Future wireless networks: key issues and a survey (ID/locator split perspective)

Chakchai So-In

Department of Computer Science,
Faculty of Science,
Khon Kaen University,
Khon Kaen, 40002, Thailand
Fax: +66-43-342-910
E-mail: chakso@kku.ac.th

Raj Jain*, Subharthi Paul and Jianli Pan

Department of Computer Science and Engineering,
Washington University in St. Louis,
St. Louis, MO 63130, USA
Fax: +1-314-935-7302
E-mail: jain@cse.wustl.edu
E-mail: pauls@cse.wustl.edu
E-mail: jp10@cse.wustl.edu
*Corresponding author

Abstract: Future wireless networks (FWNs) are expected to be a convergence of different kinds of wireless technologies, such as cellular technologies, wireless local area networks (WLANs), wireless metropolitan area networks, wireless sensor networks, and traditional wired networks. The internet protocol (IP) will be potentially adopted as the common networking protocol for diverse networking technologies including the next generation of cellular networks using system architecture evolution (SAE). However, the IP architecture has several known challenges, such as mobility, multihoming, routing scalability, location privacy, path preference selection, etc. One of the greatest problems preventing the networks from overcoming these challenges is that the IP address is contextually overloaded, both as locators and identifiers. As a result, in this paper, we describe the issues of all-IP wireless networks, and survey recent proposals focusing on IP address overloading that can be applied to FWNs.

Keywords: future internet; future wireless internet; next generation wireless networks; NGWNs; mobility; multihoming; location privacy; multi-interface selection; ID/locator split; network architectures; future wireless networks; FWNs.

Reference to this paper should be made as follows: So-In, C., Jain, R., Paul, S. and Pan, J. (2012) 'Future wireless networks: key issues and a survey (ID/locator split perspective)', *Int. J. Communication Networks and Distributed Systems*, Vol. 8, Nos. 1/2, pp.24–52.

Biographical notes: Chakchai So-In received his BEng and MEng from Kasetsart University, Thailand in 1999 and 2001. He also received his MS and PhD from Washington University in St. Louis in 2006 and 2010, all are in Computer Engineering. He is currently a faculty member at the Department of Computer Science, Khon Kaen University, Thailand. In 2003, he was an intern in a CNAP at NTU and obtained CCNP and CCDP certifications. He was an intern at Cisco Systems, WiMAX Forum and Bell Labs during summer 2006, 2008 and 2010, respectively. His research interests include architectures for future wireless networks; congestion control; protocols to support network and transport mobility, multihoming, and privacy; and quality of service in broadband wireless access networks.

Raj Jain is a Fellow of IEEE, Fellow of ACM, Winner of ACM SIGCOMM Test of Time Award and ranks among the top 50 in Citeseer's List of Most Cited Authors in Computer Science. He is currently a Professor of Computer Science and Engineering at Washington University in St. Louis. Previously, he was one of the co-founders of Nayna Networks, Inc. He was a Senior Consulting Engineer at Digital Equipment Corporation in Littleton, Mass and then a Professor of Computer and Information Sciences at Ohio State University in Columbus, Ohio. He is the author of *Art of Computer Systems Performance Analysis*, which won the 1991 'Best-Advanced How-to Book, Systems' Award from Computer Press Association.

Subharthi Paul received his BS from the University of Delhi, Delhi, India, and his Master's in Software Engineering from Jadavpur University, Kolkata, India. He is presently a doctoral student in Computer Science and Engineering at Washington University in St. Louis, MO, USA. His primary research interests are in the area of future internet architectures.

Jianli Pan received his BE in 2001 from Nanjing University in Posts and Telecommunications (NUPT) and his MS in 2004 from Beijing University of Posts and Telecommunications (BUPT), China. He is currently a PhD student in the Department of Computer Science and Engineering at Washington University in Saint Louis, MO, USA. His current research is on the next generation internet architecture and related issues. He is currently a student member of the IEEE.

