A Survey on Network Architectures for Mobility

XiuJia Jin

Abstract

Currently the most well-known mobility solution for the Internet is Mobile IPv4 and Mobile IPv6. However, new challenges are brought up by all kinds of needs in mobility such that they can hardly be handled with Mobile IP alone and standard networking architecture we have now. Thus a number of new architectures are proposed to address the issues in mobility. In this paper we discuss the requirements and current solutions of mobility, and review several novel architectures proposed for IP-based wireless networks in recent years.

Keywords: Network Architectures, Mobile IP, Mobile IPv4, Mobile IPv6, Macro-Mobility, Micro-Mobility, Host Mobility, Network Mobility, IP Mobile Multicast

Table of Contents

1 Introduction

As is widely agreed, next generation mobile communication system, fourth-generation (4G) system, will be heterogeneous networks, which provides ubiquitous access and seamless mobility across heterogeneous network technologies, such as cellular, WLAN, and broadcast access, and this relies on the all-IP architecture.

However, before adopt IP techniques in 4G wireless networks, mobility management is still one of the most important issues remain to be solved [Lach03].

Mobility in wireless networks basically refers to a node, Mobile Node (MN), or sometimes a subnet, changing its point of attachment to the network while its communication to the network remains uninterrupted. A change in the MN's point of attachment to the network is called handover. The mobility of a node is called Host Mobility, and the mobility of a subnet is called Network Mobility. In addition, there are also Personal Mobility, which refers to the ability of a user to access services regardless of the terminal or networking he/she is using, and focuses on the movement of users rather than devices; and Session Mobility, which refers to the mobility between two terminals, and it is mainly about tracking the communication sessions between two nodes as they move.

According to the scope of node movement, mobility can be divided into Micro-mobility and Macro-mobility. On the link-layer, most access networks provide mobility by having an access router keep track of the specific access point a Mobile Node (MN) is attached to [Chiussi02]. The localized mobility between pico-cells (probably heterogeneous cells) in the same subnet and the mobility between subnets in a domain is called micro-mobility, while the mobility between the domains in wide-area wireless networks is called macro-mobility. The mobility solutions like Mobile IP can provide support for macro-mobility. But they are not suitable for micro-mobility due to signaling overhead, handover latency, and transient packet loss.

To achieve mobility, both seamless connectivity and continuous reachability should be provided [Zhuang03]. Since the standard Internet combines the unique host identifier with the topology location using IP addresses, it cannot provide support for mobility. Thus, Mobile IP is proposed by IETF to support mobility in IP networks. Mobile IP supports mobility by decoupling the binding between the host identifier and topology location using a fixed indirection point. Although Mobile IP successfully extended the ability of the Internet to support mobility, it still has some limitations and is not suitable for all movement patterns of MNs.

In this paper, we will review several novel architectures proposed for macro-mobility in IP-based wireless networks. We will focus mainly on Host Mobility in this paper. [Perera04] is a survey paper on Network Mobility. The rest of paper is organized as follows: Section 2 presents the fundamentals, requirement and issues of mobility, and gives an overview of current solutions for mobility. In Section 3, several novel mobility architectures (mainly on host mobility and macro-mobility) are presented and evaluated. Section 4 is the summary of this paper.

Back to Table of Contents

---

2 Mobility of IP-based Wireless Networks

Mobility management in wireless networks involves changing the point of attachment, and hence the IP address, of a MN. This section focuses on the fundamentals, requirements and issues of mobility, and introduces current solutions for mobility.

2.1 Categories of Mobility

Although all types of mobility require efficient handoff, efficient routing, low packet loss, etc, different issues exist in different patterns of movement. In order to better understand and address the issues exist in mobility, we should first take a look at different scenarios of mobility.

2.1.1 Host Mobility and Network Mobility

Host mobility refers to an end host changing its point of attachment to the networks while the communication between the host and its correspondent node stays uninterrupted.

Network mobility refers to a mobile IP subnet changing its point of attachment to an IP backbone. [Lach03] In a simple scenario of network mobility, a mobile network contains a mobile router and a set of mobile nodes and the internal structure of a mobile network is a relatively stable internal topology. While in a complex mobility scenario, a mobile network may itself be visited by mobile nodes or other mobile networks.

2.1.2 Macro-mobility and Micro-mobility

Macro-mobility refers to the inter-domain movement. A number of well-known proposals like Mobile IP are developed to address the issues in macro-mobility. These proposals are well suited for macro-mobility due to their mechanisms for achieving efficient handoff, low rate of packet loss, efficient routing of packets, etc. However, these proposals always have relatively large overhead.

Issues will rise when mobility solutions for macro-mobility such as Mobile IP are adopted for micro-mobility. Base Mobile IP mechanism introduces significant network overhead in terms of delay, packet loss, and signaling. For example, the real-time wireless applications such as Voice IP (VoIP) would suffer degradation of service due to frequent handoff [Inayat03]. Micro-mobility solutions are proposed for localized mobility in a domain. These proposals focus on reducing the handoff latency by inducing those additional overheads due to control traffic as they have to maintain routing information at the local network and are also heavy on the address space [Campbell02].

2.1.3 IP Mobile Multicasting

Instead of sending data to a single node, multicasting delivers data to a set of selected receivers. In IP multicast, a source sends a single copy of a packet and the network duplicates the packet as needed until the packet reaches all the selected receivers. This avoids the overheads associated with both replication of packets at the source and sending duplicated packets over the same link. [Romdhani04] is a good survey paper for IP mobile multicast.

2.2 Requirements of Mobility

The main goal of the mobility solutions is to continue the communication of the MN and the networks while it moves, and avoid the disrupting of the connections. When a MN moves from one place to another, in order to support seamless connectivity and continuous reachability, a mobility solution should provide mechanism to handle the handover and the routing of packets thereafter. The proposals for mobility should have the following properties [Atiquzzaman05][Zhuang03][Henderson03]:

Efficient Handoff: The performance of a mobility scheme mainly depends on the type of handoffs it uses. There are two types of handoffs: soft handoff and hard handoff. Soft handoff makes a new connection before disconnecting the previous connection. It allows the mobile node to communicate with multiple interfaces during handoff, and the communication with the old interface is dropped when the signal strength between the old access point drops below a certain threshold. Hard handoff drops the previous connection before making a new connection. Handoffs should be handled efficiently in order to reduce or avoid the loss and delay of packets as possible.

