2.7 IP-Based Mobility

IP-based mobility management techniques can be implemented in several layers of the protocol stack, such as the network layer, transport layer, and application layer. IP-based mobility protocols can be used to take care of mobility for 3G- and 4G-based systems. MIPv4 (Mobile IPv4) (Perkins, 2002b) and its several variants, namely MIP-RO (MIP with Route Optimization), MIP-RR (MIP with Regional Registration) (Perkins, 2002c), MIP-LR (MIP with Location Registers) (Jain et al., 1999), MIPv6 (Johnson et al., 2004), and MOBIKE (S. Eronen, 2006), are a few of the network layer mobility protocols that were defined by the IETF. Cellular IP (Campbell et al., 2000), HAWAII (Handoff Aware Wireless Access Internet Infrastructure) (Ramjee et al., 2000), Proxy MIPv6 (Gundavelli et al., 2008), and IDMP (Intra Domain Mobility Protocol) (Das et al., 2002) are the network layer micromobility protocols suitable for intradomain mobility. Intradomain mobility refers to a movement scenario in which the mobile's movement is confined to one administrative domain. MSOCKS (Maltz and Bhagwat, 1998), TCP-Migrate (Snoeren and Balakrishnan, 2000), and SCTP (Stream Control Transport Protocol) (Koh et al., 2003) have been designed to take care of mobility in the transport layer. SIP-based mobility (Schulzrinne and Wedlund, 2000a) takes care of mobility by means of application layer signaling, such as by SIP (Session Initiation Protocol) (Rosenberg et al., 2002). HIP (Host Identity Protocol) (Moskowitz and Nikander, 2006) defines a shim layer between the network layer and transport layer to provide terminal mobility in a way that is transparent to both the network layer and the transport layer.

We have provided a survey of the related mobility protocols and issues (Dutta et al., 2001, 2002a). We have also experimented with several mobility protocols, namely MIPv4, MIPv6, SIP-based mobility, MIP-LR, and ProxyMIPv6, and verified that these mobility protocols in their current form are not adequate to meet the delay and packet loss performance requirements for real-time traffic (Dutta et al., 2005c, 2007d), and hence these protocols will benefit from overall systems optimization.

We now briefly describe some of these IP-based mobility protocols and categorize them into network layer macromobility, network layer micromobility, application layer mobility, and transport layer mobility.

2.7.1 Network Layer Macromobility

Network layer mobility can be categorized into two types: *macromobility* and *micromobility*. The macromobility mechanism takes care of global mobility where the mobile moves between administrative domains. We describe two types of network layer macromobility, namely Mobile IPv4 and Mobile IPv6.

2.7.1.1 Mobile IPv4

Mobile IP is a mechanism developed for the network layer to support mobility (Perkins, 2002a). Originally it was intended for travelers with laptops to provide portability, and was later adopted by the wireless community. It supports transparency above the IP layer, including the maintenance of active TCP connections and UDP port bindings. A mobile host is identified by a node identifier such as a fixed IP (home IP address). When the mobile host connects to a visited network that is different from the one that its IP address belongs to, its home network forwards packets to the mobile. A router (or an arbitrary node), which is usually known as the home agent, on the user's home network forwards the packets. There are two different methods to deliver packets to a mobile host when it is on a foreign network. With the first method, the mobile host adopts a second (temporary) IP address known as the care-of address (CoA) and registers it with its home agent. When the home agent receives a packet for this user, it encapsulates the packet in another IP packet with the care-of address as the destination address and sends it to the foreign network (Perkins, 1996a,b). Encapsulating a packet within another packet until it reaches the care-of address is known as tunneling. Note that encapsulation adds between 8 and 20 bytes of overhead, which can be significant for voice packets of this size.

The care-of address in the first method is said to be co-located, and it can be acquired via services such as DHCP (Dynamic Host Configuration Protocol) (Droms 1997) or an optimized version such as DHCP with rapid commit (Park et al., 2005) in a local area network, or via PPP (Simpson, 1994) in a point-to-point networking environment. With the second method, the mobile host first registers with a foreign agent (FA) in the network it is visiting. The foreign agent sends (registers) its address to the mobile host's home agent as the care-of address of the mobile host. Packets that are intended for the mobile host are sent to the foreign agent after the home agent has encapsulated them with the IP address of the foreign agent. After decapsulating these packets, the foreign agent delivers them to the mobile host.

