, and N. Passas. Media-independent handover for seamless service provision in heterogeneous networks. *IEEE Communications Magazine*, pages 64–71, 2008.

[45] D. Le, X. Fu, and D. Hogrefe. A review of mobility support paradigms for the Internet. *IEEE Communication Surveys & Tutorials*, pages 38–51, 2006.

[46] Q. Le-Trung, P. E. Engelstad, V. Pham, T. Skeie, A. Taherkordi, and F. Eliassen. Providing internet connectivity and mobility management for MANETs. *International Journal of Web Information Systems*, pages 239–263, 2009.

[47] Q. Le-Trung, P. E. Engelstad, T. Skeie, and A. Taherkordi. Load-balance of intra/inter-MANET traffic over multiple Internet gateways. *Proc. of the 6th International Conference on Advances in Mobile Computing and Multimedia*, pages 50–57, 2008.

[48] Q. Le-Trung and G. Kotsis. Reducing problems in providing Internet connectivity for mobile ad hoc networks. *Proc. of the 4th International Workshop of the EuroNGI/EuroFGI Network of Excellence on Wireless Systems and Mobility in Next Generation Internet*, pages 113–127, 2008.

[49] H. Levkowetz and S. Vaarala. Mobile IP NAT/NAPT traversal using UDP tunneling. IETF Internet Draft draft-ietf-mobileip-nat-traversal-07.txt, 2007.

[50] J.H. Li, S. Luo, S. Das, and *et al.* An integrated framework for seamless soft handoff in ad hoc networks. *Proc. of the IEEE Military Communications Conference (MILCOM)*, pages 1–7, 2006.

[51] Y. Liao and L. Gao. Practical schemes for smooth MAC layer handoff in 802.11 wireless networks. *Proc. of the International Symposium on World of Wireless, Mobile and Multimedia Networks*, pages 181–190, 2006.

[52] C. Makaya and S. Pierre. *IEEE Transactions on Wireless Communications*, 7(2).

[53] K.E. Malki. Low Latency Handoffs in Mobile IPv4. IETF Internet Draft draft-ietf-mobileip-lowlatency-handoffs-v4-11.txt, 2005.

[54] K.E. Malki and H. Soliman. Simultaneous bindings for mobile IPv6 fast handoffs. IETF Internet Draft draft-elmalki-mobileip-bicasting-v6-00.txt, 2001.

[55] T. Manodham and T. Miki. A novel AP for improving the performance of wireless LANs supporitng VoIP. *Journal of Networks*, pages 41–48, 2006.

[56] P. McCann. Mobile IPv6 fast handovers for 802.11 networks. IETF RFC 4260, 2005.

[57] B. McCarthy, C. Edwards, and M. Dunmore. Using NEMO to support the global reachability of MANET nodes. *Proc. of IEEE INFOCOM*, pages 2097–2105, 2009.

[58] T. Melia, G. Bajkp, S. Das, N. Golmie, and J.C. Zuniga. IEEE 802.21 mobility services framework design (MSFD). IETF Internet Draft draft-ietf-mipshop-mstp-solution-12.txt, 2009.

[59] A. Mishra, M. Shin, and W. Arbaugh. An empirical analysis of the IEEE 802.11 MAC layer handoff process. *ACM SIGCOMM Computer Communication Review*, pages 93–102, 2003.

[60] A. Mishra, M. Shin, and W. Arbaush. Context caching using neighbor graphs for fast handoffs in a wireless network. *Proc. of IEEE INFOCOM*, pages 351–361, 2004.

[61] G. Mona, H. Philipp, P. Christian, F. Vasilis, and A. Hamid. Performance analysis of Internet gateway discovery protocols in ad hoc networks. *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 120–125, 2004.

[62] J. Montavont and T. Nöel. IEEE 802.11 handovers assisted by GPS information. *Proc. of the 2006 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 166–172, 2006.

[63] T. Narten, E. Nordmark, W. Simpson, and H. Soliman. Neighbour discovery for IP version 6 (IPv6). IETF Internet Draft draft-ietf-ipv6-2461bis-02.txt, 2005.

[64] E. Natsheh and T. C. Wan. Adaptive and fuzzy approaches for nodes affinity management in wireless ad-hoc networks. *Mobile Information Systems*, pages 273–295, 2008.

[65] C. Ng, F. Zhao, U.C. Davis, M. Watari, and P. Thubert. Network mobility route optimization solution space analysis. IETF Internet Draft draft-ietf-nemo-ro-space-analysis-03.txt, 2006.

[66] N. Niebert, A. Schieder, J. Zander, and R. Hancock. *Ambient Networks: Cooperative Mobile Networking for the Wireless World*. John Wiley & Sons, Ltd., 2007.

[67] S. Pack, J. Choi, T. Kwon, and Y. Choi. Fast-handoff support in IEEE 802.11 wireless networks. *IEEE Communication Surveys & Tutorials*, pages 2–12, 2007.

[68] S. Pack and Y. Choi. Fast inter-AP handoff using predictive authentication scheme in a public wireless LAN. *Proc. of IEEE Networks Conference*, pages 15–26, 2002.