Location Management: If a mobile host offers services to other nodes, it must be able to be located by these nodes as it moves as well as keeping the privacy of its topological location.

Efficient routing: Packets should be routed with the latency as low as possible, optimally close to the shortest path provided by IP routing.

Security: Security is a crucial issue in a wireless environment. Mobility management schemes should not introduce additional security issues to the network. Also, the interruption of connectivity due to the time required for authentication process should be avoided.

Scalability: A mobility scheme is said to be scalable if its performance does not drop as the number of nodes (MNs and CNs) increases.

Fault tolerance: A scheme should be able to function even in the presence of failure. A mobility scheme should make the communication between mobile nodes as much tolerant to fault as the communication between stationary nodes.

Simultaneous mobility: end hosts may move simultaneously, and the communication between them should not be interrupted.

Link layer independence: User should be able to seamlessly operate across heterogeneous link layer technologies, not all of which support the same link layer mobility scheme.

Compatibility with IP routing: Mobility management must work well with IP routing, such as acquiring a new topologically correct IP address upon moving, since full host routes are not propagated in the Internet.

Transparency: The mobility scheme should be transparent to applications so that the applications are not aware of the handoff, and thus do not need to be modified for mobility.

Quality of Service: QoS should not be reduced as the MH moves and performs handoff.

2.3 Current Solutions for Host Mobility

The Internet host mobility is mainly approached from three angles: data link layer mobility, network layer mobility, and other higher layer mobility [Snoeren00]. IEEE 802.11b, Mobile IP, and MSOCKS are three example schemes of these three layers respectively [Atiquzzaman05].

Link layer technologies which support mobility include Ricochet [Ritter01], 802.11b, GSM, etc. Hiding mobility in the link layer results in the reinvention of mobility support in each new wireless system.

Most solutions proposed so far (e.g. Mobile IP) are based on the idea of indirection points between MN and CN so that the CN does not need to know the topology location of the MN by sending packets to the indirection points. These approaches do not require changing fixed hosts in the Internet, but they require changing the underlying IP substrate. Some other solutions emphasize on end-to-end architecture (e.g. TCP Migrate). These solutions do not require change to the underlying IP substrate [Snoeren00].

Here we will introduce some mobility solutions which has already existed for a period of time and is widely used or referred to.

2.3.1 Mobile IP

The most widely known mobility solution today is Mobile IP, which were developed by IETF to support mobility on the Internet. Mobile IP aims to allow a MN to continue the communication with its CN during its movement. It supports network layer mobility so that TCP is not aware of the mobility.

In Mobile IP, Host Agent (HA) is used as indirection points. HA is in the MN's home network and it intercepts and tunnels packets to the MN. A Mobile Node (MN) has a permanent Home Address (HoA) from its home network and obtains a temporary Care-of-Address (COA) which is routable within the foreign network when it moves to a new network. The MN registers its COA to the HA in its home network every time it obtains a new COA, and this process is called registration process. For maintaining the transport and higher-level communications when moving, the MN maintains its HoA and uses the COA for routing purpose. A binding associates these two addresses on both the MN part and the HA part.

Mobile IP ensures the delivery of packets which destine to a MN's home address by creating a routing tunnel between the MN's home network and its COA. Each time when a packet is received in the home network for the MN, the HA will intercept the packet and then encapsulate it inside a packet and send it to the COA of the MN. Thereafter packets sent from the MN addressed to the CN may either be routed directly from the foreign network to the CN, which is also known as triangle routing, or be tunneled back to the HA and routed from HA to CN, which is known as reverse tunneling. Triangle routing may not be allowed by the security infrastructure in the foreign network, while reverse tunneling solves this issue.

Both Mobile IPv4 and Mobile IPv6 are based on the above ideas, and they share many features. However, Mobile IPv6 offers some improvements:

1) There is no need for an FA in Mobile IPv6. In IPv4, a Foreign Agent (FA) is deployed. FA is an agent in the foreign network which the MN is visiting, and when a MN visits the foreign network, it obtains a COA from the FA of that network. In Mobile IPv6, since it offers address auto-configuration capabilities, there is no need to deploy FAs in foreign networks. Most packets sent to the MN are sent using an IPv6 routing header rather than IP encapsulation, which reduces the amount of overhead compared to Mobile IPv4. Also, the registration on Mobile IPv6 is direct while in IPv4 it may either be direct or through the FA.

2) Optimal routing. Another important feature in Mobile IPv6 is that it offers support for optimal routing of data packets between the CN and the MN, bypassing the HA, in order to avoid triangle routing. With route optimization, the MN informs the CN of its COA using a Binding Update (BU), and then an IPv6 routing header will be used to send packets directly from the CN to the COA of the MN. However, route optimization compromises location privacy by exposing the COA, and hence its location, to the CN.

3) Dynamic HA discovery. Mobile IPv4 uses a broadcasting mechanism to dynamically discover the HA in the home network, while Mobile IPv6 uses the IPv6 Neighbor Discovery Protocol [Narten98].

Mobile IP also has the following limitations:

1) The dependence in Mobile IP on a fixed HA reduces fault tolerance. If the HA or the home network fails or is overloaded, the MN will be unreachable. To address this issue, the notion of dynamic home agents is proposed for MIPv4. However, the actual algorithm used to discover and allocate a nearby home agent is still under investigation [Zhuang03]. MIPv6 provides a dynamic home agent address discovery mechanism that allows a MN to dynamically discover the IP address of a HA in its home network.

2) The routing efficiency would be degraded by routing through HA when the MN is far away from HA.

3) Handoff performance. There have been two mechanisms proposed to increase handoff performance in MIPv4 and MIPv6: low latency handoff and fast handover. In low latency handoff, a BU is sent in advance of an actual link-layer handoff. However, it must be guaranteed that the BU completes before the actual handoff does, which is difficult to achieve in practice. Fast handover sets up a bi-directional tunnel between an anchor FA and the current FA. This allows the MN to delay a formal BU to the HA and minimizes the impact on

real-time applications. However, this mechanism requires the existence of a FA in each network the MN visits.