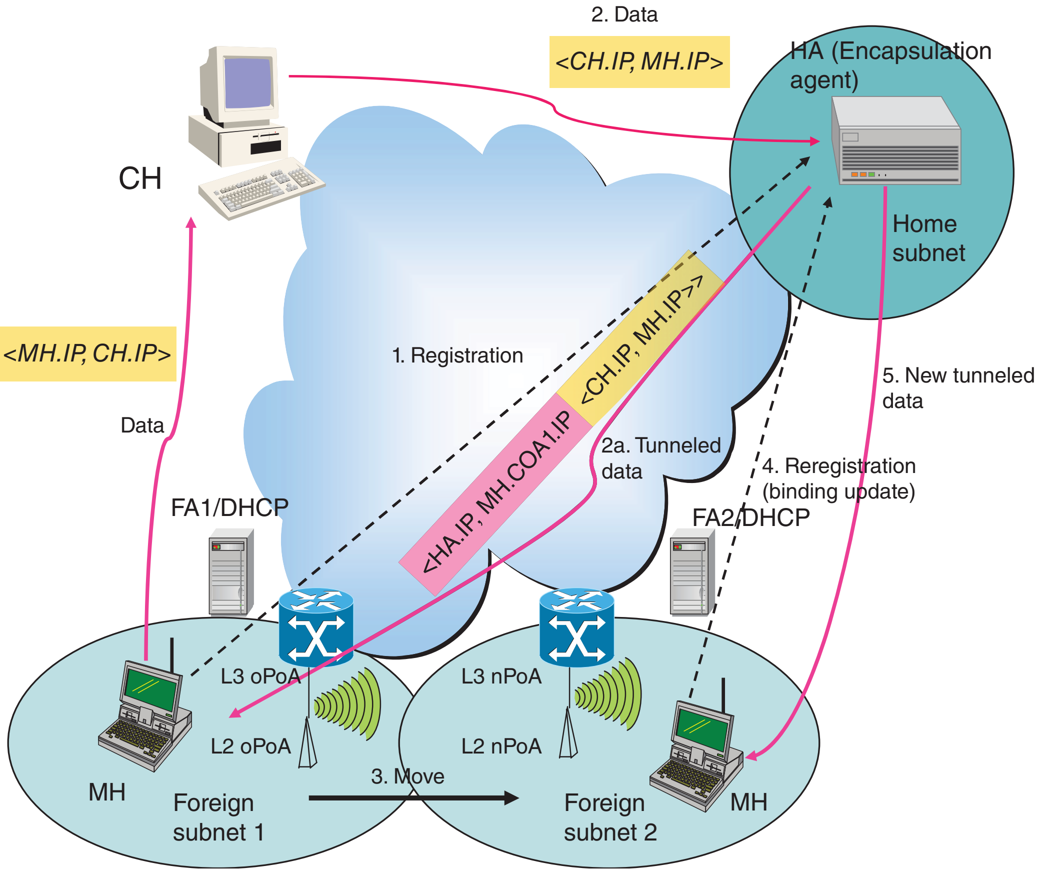

Figure 2.13 Mobile IPv4

Figure 2.13 shows the functional details of Mobile IPv4. In this specific figure, the mobile node moves from subnet 1 to subnet 2 and in the process changes its layer 2 and layer 3 points of attachment and either reconfigures itself with a new care-of-address from a DHCP server or uses the FA's address as its new CoA.

For the methods outlined above to be able to work, a mobile host needs to be able to learn that it has moved from its home network to a foreign network. For this purpose, home agents and foreign agents advertise their presence periodically in their own broadcast domains. A mobile host can also solicit agent advertisements if these advertisements are absent. Packets from the correspondent host must first travel to the home agent and then be later forwarded to the mobile host either by way of the foreign agent or directly. Packets from the mobile host do not have to traverse the home agent; the mobile host sends them as usual with its home IP address as the source address, which is known as triangular routing.

Routing of all incoming packets via the home network may cause additional delays and waste of bandwidth. However, if the correspondent host knows where the mobile host is, it can send packets directly to the care-of address of the mobile host. This is achieved by a route optimization (Perkins, 2002a) process that enables the mobile to send mobility binding updates directly to the corresponding host. Binding updates are sent from the home agent upon request from the mobile, or can be sent upon receiving a warning message from a foreign agent if the mobile host changes location during a communication session. In the second case, the former foreign agent will keep forwarding packets to the new foreign agent until the correspondent host updates its mobility binding cache (this is known as smooth handoff).