[69] S. Pack, H. Jung, T. Kwon, and Y. Choi. SNC: a selective neighbor caching scheme for fast handoff in IEEE 802.11 wireless networks. *ACM Mobile Computing and Communications Review*, pages 39–49, 2005.

[70] C. Perkins and P. Bhagwat. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. *Proc. of the SIGCOMM Conference on Communications Architectures, Protocols and Applications*, pages 234–244, 1994.

[71] C. Perkins, Belding-Royer E., and S. Das. Ad hoc on-demand distance vector (AODV) routing for IP version 6. IETF Internet Draft draft-perkins-manet-aodv6-01.txt, 2000.

[72] C. Perkins, Belding-Royer E., and S. Das. Ad hoc on-demand distance vector (AODV) routing. IETF RFC 3561, 2003.

[73] C. Perkins, J.T. Malinen, R. Wakikawa, A. Nilsson, and A.J. Tuominen. Internet connectivity for mobile ad hoc networks. *Wireless Communications and Mobile Computing*, pages 465–482, 2002.

[74] E. Piri and K. Pentikousis. IEEE 802.21: Media-independent handover services. *Internet Protocol Journal*, pages 7–27, 2009.

[75] E. Piri and K. Pentikousis. Towards a GNU/Linux IEEE 802.21 implementation. *Proc. of IEEE International Conference on Communications (ICC)*, pages 1–5, 2009.

[76] K. Ramachandran, S. Rangarajan, and J. C. Lin. Make-before-break MAC layer handoff in 802.11 wireless networks. *Proc. of IEEE International Conference on Communications (ICC)*, pages 4818–4823, 2006.

[77] I. Ramani and S. Savage. SyncScan: practical fast handoff for 802.11 infrastructure networks. *Proc. of IEEE INFOCOM*, pages 675–684, 2005.

[78] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and S.Y. Wang. HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks. *IEEE/ACM Transactions on Networking*, pages 396–410, 2002.

[79] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser. An analysis of the optimal node density for ad hoc mobile networks. *Proc. of IEEE International Conference on Communications (ICC)*, pages 857–861, 2001.

[80] P. M. Ruiz and A. F. Gomez-Skarmeta. Adaptive gateway discovery mechanisms to enhance Internet connectivity for mobile ad hoc networks. *Ad Hoc & Sensor Wireless Networks*, pages 159–177, 2005.

[81] S. Sharma, N. Zhu, and T-C. Chiueh. Low-latency mobile IP handoff for infrastructure-mode wireless LANs. *IEEE Journal on Selected Areas in Communications (JSAC)*, pages 643–652, 2004.

[82] M. Shin, A. Mishra, and W. Arbaugh. Improving the latency of 802.11 hand-offs using neighbor graphs. *Proc. of the 2nd International Conference on Mobile Systems, Applications, and Services*, pages 70–83, 2004.

[83] S. Shin, A. G. Forte, A. S. Rawat, and H. Schulzrinne. Reducing MAC layer handoff latency in IEEE 802.11 wireless LANs. *Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pages 19–26, 2004.

[84] A. C. Snoeren and H. Balakrishnan. An end-to-end approach to host mobility. *Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pages 155–166, 2000.

[85] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier. Hierarchical mobile IPv6 mobility management (HMIPv6). IETF RFC 4140, 2005.

[86] A. G. Valko. Cellular IP: a new approach to Internet host mobility. *ACM SIGCOMM Computer Communications Review*, pages 50–65, 1999.

[87] H. Velayos and G. Karlsson. Techniques to reduce the IEEE 802.11b handoff time. *Proc. of IEEE International Conference on Communications (ICC)*, pages 3844–3848, 2004.

[88] R. Wakikawa and S. Gundavelli. IPv4 support for proxy mobile IPv6. IETF Internet Draft draft-ietf-netlmm-pmip6-ipv4-support-13.txt, 2009.

[89] R. Wakikawa, J. T. Malinen, C. E. Perkins, A. Nilsson, and A. J. Tuominen. Global connectivity for IPv6 mobile ad hoc networks. IETF Internet Draft draft-wakikawa-manet-globalv6-00.txt, 2001.

[90] WEIRD. WiMAX Extension to Isolated Research Data Network Project, `http://www.ist-weird.eu/`.

[91] K. Weniger and M. Zitterbart. Address autoconfiguration in mobile ad hoc networks: current approaches and future directions. *IEEE Networks*, pages 6–11, 2004.

[92] J-C-S. Wu, C-W. Cheng, N-F. Huang, and G-K. Ma. Intelligent handoff for mobile wireless Internet. *Mobile Networks and Applications*, pages 67–79, 2001.

[93] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia. Fast handovers for proxy mobile IPv6. IETF Internet Draft draft-ietf-mipshop-pfmipv6-09.txt, 2009.

[94] H. Yokota, A. Idoue, T. Hasegawa, and T. Kato. Link layer assisted mobile IP fast handoff method over wireless LAN networks. *Proc. of the ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pages 131–139, 2002.