2.3.2 Other Proposals

There are also some other well known proposals for mobility management in IP-based wireless networks, such as MSOCKS for transport layer mobility and HAWAII, Cellular IP for micro-mobility.

Transport layer schemes are based on an end-to-end approach to mobility that attempt to keep the Internet infrastructure unchanged by allowing the end hosts to take care of mobility. MSOCKS, SIGMA, and TCP Migrate are transport layer schemes. A good survey paper on transport layer mobility management schemes can be found in [Atiquzzaman05].

MSOCKS [Maltz98]: MSOCKS is a transport-layer mobility architecture which also uses home-agent-based approach. In MSOCKS, connection redirection is achieved by using a split-connection proxy. MSOCKS divides a TCP connection at a proxy and thus the host-to-host communication is divided into host-proxy and proxy-host communications, which is called TCP Splice. MSOCKS uses TCP Splice to migrate the connection from the old address to the new one. It supports multiple IP addresses for multiple interfaces. When a MN moves to a new location, it obtains a new IP address and establishes a new connection using the new interface with the proxy. Since the connection between its CN and the proxy remains unchanged, the connection will not be interrupted and the CN will not be aware of the mobility.

TCP Migrate [Snoeren00]: TCP migrate decouple the binding of host identifier and topology location by redirecting through the DNS. In TCP Migrate, both the MN and CN use a modified form of TCP which can tolerate a change in IP address for communication. The CN uses DNS to learn the current address of the MN, which updates DNS every time it moves. Since it does not use an indirection point, TCP Migrate can achieve an optimal latency stretch and is as fault tolerant as IP routing. It has the following limitations: 1) it lacks simultaneous mobility support, and 2) it needs modification of the TCP implementations on both the MN and the CN, 3) it does not preserve location privacy.

There are also a number of micro-mobility schemes, such as HAWAII, Cellular IP, IDMP, etc. These schemes provide intra-domain mobility by using source based routing or dynamic home agent.

2.4 Current Solutions for Network Mobility

Although Mobile IP can suitably handle node mobility, including hosts and routers, it is not explicitly suitable for network mobility. [Lach03] The basic approach of supporting network mobility is bidirectional tunneling. In Mobile IPv6, a MN is supported by its HA using a bidirectional MN-HA tunneling. Similarly, while apply this to network mobility, a Mobile Router (MR) will have an HA, and a bidirectional MR-HA tunneling will be used to support the session continuity of the nodes within its mobile network while the MR and its mobile network moves.

However, the basic approach is not sufficient to address advanced issues [Lach03], such as: route optimization for nested mobility, which refers to allowing packets sent between a CN and a MN within a mobile network to be routed through optimal path no matter how deep the mobile network is nested; optimization for a standalone mobile network, which means that a visiting node in a mobile network should be reachable by the other nodes in the same mobile network even when the mobile network is not attached to the IP infrastructure; seamless mobility, which means the packet loss and handover delay should be minimized; support of dynamic addressing and routing mechanisms, which means that the MR should be able to take part in routing and network management operations with its home network; and etc.

A concept which is called Prefix Scope Binding Update (PSBU) [Thubert02] is proposed to support advanced network mobility in IPv6. It suggests that an MR advertise its mobility through both the classic Mobile IPv6 binding update and PSBU. A PSBU binds the mobile network prefix with the COA of the MR rather than HoA with COA of the MN. With this approach, upon reception of a packet from MR's HA, the MR will send a PSBU to the CN which enables the CN to bind the mobile network prefix to the MR's COA. Thus, all the subsequent packets addressed to the nodes in the mobile network will be sent toward the MR's COA to achieve route optimization.

Another project named OverDRiVE is also in developing. This project has strong influence on the ongoing work within the IETF NEMO group. Detailed information about OverDRiVE can be found in [Ronai03].

Research on network mobility is just starting, and much research still needs to be conducted to develop optimized solutions to enable seamless network mobility.

2.5 Current Solution for IP Mobile Multicast

In multicast communication, the scenario of handover is particularly challenging and several issues emerge with most solutions due to the handover impacts. When a multicast receiver is mobile, it will experience additional delay in receiving packets due to handover delay, joint latency, and increased propagation delay to the new location [Romdhani04]. When a multicast receiver move from one location to another, it may either use the basic way in Mobile IP to receive the multicast packets from its home network, or it may use the multicast infrastructure of the foreign network to subscribe to multicast groups. However, both ways can induce problems: if it receives packets from home network after it moves, the multicast routing may be sub-optimal and the delay may increase; while if it re-subscribe every time it moves to a new network, some multicast packets may be missed. Also, the mobility of the multicast source can cause disruption of the multicast session or a reconstruction of the entire multicasting tree.

There are several types of solutions for mobile receivers [Romdhani04]:

1) Home subscription-based solutions, which rely on the Mobile IP architectural entities, HA and MN, and uses a multicast router located in the home network. The MN will then follow the common way in Mobile IP by setting up a bi-directional tunnel to receive multicast packets from the HA.

2) Remote subscription-based solutions, where the MN joins the multicast group via the foreign network.

3) Hybrid solutions, which use both home and remote subscription-based solutions simultaneously.

4) Non-IP multicast-based solutions, which rely on the techniques such as explicit multicast and recursive unicast rather than use IP Multicast.

There are also several solutions for mobile sources [Romdhani04], which may either use home-based approaches in which a mobile source uses its HoA, or use remote-based approaches the mobile source sends the multicast packets from the foreign network.

Multicasting-based architecture can also be used for mobile host in Host Mobility. Some proposals which use IP multicasting for mobility support hide the indirection point within the multicast address. Several studies [Mysore98] [Helmy00] [Helmy01] have shown that using multicast mobility can cut the latency stretch of Mobile IP in half and significantly reduce packet loss due to handoffs.

Mobility Support using Multicasting in IP (MSM-IP) is an architecture implements mobility using IP multicast. Its main advantages are that it can have low handoff latency and packet loss. However, MSM-IP is

a single point of failure and is vulnerable to overloading, network faults and host faults.