Another optimization that has been proposed is regional registration (Calhoun et al., 2003b), where the mobile registers locally in a visited domain. In base Mobile IP, a mobile host is required to register with its home agent each time it changes care-of address, thus causing signaling delay to the registration if the mobile host is far away from its home agent. Regional registration attempts to decrease the number of home registrations by maintaining a hierarchical structure of foreign agents. As long as a mobile host's foreign agent is located hierarchically under a so-called gateway foreign agent (GFA), it is unnecessary to relay registration messages back to the home agent, since the home agent has already registered the GFA's address as the care-of address. To make Mobile IP handoffs (i.e., the registration process) more suitable for real-time and delay-sensitive applications, Malki (2004) proposed two additional methods. With the first of these methods, called the network-assisted mobile and network-controlled (NAMONC) handoff method, the mobile host is informed (assisted) by the network that a layer 2 handoff is anticipated. Here, it is proposed to use simultaneous bindings (multiple registrations at a time) in order to send multiple copies of the traffic to potential points of attachment before the actual movement. The other method, called the network-initiated, mobile-terminated (NIMOT) handoff method, proposes extensions to the base Mobile IP so that foreign agents can utilize information from layer 2. Specifically, foreign agents use layer 2 triggers to initiate a preregistration prior to receiving a formal registration request from the mobile host. Both methods assume considerable involvement of information from layer 2.

We present many of the Mobile IP-related optimization techniques in Chapter 6.

2.7.1.2 MIP-LR

Mobile IP with Location Registers (MIP-LR) avoids encapsulation of packets (Jain et al., 1999) and provides survivable features in the case of failure of location registers. In MIP-LR, each subnet may contain a host that functions as a visitor location register and/or a host that functions as a home location register. Each mobile host can be served by multiple HLRs. Each VLR advertises its presence on its local subnet using agent advertisement messages similarly to Mobile IP. When a mobile host is located on its local subnet, it is not registered at either the HLR or the VLR. When the mobile moves to a foreign network, it obtains a care-of address from the pool of addresses that the VLR has. The mobile host registers with the foreign VLR using the CoA it has obtained, which in turn relays the registration to the mobile host's HLR. The HLR returns a registration reply containing the allowed lifetime for this registration; the VLR records the mobile host's CoA and the lifetime and forwards the reply to the mobile host. A correspondent host wishing to send a packet to the mobile host for the first time issues a query to the HLR, which returns the mobile host's CoA as well as the remaining registration lifetime. The correspondent host then sends the packet directly to the mobile host's CoA. The correspondent host caches a binding for the mobile host's CoA and uses this binding for subsequent packets destined for the mobile host. The correspondent host must refresh its binding cache by querying the HLR again before the mobile host's remaining registration lifetime expires. In MIP-LR, unlike Mobile IP, the HLR can be geographically distributed anywhere. We have implemented an extension of MIP-LR using an application layer module that does not require any kernel changes. Having an application layer module, this allows a mobile to use a policy-based approach to trigger MIP-LR for certain types of application. Figure 2.14 shows the functionalities of Mobile IP with Location Registers when the mobile moves from one subnet to another and in the process changes its layer 2 and layer 3 points of attachment. In this case, there is no foreign agent in the visited network and it is also not a requirement that the location register needs to be in the home network.