1 Introduction

Future wireless networks (FWNs) are expected to be a large-scale convergence of many wireless technologies, e.g., cellular networks, wireless local area networks (LANs) or WLANs, wireless sensor networks, wireless broadband access networks, and traditional wired networks. With the advances in networking technologies, the concept of a single user, single host, single interface, and single network will no longer be valid in the context of FWNs.

Users and/or customers in FWNs may be composed of many multi-interface wired/wireless devices leveraging a variety of networking interfaces, such as WLANs, 2G/3G/4G, long term evolution (LTE), (mobile) worldwide interoperability for microwave access (WiMAX) and Ethernet.

In FWN, the users need not be aware of the different behaviours and/or characteristics of the networking media underneath their applications (IEEE Std. 802.21-2008, 2009). The users will be in a high speed networking environment that can inherently support mobility: mobility over large geographic topologies and mobility of users (mobile users) over devices, device multihoming and concurrent multi-interface sessions. In addition, with various networking connections, the service providers and mobile users should be able to choose the best connection (path preference selection) based on cost and quality of service (QoS) requirements.

Due to the nature of a wireless medium, the wireless channel is unpredictable so the channel capacity (effective bandwidth) may be changed over time and distance. Notice that in FWNs, since the users tend to be mobile users, the issue of power consumption is critical. Mobile users or mobile nodes tend to go to sleep, or default to an idle mode when not in use. Unlike WLANs, FWNs will provide a long range, high speed networking capability with low latency, and guaranteed QoS. Similar to a traditional cellular service, air-time charges will be based on the QoS provided.

Moreover, multiple networking interfaces should provide load sharing, load balancing, and higher availability; path characteristic information recommended and/or provided by the service provider can help select the interfaces. Also, the mobile users should be able to maintain their (location) privacy, while the networking environment should provide inherent security.

With the emergence of billions of network-able mobile wireless devices, which may outnumber wired personal computer (PC)'s as early as 2010 (Raychaudhuri and Gerla, 2005), the problem of scalability looming in the current network will be exacerbated. Apart from all these requirements, FWNs' designers also need to evaluate and plan the transition steps from the current network to FWNs, possibly through an incremental deployment of interoperable FWNs' mechanisms over the current network.

The features described above, such as mobility, multihoming, path preference selection, privacy, security, scalability, and deployability, are only a fraction of those needed to be achieved in FWNs. Given this diverse set of requirements, it is extremely difficult to predict the direction of FWNs' evolution.

A common well-known networking protocol used in wired/wireless networks is the internet protocol (IP) (DARPA Internet Program, 1981a; Deering and Hinden, 1998). IP will be potentially used as the networking protocol in FWNs. The two main reasons to select IP in FWNs are as follows: first, IP has been used in the current internet for years, and second, IP will be potentially adopted as the common networking protocol for system architecture evolution (SAE), which is the core networking architecture developed by the 3rd generation partnership project (3GPP) (The 3rd Generation Partnership Project, 2007a, 2007b) for the next generation of cellular wireless networks. SAE will be an all IP mobile wireless network.

In SAE, mobile IP (Perkins, 2002; Johonson et al., 2004), proxy mobile IP (PMIPv6) (Gundavelli et al., 2008) and dual stack Mobile IPv6 (DSMIPv6) (Soliman, 2007) are used for vertical handover to other non-3GPP networks, such as from 3G to WiMAX or to WLANs.

One of the greatest issues of the current IP architecture is the overloading of IP address semantics stated in Jain (2006), Meyer et al. (2007), and ITU-T Y.2015 (2007). The problem is that the IP address acts as a host or a node identifier as well as a node

locator in the routing space. This contextual overloading implicitly binds a host to its point-of-attachment into the network, and there is no independent naming space to represent the end host itself. Thus, every time the end host moves to a new network and obtains a new IP address, it has to do a network handover resulting in binding the host to a new IP address. As a result, transport layer sessions bound to IP addresses are routinely invalidated.

Additionally, with a multi-interface feature, a node or a host may have many different networking interfaces with different types of QoS controls, such as cellular networks (2G/3G/4G) and wireless broadband networks (e.g., WiMAX and LTE) for various applications, including voice, video, TV broadcasting, online games, etc. This multihoming phenomenon offers many advantages, such as seamless mobility (handover/handoff), enhanced availability (fault-tolerance), and traffic engineering (load balancing and load sharing). However, the traditional IP architecture cannot make use of these capabilities due to the address semantic overloading problem.