As we can see from this section, the issues in mobility management in wireless networks are still not well solved, which motivates continuous research effort on new architectures for mobility.

Back to Table of Contents

---

3 Novel Architectures for Mobility

In this section, we will present several novel architectures proposed in recent years. However, it is not possible to present all the architectures for mobility. The architectures we choose to present below mainly focus on the supporting host mobility in IP-based wireless networks. There are also some other proposals address other mobility issues: for example, an architectures for micro-mobility can be found in [Chiussi02]; [Carter03] is an interesting localized mobility system which provides support to the local communication of MN in the foreign network even when it is disconnected from the Internet; [Lenders05] proposes an architecture for mobile ad-hoc networks.

3.1 ROAM

Robust Overlay Architecture for Mobility (ROAM) [Zhuang03] is an architecture proposed to provide seamless mobility for Internet hosts. It builds on top of the Internet Indirection Infrastructure (i3) [Stoica02].

i3, which uses an overlay indirection infrastructure that gives end-hosts control over the placement of the indirection points, provides a powerful and flexible rendezvous-based communication abstraction. i3 is implemented as an overlay network on top of IP. It achieves the mobility by deploying a trigger to associates a logical identifier and the address of a receiver, in other words, the trigger plays the role of indirection point. In i3, instead of sending a packet to a certain destination address, each packet is sent to an identifier. A receiver inserts a trigger into an i3 node (server) to receive the packet. Fig.1 shows an example of how a packet is sent from a sender to a receiver. At first the receiver R inserts a trigger (id, R), which binds the id with its address, into an i3 sever. When a packet (id, data) is sent from the sender, it would be routed through the overlay network to the i3 server. The i3 server then looks up the matching trigger and forwards the packet via IP to the receiver with the address specified in the trigger. In i3, delivery is best-effort with no guarantees about packet delivery [Zhuang03].

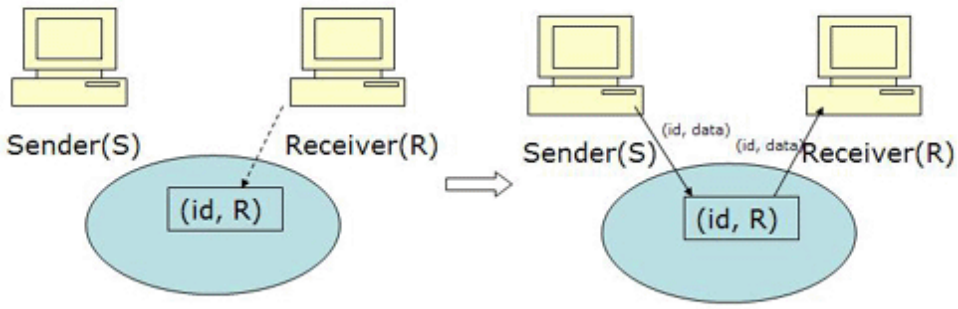

Fig. 1: An example of communication between two nodes

i3 provides a general-purpose indirection infrastructure that supports different communication services such as multicast and transcoding [Zhuang03].

ROAM is build on top of i3 and provides an end-to-end architecture for Internet host mobility. It gives end hosts control over the placement of indirection points, which allows the end hosts to optimize the routing and handoff efficiency. In ROAM, after a mobile host changes its address from R to R', it simply updates the trigger from (id, R) to (id, R'). As we can see from Fig. 2, the change of the receiver's address is transparent to the sender.

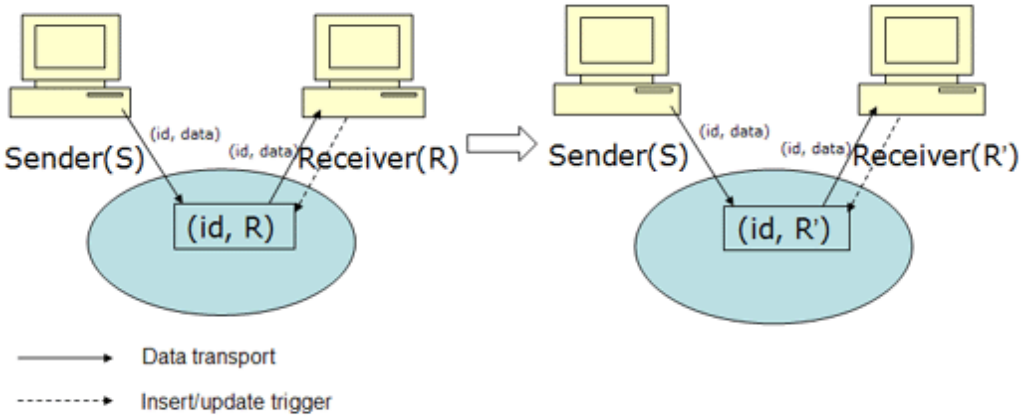

Fig. 2: An example of the receiver changing its address from R to R'

ROAM achieves the following properties desired in mobility:

Efficient routing: ROAM improves the ability of i3 to provide low latency stretch by taking advantage of the mobility pattern. The delay on each hop in i3 is sometimes unacceptably high. To reduce the delay, i3 introduces trigger server caching and trigger sampling techniques [Stoica02]. With the first technique, both end hosts caches the server storing the particular trigger so that the packets which match the trigger can be sent directly to that server via IP. Trigger sampling technique is that an end-host picks random sample triggers from nearby servers, measures the round trip delay to the servers storing those triggers, and uses the triggers with lowest delay. ROAM takes into consideration the mobility pattern where some moves are short and some moves can be very far (both in geographic distance and network latency). As the mobile host moves, it randomly samples nearby i3 servers, and if the cache is full and the new sample is much closer than the closest sample exists in the cache, then the least recently used trigger will be removed from the cache. If the new sample is not much closer than the closest sample existing in the cache, then the closest sample in the cache will be replaced by the new sample.

Efficient handoff: One of the issues in Mobile IP is that the address Binding Updates (BUs) are propagated all the way to a HA or CN. As a result, a potentially large number of packets may be in flight when the path latency from the MH to the HA or CN is high. Those in-flight packets could be lost if the MH performs hard handoff. This situation is improved in ROAM. Because of the number of packets that are lost during a hard handoff is proportional to the delay between the MH and the indirection point [Zhuang03], and that ROAM lets the MH to choose the indirection points, i.e. triggers, that map onto nearby i3 servers, the packet loss is reduced in ROAM.