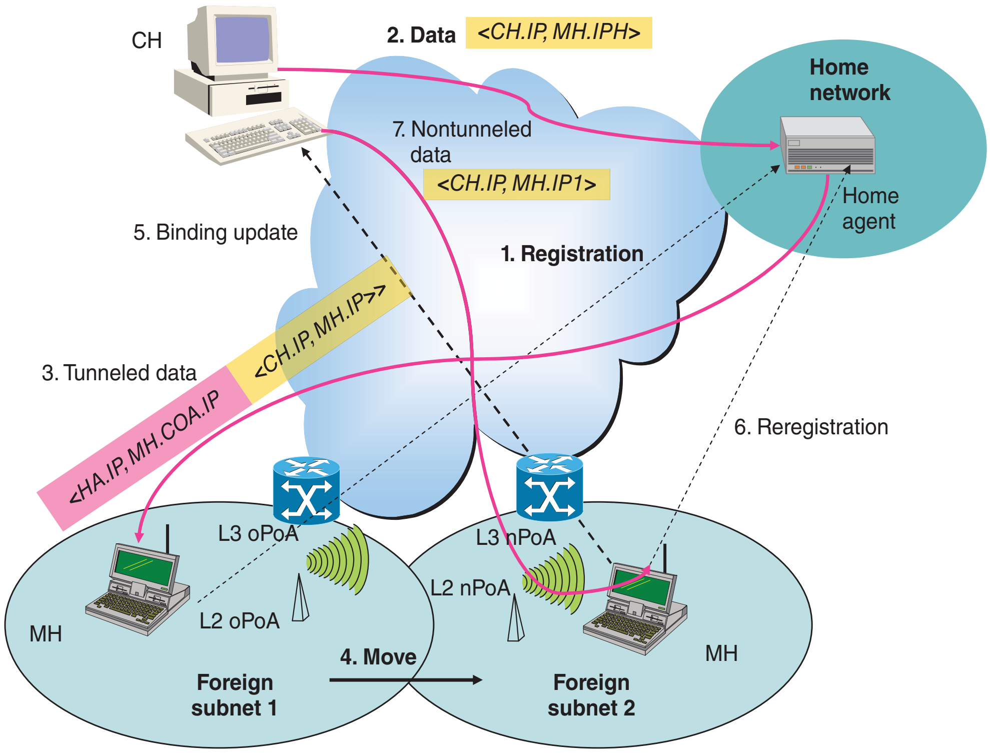

Figure 2.14 Mobile IP with Location Registers

After a mobile host moves, if the mobile host was previously registered at some other foreign VLR, the new VLR deregisters the mobile host at the old VLR. This deregistration is required so that the mobile host's old CoA can eventually be released for use by some other mobile host. If a VLR runs out of CoAs temporarily, it can still issue its own IP address as a CoA and, when a mobile host registers using this CoA, inform the HLR accordingly.

2.7.1.3 Mobile IPv6

IPv6's (Deering and Hinden, 1998) increased address space and inherent support for security and autoconfiguration have made it an attractive candidate to support mobility for the next-generation Internet. Mobile IPv6 is the protocol to support mobility for IPv6 nodes. Since address autoconfiguration is a standard part of MIPv6, the MH will always obtain a CoA routable to a foreign network. Thus, there is no need to have a foreign agent in MIPv6. When the mobile node moves to a new foreign network, it acquires a temporary care-of-address using stateless autoconfiguration (Thomson and Narten, 1998) or via DHCPv6 (Droms et al., 2003).

Figure 2.15 shows the functional components of Mobile IPv6. Unlike Mobile IPv4, the visited networks do not have any foreign agents. MIPv6's route optimization feature also enables direct data delivery from the correspondent host (CH) to the mobile node.

Although Mobile IPv6 is defined as a network layer approach and one needs to install an MIPv6 stack so as to support mobility in an IPv6 space, any standard operating system will in future come with inherent Mobile IPv6 support.