Such an implicit overloading makes it difficult to support full mobility, multihoming, traffic engineering, privacy, security, etc. As a result, in this paper, we survey recent proposals on identity (ID)/locator split. Our primary focus is on solutions proposed within this architectural paradigm for mobility and multihoming. We also briefly describe the effect of the overloading problem on the routing scalability.

We categorise the mobility techniques by the protocol layers, e.g., network, transport, and application layers. For multihoming, we categorise the techniques into three groups: network-based, host-based, and the combination (hybrid) approaches. We also briefly state and survey multi-interface selection mechanisms for full multihoming support.

This paper is organised as follows. In Section 2, we revisit and briefly describe the four main issues: mobility, multihoming, routing scalability, and deployability. Then, in Section 3, we survey recent proposals and/or techniques in terms of mobility and multihoming aspects, as well as the multi-interface selection issue. We discuss the feasible solutions in Section 4. Finally, the conclusions are drawn in Section 5.

2 Key issues

In general, the four key issues relevant to FWNs are mobility, multihoming, routing scalability, and deployability. Notice that another important issue is security; however, discussions on security mechanisms for FWNs are beyond the scope of this paper.

- *Mobility:* The current internet, designed for stationary end-hosts, does not handle mobility easily within the internet architecture. The issue of mobility relates to handling changes in location and underlying network connectivity of mobile end-systems at each protocol layer. Note that in this paper, we focus on host mobility. Network mobility (NEMO) (Devarapalli et al., 2005), or site mobility, is out of scope of this paper. However, some of the techniques for host mobility can be extended to support NEMO and site mobility as well.
- *Multihoming:* In the past, most hosts/nodes or computers had only one networking interface. Hosts stayed within one network with one egress path. However, multihomed hosts or devices having multiple networking interfaces are becoming

more common. Additionally, users may be multihomed too. Each user can be reached through many different hosts, such as computers, personal digital assistant (PDAs), and cellular phones. We call this user multihoming. Finally, the network that users reside in may have several egress paths as well. This is the so-called site multihoming. All these use the multihoming functionality to support fault-tolerance, load sharing and/or load balancing, and traffic engineering.

- *Routing scalability:* A common solution for IP network sites to allow changing their service providers is to use provider independent (PI) addresses. However, these addresses are not aggregatable and lead to an exponential increase in size of the routing table in default free zones (DFZs) (Meyer et al., 2007).
- *Deployability:* Deployability of new mechanisms is an extremely important factor. The literature is rife with examples of technically superior proposals that have seen limited or no deployment in the real world owing to the lack of a proper and practical deployment plan.

3 Classification and a survey

This section surveys recent techniques and proposals that can be potentially applied to define the architecture of FWNs using the ID/locator split concept. Some proposals may involve or resolve more than just one issue. In general, the ID/locator split techniques are based on an internet indirection model proposed by Stoica et al. (2004) that can potentially resolve one of the greatest issues of the current IP architecture, i.e., the overloading of IP address semantics.

In the current network, the IP address acts as a host or a node identifier in the transport layer protocol, and a locator in the routing space. This contextual overloading prevents the current network from preserving the sessions interfaced to the IP addresses when the hosts change their networks. Such an implicit overloading makes it difficult to support full mobility, multihoming, traffic engineering, etc.

Therefore, in this section, we summarise and point out advantages and/or disadvantages and potential modifications for various proposals.

- *Mobility:* We mainly focus on mobility at the network layer, more specifically, mobile IP and its extensions, i.e., Perkins (2002), Johonson et al. (2004), Gundavelli et al. (2008), Soliman (2007), Ahlund and Zaslavsky (2003), Soliman et al. (2008), Koodli et al. (2005), Le et al. (2006), Campbell et al. (2002), Wakikawa et al. (2009) and Soliman et al. (2009). These techniques are used to preserve and/or maintain the connection or session connectivity regardless of the change in IP addresses. At the transport layer, modifications of transmission control protocol (TCP) (DARPA Internet Program, 1981b) states and new transport protocols have been presented as potential mobility solutions, such as Snoeren and Balakrishnan (2000), Sultan et al. (2002), Stewart et al. (2000), Kohler et al. (2006), Goff et al. (2000), Liu and Singh (2001), Maltz and Bhagwat (1998), Funato et al. (1997), and Haas (1997).