Fault tolerance: ROAM is tolerant to server failure by periodically refreshing triggers. The triggers stored at a failing server can be inserted to another server when the triggers are refreshed.

Simultaneous mobility: Both end hosts can move simultaneous since the communication can always be kept through the i3 servers.

Location privacy: The flexibility of the end hosts to choose the triggers guarantees the privacy of host

location, and it also allows each application to make the tradeoff between the privacy and efficiency of routing.

ROAM has a low latency stretch and it is highly robust, and it achieves most mobility requirements. However, ROAM has a large header size (117 Bytes) [Zhuang03], and header compression is needed to reduce the packet header overhead.

3.2 Hi3

Hi3 [Nikander04] is a combination of Secure-i3 and Host Identity Protocol (HIP). It is more efficient and secure than Secure-i3, and more flexible and denial-of-service resistant than HIP.

Secure-i3 is derived from i3 and provides more robust protection against Denial-of-Service (DoS) attacks by hiding the IP addresses of the end hosts from other users of the network. The indirection approach in Secure-i3 adopts two types of triggers: public and private triggers, instead of a single type of trigger in i3. Public trigger are used to announce the existence of a service and is known by other users, while private triggers are used for actual communication between two end-hosts and are only known by the end-hosts involved in the communication.

HIP [Moskowitz05] provides an end-to-end mobility by introducing another way of breaking the binding between identifiers and topological locations of mobile nodes by mapping IP address to new cryptographic identifiers. [Nikander04] HIP adds a new layer to the IP stack, and the new layer is located within the IP layer. In HIP, the cryptographic identifier, together with the public/private key, are used to replace the IP address and to provide identities for the hosts. To address the issues of initial rendezvous, simultaneous movement, and location privacy, HIP is extended with a rendezvous server, which is used to forward HIP control packets to a registered HIP host. HIP provides good security by allowing any pair of hosts to authenticate the public key of their peers about the current IP addresses in use, and it also provides CPU and memory exhausting denial-of-service resistance, mobility, and multi-address multi-homing. However, it does not address the type of flooding denial-of-service attacks and lacks support for multicast or anycast.

Hi3 is developed based on the observation of the similarity of HIP rendezvous server and the basic Secure-i3 infrastructure. It allows direct, IPsec-protected end-to-end traffic while routing HIP control packets in i3, and it applies the DoS protection idea of hiding IP addresses from Secure-i3. In Hi3, the IPsec envelopes and Security Parameter Indexes (SPIs) is used to implement similar DoS protection with Secure-i3 [Nikander04].

In Hi3, basic mobility can be handled at the HIP level. However, when the hosts lose direct reachability information, they need to update their public registrations for initial reachability and keep updating their current location information at the infrastructure. Registration of private triggers is not needed in Hi3 in order to reduce signaling overhead. Because the private triggers are created by the infrastructure during the initial session setup process, the update of private trigger can always be easily distributed to the servers hosting the private triggers by the infrastructure with the information provided by the end hosts.

By taking the advantage of end-to-end mobility provided in HIP and secure indirect mobility provided by secure-i3, Hi3 is highly efficient and robust [Nikander04].

3.3 SPINAT

Hi3 [Nikander04] is a combination of Secure-i3 and Host Identity Protocol (HIP). It is more efficient and secure than Secure-i3, and more flexible and denial-of-service resistant than HIP.

In SPINAT, SPINAT device acts as an overlay router for the end-point identifiers by translating the IP

addresses base on the SPI value and the destination IP address carried in the IPsec payload packets. As IPsec is used, the SPI value can then be used together with the destination IP as an index for end-point identifiers.

SPINAT uses IPsec control plane signaling in registering triggers and establishing communication contexts at SPINAT devices. A trigger consists of an identifier and the IP address of the end host. The communication context includes the information required for address translation, such as SPI values.

A receiver registers a trigger to receive IPsec control plane messages which contain the identifier of the end host in the header. A trigger can be registered either explicitly or transparently. To explicitly register a trigger means that a trigger is registered at a public SPINAT device using Security Association (SA) establishment signaling (i.e. the key exchanging); to transparently register a trigger means that a trigger is registered at an on-path SPINAT device during end-to-end SA update signaling, i.e. the mobility exchange.

Fig. 3 shows an example of trigger registration. In this example, Laptop1 is initially in the Internet and has registered the trigger explicitly at the public SPINAT device, and the public SPINAT device has learned the mapping between the identifier and IP Address of Laptop 1. Then laptop 1 moves from Internet to a private network, so it transparently registers a trigger at the on-path SPINAT device. After the registration process, laptop 1 can start to receive traffic. Once laptop 2 has data to send to laptop 1, it will start an end-to-end key exchange process, and then the SPINAT devices may either establish a communication context by themselves or require laptop 1 to explicitly register an IPsec traffic filter for that session.