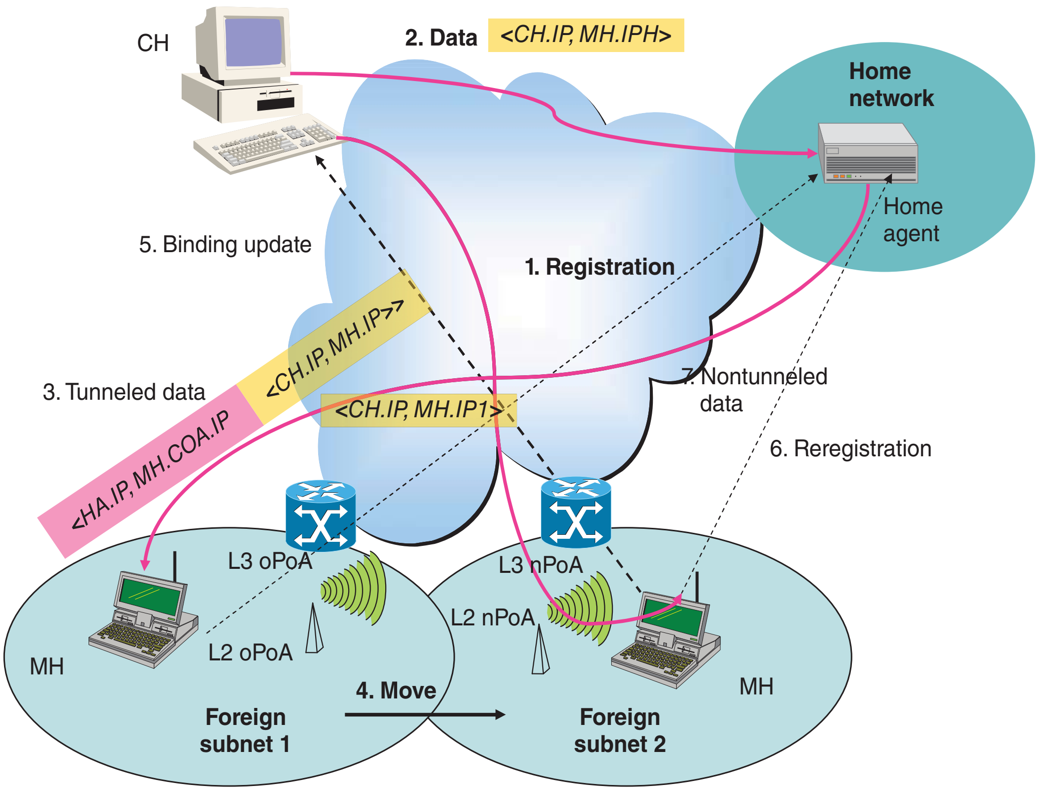

Figure 2.15 Mobile IPv6

While Mobile IPv6 provides a way of making sure of the uniqueness of an address as a mobile moves to a new router space, it also adds delay to the binding update and binding acknowledgement as in Mobile IPv4. However, compared with regular Mobile IP, there are inherent advantages to MIPv6. Route optimization is a standard feature of MIPv6, and thus there is no need for the CH to be equipped with additional software like MIP-RO. The MH sends a binding update directly to the CH and makes use of the home address destination option as part of the binding update. This allows the correspondent host to keep a binding cache that maps the care-of address of the mobile to the mobile's home address. For the ongoing traffic, this avoids triangular routing, and thus packets from the CH to the MH need not be encapsulated but are sent directly to the MH with its CoA as the source route. However, when a new CH needs to communicate with the mobile for the first time, the packets from the CH need to travel to the home agent and be tunneled to the mobile host. As the mobile moves during the packet transfer process, the subsequent packets are tunneled directly to the mobile host without being routed via the home agent.

2.7.1.4 Fast Mobile IPv6

While Mobile IPv6 takes care of session continuity during handoff, by itself it lacks the ability to provide the low-latency handoff and reduced packet loss that are essential for many interactive applications such as Voice over IP, gaming, and conferencing. Most of the handoff delays observed in Mobile IPv6 are due to IP address configuration and binding update delay when the home agent is far away. Fast Mobile IPv6 (FMIPv6) (Koodli 2008) proposes mechanisms to reduce the handoff delay by way of localizing the binding updates to the edges of the network, reducing the delay due to IP address acquisition, and buffering at the edge routers. This involves additional protocol exchange between the mobile host, the current router (pAR), and the next access router (nAR). These mechanisms can be categorized into two types of handover, namely predictive and reactive. The FMIPv6 protocols work in conjunction with the existing MIPv6 stack. Figure 2.16 shows the interaction among several network elements. For brevity, it does not reflect the MIPv6-related signaling, however. Figure 2.17 and Figure 2.18 describe the call flows for predictive and reactive handovers, respectively.

We now briefly describe the predictive operation of FMIPv6. The mobile host sends a router solicitation for a proxy (RtSolPr) message to its default access router (pAR) in order to obtain information related to the link layer addresses of the neighboring access points discovered during the layer 2 scanning process, and the prefixes associated with the neighboring access router (nAR). The current access router (pAR) communicates with the nAR using protocols such as Candidate Access Router Discovery (CARD) (Liebsch et al., 2005) to obtain the relevant information about the neighboring network elements. The pAR serving the user responds with a proxy router advertisement (PrRtAdv) containing the requested information, thus allowing the mobile host to perform address autoconfiguration prior to its movement to the new network. The host, after formulating a prospective new CoA, sends a fast binding update (FBU) to its default router instructing it to tunnel packets addressed to its old CoA (oCoA) towards its new CoA (nCoA). The access router currently serving the host (pAR) starts buffering newly arriving packets with the oCoA as their destination and exchanges handover initiate (HI) and handover acknowledge (HAck) messages with the nAR to initiate the process of the MH's handover. This HI/HAck message exchange can also serve for validation of the nCoA already formed by the host. The pAR responds to the MH with a fast binding acknowledge (FBack) message on both links (old and new) and starts the tunneling of buffered and arriving data to the nCoA. These