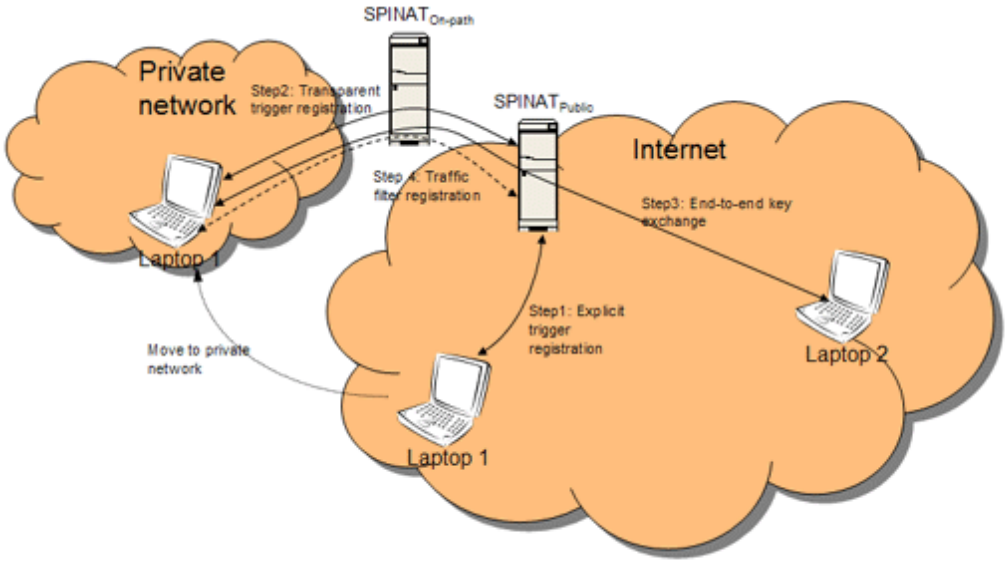

Fig. 3: Trigger registration at public and on-path SPINAT devices

After trigger registration, once the receiver replies to the initial trigger message, either the SPINAT devices on the path will establish the communication context by themselves, or the receiver can explicitly establish the communication context on the SPINAT device where it registers its trigger on.

SPINAT makes it possible to implement scalable overlay routing architectures while keeping the existing security level of IPsec. It can be integrated into middle-boxes in any overlay routing architectures, such as i3 nodes.

3.4 FARA

Forwarding directive, Association, and Rendezvous Architecture (FARA) [Clark03], which is a new organization of network architecture concepts, is also based upon the idea of decoupling of host identifiers

from the network addresses.

FARA defines an abstract set of components and the modular relationships among them, while it leaves undefined many detailed technical mechanisms and design decisions. It allows different forwarding mechanisms to co-exist in a network. M-FARA, which is one particular instantiation of FARA, is designed for mobility and addressing domains.

In FARA [Clark03], host-to-host communication is considered to be a communication between entities through associations, and the packets are exchanged over a communication substrate. In this concept, entity is abstract and may refer to a process, a thread, a set of processes, a computer, a cluster of computers, etc. Association is a logical linkage between two entities, which implies persistent communication state. Communication Substrate refers to the lower-level supporting systems like operating systems and networks.

In FARA, FDs are used to provide routing information. When an entity has a packet to send, it hands that packet to its communication substrate with the header field called destination Forwarding Directive (FD), which contains information needed for eventual delivery of the packet to the destination entity. Also, a reply FD may also be contained for delivery of a return packet to the source entity.

One particular instantiation of FARA: M-FARA architecture is designed for mobility and addressing domains. Thus a specific set of mechanisms for network addressing, forwarding, FD management, and security is provided in M-FARA.

In M-FARA, it assumes that there are multiple domains, and there is a distinct addressing realm in each domain. In this case, an FD is composed of a series of sub-FDs, each of which has meaning within its own domain, in order to traverse each realm along the path. When an entity moves to a new location, a new FD will be computed for the other ends of each existing association. A mechanism should be defined for transforming the FDs so that they can be meaningful at the new location.

To support mobility, M-FARA uses M-agents (mobility agents) as rendezvous points and to update FDs to handle mobility. Each mobile entity registers itself to a M-agent at startup, and the entity informs its M-agent every time it moves so that the M-agent can keep track of the entity's location. Also, the entity sends packets which contain updated reply FD to the remote entities which it has association with so that the remote entities know the location of the mobile entity.

The security of M-FARA is relatively weak. It provides authentication for the initial packet exchange which establishes the association, but it does not verify every packet.

3.5 MOON

MObile Overlay Network (MOON) [Albertengo05] is a centralized WLAN architecture which operates at the session layer and is fully compliant with the Internet Protocol Stack. MOON aims to provide: preservation of communications, higher layers independence from lower layers, efficient routing, efficient handoff, simultaneously mobility, flexibility, and security.

As is shown fig. 4, MOON is formed in a hierarchical structure which is similar to the cellular networks. In this structure, two routing entities are defined:

Enhanced Gateway Router (EGR): located at the border of the domain, connects the domain to the Internet, is the highest hierarchical entity in the structure.

Enhanced Access Router (EAR): an edge router with enhanced authentication and location functionalities, is

the entry point to the overlay network.

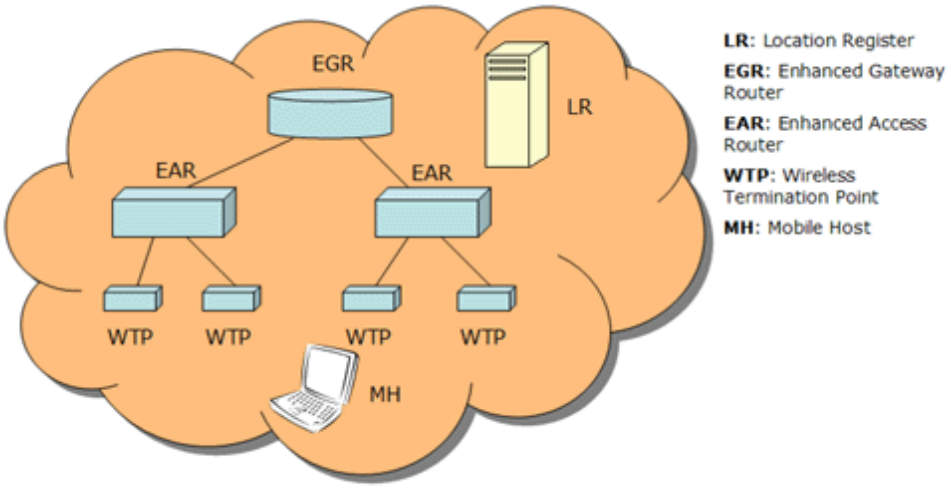

Fig. 4: MOON Architecture

Inside a MH, a session layer handler is used to manage cross bound traffic and handoffs. All the packets sent from host are encapsulated by the handler and delivered to the destination in the overlay network. All the packets directed to a mobile host do not need to be addressed to its home network, but are directly sent to the target.

When a node changes its point of attachment to the network, a virtual interface, which has a fixed IP address, is used to communicate. All the applications refer to this virtual interface instead of the real interface used to communicate so that they are not aware of mobility. The session layer handler even allows both horizontal handoffs and vertical handoffs between two different real interfaces, such as Wi-Fi and UMTS.

A session layer address space is defined in MOON to separate the unique identifier and the topological location of the end host. Since a session between two end hosts is defined only by the identifier, which never changes during handoffs, simultaneous mobility can be handled as well. Also, it achieves efficient routing by directly sending the packets to the destinations.

FAIWL [Albertengo05] (Fast Authentication in Interconnected Wireless LANs) is adopted in MOON as the authentication mechanism. In FAIWL, the MN is assumed to be off and is in the foreign network at the beginning. When the MN is switched on, it fully authenticate itself to gain access to the foreign network, and an authentication context for the MN is created at the end of the authentication process and will be stored in the RADIUS server, thus a key called pairwise master key which is derived from master key will be created for the EAR which currently hosting the MN. A full authentication is not always needed in FAIWL, and the kind of authentication can be decided according to the movement scenarios of the MN. The integration of FAIWL and MOON leads to fast handoff.

3.6 MAT

In Mobile IP with Address Translation (MAT) [Inayat03] is an end-to-end mobility architecture. The address translation in MAT is done at the IP layer. Dedicated servers are used to store the location management information, and the end hosts update the location information in the servers through a secure transmission once they moves to a new location.

Fig. 5 shows the basic communication model of MAT. Two types of addresses are adopted in MAT: Home Address (HoA), which specifies the nodes's identity, and Mobile Address (MoA), which represents the current

location of the node. The network layer is divided into two sub-layers: MAT sub-layer and Delivery sub-layer. Upon receiving a packet from the transport layer, the MAT sub-layer determines the IP address of the IP address Mapping Server (IMS) corresponding to the destination MAT node, and then translate the home address to the destination mobile address according to the address mapping.

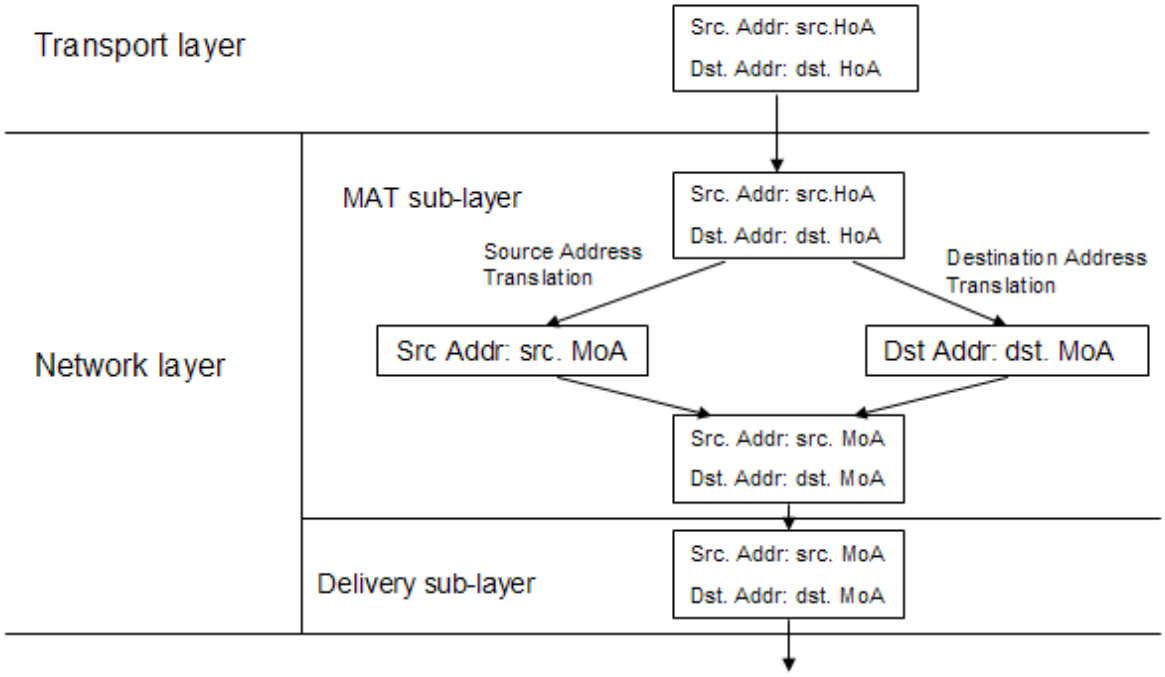

Fig. 5: Basic communication model of MAT

MAT performs a soft handoff and it allows attachment to multiple interfaces during handoff. The priority of mobile address changes while the mobile node crosses the bound between two domains. The CN is informed by sending a packet to it with Home Address Option (HAO) so that the CN will fetch the new mapping association from the IMS.

Back to Table of Contents

---

4 Summary

In this paper we first presented an overview of mobility and the requirement of the host mobility in the Internet, and then we reviewed several novel architectures which are proposed in the very recent years. Some of these architectures are still work in progress, and none of the proposals have gained major acceptance. They all aim at achieving some of the mobility requirements which are not met by Mobile IPv4 and Mobile IPv6. In these proposals, i3 is one of the most prominent proposals for mobility. It introduces a new idea of indirection point by using i3 servers to store the mapping of identifiers and IP addresses. Hi3, SPINAT are both based on the work of i3, although SPINAT is more a block over an architecture, which can be embedded into different architectures. FARA is also a rendezvous-based scheme like i3, and it is more flexible since it leaves some details undefined so that a specific architecture can be built on the base of it. MOON is a session layer architecture, and it can be used to avoid some issues in network layer mobility. MAT is basically an address translation architecture, which translates the destination address according to the address mapping before sending it.

Back to Table of Contents

References

[Perera04] Eranga Perera, Vijay Sivaraman, and Aruna Seneviratne, "Survey on Network Mobility Support," SIGMOBILE Review, 2004.

[Atiquzzaman05] Mohammed Atiquzzaman and Abu S. Reaz, "Survey and Classification of Transport Layer Mobility Management Schemes," 16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications, September 11 - 14, 2005, http://www.cs.ou.edu/~netlab/Pub/TLAYER-PIMRC05-invited-CR.pdf

[Tripathi98] N. D. Tripathi, J. H. Reed, and H. F. VanLandinoham, "Handoff in cellular systems," IEEE Personal Communications, vol. 5, no. 6, pp. 26 – 37, December 1998.