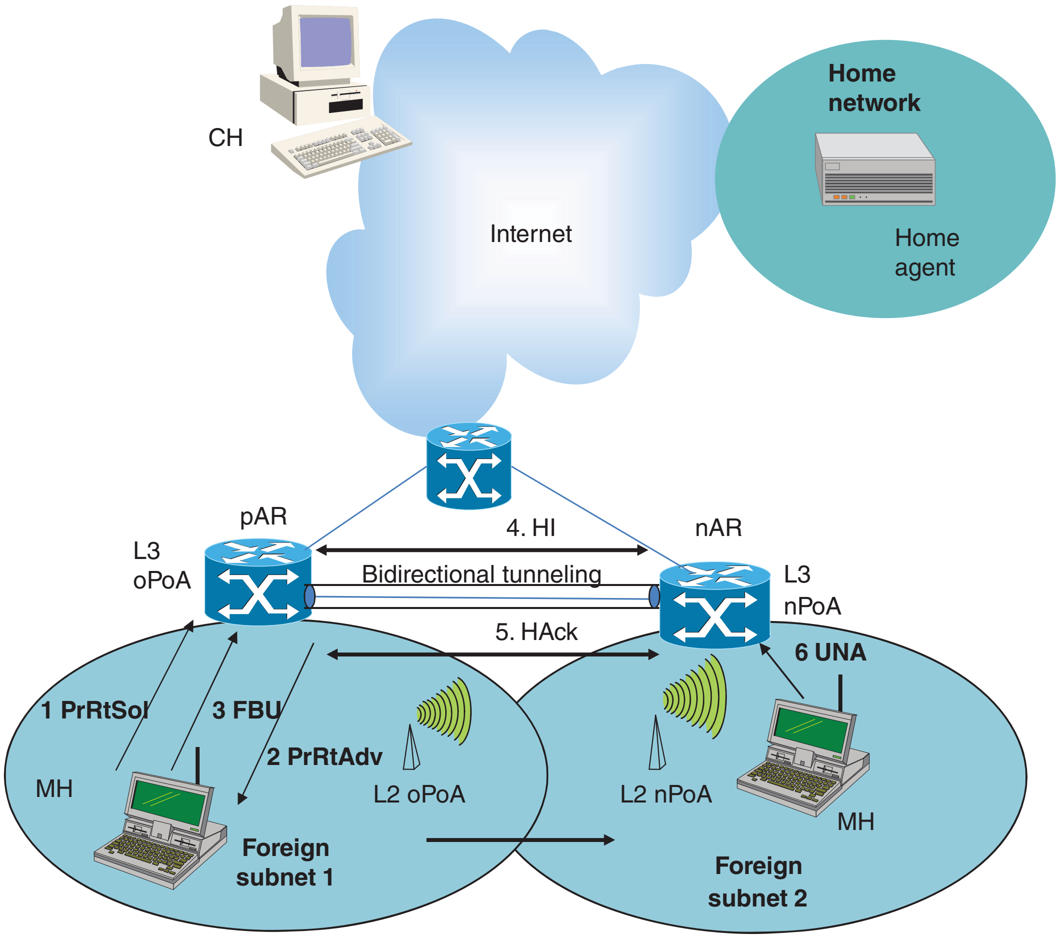

Figure 2.16 Fast Mobile IPv6

packets are also buffered at the nAR until the mobile arrives at the new point of attachment. The MH, as soon as it attaches to the new link, transmits a UNA (unsolicited network advertisement) to inform the nAR of its presence. Buffered packets at the nAR can be delivered immediately to the MH on the new link.

In the reactive mode of FMIPv6, the FBU is sent after the mobile connects to the new network. Thus, the FBU is routed through the nAR but is processed at the pAR. Unlike predictive handoff, the packets destined for the previous address of the mobile are forwarded to the nAR instead of being buffered at the nAR. Packet loss is minimized in predictive handover owing to buffering at the nAR.

2.7.2 Network Layer Micromobility

There are several network layer micromobility protocols that are meant to optimize mobility when the mobile's movement is confined within a domain. These protocols avoid the overhead associated with tunneling over the air, and reduce the signaling overhead when the mobile's movement