[Perkins02] Ed. C. Perkins. "IP Mobility Support for IPv4," Internet RFC, RFC 3344, August 2002.

[Zhuang03] Shelley Zhuang, Kevin Lai, Ion Stoica, Randy Katz, Scott Shenker, "Host Mobility Using an Internet Indirection Infrastructure," in Proceedings of MOBICOM 2003.

[Stoica02] Ion Stoica, Daniel Adkins, Shelley Zhuang, Scott Shenker, and Sonesh Surana, "Internet Indirection Infrastructure," in Proceedings of SIGCOMM 2002.

[Snoeren00] Alex C. Snoeren and Hari Balakrishnan, "An End-to-End Approach to Host Mobility," in Proceedings of MOBICOM 2000.

[Ylitalo05] Jukka Ylitalo, Patrik Salmela, and Hannes Tschofenig, "SPINAT: Integrating IPsec into Overlay Routing, " 2005.

[Moskowitz05] R. Moskowitz, P. Nikander, "Host Identity Protocol Architecture", Network Working Group Internet-Draft, August 2005.

[Clark03] David Clark, Robert Braden, Aaron Falk, Venkata Pingali, "FARA: Reorganizing the Addressing Architecture," in Proceedings of SIGCOMM 2003.

[Albertengo05] Guido Albertengo, Claudio Pastrone, and Giacomo Tolu, "MOON: a New Overlay Network Architecture for Mobility and QoS Support," IEEE, 2005.

[Inayat03] Riaz Inayat, Reiji Aibara, Kouji Nishimura, Takahiro Fujita, Yoshihiro Nomora, and Kaori Maeda, "MAT: An End-to-End Mobile Communication Architecture with Seamless IP Handoff Support for the Next Generation Internet," 2003.

[Albertengo05_2] G. Albertengo, C. Pastrone and G. Tolu, "Fast Authentication in Interconnected Wireless LANs," submitted to IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, June 2005.

[Chiussi02] Fabio M. Chiussi, Denis A. Khotimsky, and Santosh Krishnan, "A Network Architecture for MPLS-Based Micro-Mobility", In Proceedings of IEEE Wireless Communication & Networking, WCNC02, March 2002.

[Henderson03] Thomas R. Henderson, "Host Mobility for IP Networks: A Comparison", IEEE Network, vol. 17, no. 6, pp. 18-26, November/December 2003.

[Lach03] Hong-Yon Lach, Christophe Janneteau, and Alexandru Petrescu, "Network Mobility in Beyond-3G Systems", IEEE Communications Magazine, July 2003.

[Campbell02] Campbell A. T., Gomez J., Sanghyo K., Chieh-Yih W., Turanyi Z. and A.G.Valko, "Comparison of IP Micromobility Protocols", IEEE Wireless Communications, vol. 9, no. 1, pp. 72 – 82, February 2002.

[Ritter01] Mike Ritter et. al., "Mobile Connectivity Protocols and Throughput Measurements in the Ricochet MCDN System", in Proceedings of MobiCom, 2001.

[Mysore98] Jayanth Mysore and Vaduvur Bharghavan, "Performance of Transport Protocols Over a Multicasting-based Architecture for Internet Host Mobility", in IEEE ICC, 1998.

[Helmy00] Ahmed Helmy, "A Multicast-based Protocol for IP Mobility Support", in Proceedings of NGC, 2000.

[Helmy01] Ahmed Helmy and Muhammad Jaseemuddin, "Efficient Micro-Mobility using Intra-domain Multicast-based Mechanisms", Tech. Report, USC, August 2001.

[Maltz98] David A. Maltz, Pravin Bhagwat, "MSOCKS: An Architecture for Transport Layer Mobility", in Proceedings of IEEE Infocom '98, March 1998.

[Romdhani04] Imed Romdhani, Mounir Kellil, and Hong-yon Lach, "IP Mobile Multicast: Challenges and Solutions", IEEE Communications Surveys & Tutorials, First Quarter 2004, vol 6, no. 1, pp. 18 – 41.

[Narten98] T. Narten, E. Nordmark, and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", IETF Standards Track, RFC 2461, December 1998.

[Thubert02] P. Thubert, and M. Molteni, "IPv6 Reverse Routing Header and Its Application to Mobile Networks", draft-thubert-nemo-reverse-routing-header-01.txt, work in progress, October 2002.

[Ronai03] M.Ronai, A. Petrescu, R. Tonjes, and M. Wolf, "Mobility Issues in OverDRiVE Mobile Networks", IST Mobile Summit 2003, Aveiro, Portugal, June 2003.

[Carter03] Casey Carter, Robin Kravets, Jean Tourrilhes, "Contact Networking: A Localized Mobility System", in Proceedings of MobiSys '03, pp. 145-158, May 2003.

[Lenders05] Vincent Lenders, Martin May, and Bernhard Plattner, "Towards a New Communication Paradigm for Mobile Ad Hoc Networks", Mobile Ad hoc and Sensor Systems Conference, pp 67- 73, November 2005.

Back to Table of Contents

List of Acronyms

MN	Mobile Node

CN	Correspondent Node
VoIP	Voice IP
QoS	Quality of Service
COA	Care of Address
HA	Home Agent
FA	Foreign Agent
BU	Binding Update
MR	Mobile Router
MSM-IP	Mobility Support using Multicasting in IP
ROAM	Robust Overlay Architecture for Mobility
HIP	Host Identity Protocol
DoS	Denial-of-Service
SPIs	Security Parameter Indexes
SPINAT	Security Parameter Index (SPI) multiplexed NAT
SA	Security Association
FARA	Forwarding directive, Association, and Rendezvous Architecture
FD	Forwarding Directive
MOON	MObile Overlay Network
EGR	Enhanced Gateway Router
EAR	Enhanced Access Router
MAT	Mobile IP with Address Translation
HoA	Home Address
MoA	Mobile Address
IMS	IP address Mapping Server

Back to Table of Contents

Last Modified: April 24, 2006.

Note: This paper is available on-line at
http://www.cse.wustl.edu/~jain/cse574-06/ftp/%directory%/index.html.