  We also describe the proposals that apply an indirection layer mechanism below the transport layer (Moskowitz et al., 2006; Koponen et al., 2005). Additionally, we

briefly illustrate the use of session initiation protocol (SIP) (Rosenberg et al., 2002; Schulzrinne and Eddy, 2000), adopted by 3GPP (for voice signalling), in terms of connection continuity based on TCP (Vakil et al., 2001). Note that some of these proposals may resolve problems of multihoming as well.

- *Multihoming:* We discuss the multihoming issue and survey recent proposals that solve multihoming using ID/locator split. We focus on Shim6 or Level-3 Shim for IPv6 (Nordmark and Bagnulo, 2007); LISP or locator identifier separation protocol (Farinacci et al., 2007; Meyer, 2008); HRA or hierarchical routing architecture (Xu and Guo, 2008); and MILSA or Mobility and multihoming supporting identifier locator split architecture (Pan et al., 2009; Paul et al., 2009). Notice that the ID/locator separation concept naturally supports both mobility and multihoming.
- *Routing scalability:* To avoid the scalability problem, provider aggregatable (PA) addresses must be deployed.
- *Deployablity:* We discuss this issue within the description of each technique.

Again, in general, the indirection idea is commonly used to resolve all issues we mentioned above. Internet Indirection Infrastructure (Stoica et al., 2004), or i3, is one of the pioneering works in this direction. Briefly, i3 abstractly introduced the triggering concept on overlay networks. The basic idea is that end-hosts transmit the packet with the unique host identities, or IDs. The servers help translate the identities to current locators. The packets are then forwarded using these locators.

Many recent proposals have been derived from the concept of the i3 indirection technique. In general, the proposals can be categorised in several ways. For example:

1 the use of either host-based or network-based approach; the host-based approach is oblivious to the change of network infrastructure underneath; however, all end-hosts need to be modified

2 a hierarchical naming space (domain name system or DNS-like) vs. a flat naming space (distribute hash table or DHT-like) (Paul et al., 2009); each has its own pros and cons in terms of delay latency and scalability

3 an IP address like identity vs. a new naming space. The use of an address-like identity is that it is easy to deploy.

3.1 Mobility

The mobility support makes cellular networks different from traditional wired networks. In FWNs, users and/or customers will frequently move at a high speed from one place to another place, i.e., different networks. Due to the nature of the wireless medium, the wireless channel condition is unpredictable so the channel capacity (effective bandwidth) may change over time and distance. The disconnection also occurs frequently. The mobility issue is about how to maintain the connection/session connectivity and/or how to avoid the disruption operation.

Mobility can be applied to any specific layer (Eddy, 2004) from a physical layer to an application layer. At the physical layer, or Layer 1 (L1), *auto-dial* to select the best connection, or a redundant path, is an example of a L1 mobility technique. There is no concept of connection/session continuity at this layer.

Forward error control (*FEC*) and (*Hybrid*) *automatic repeat request* (*ARQ*) (Jeffrey et al., 2007) are two techniques at a link layer, or Layer 2, that help keep the connection in spite of varying link quality with mobility.

Consider the network layer. When users or hosts move from one network to another, IP addresses are changed. One solution for NEMO (Layer 3) is to keep a permanent node identity regardless of the location. There are many proposals based on ID/locator split ideas used to maintain the mobility at this layer.

With the ID/locator split concept, both mobility and multihoming are generally supported. The main difference among various ID/locator split techniques is either the way the ID/location separation is introduced, or the way IDs are represented.

At Layer 3 and above, the mobility techniques can be categorised into four categories: network layer, transport layer, sub-layer between network and transport, and application layer techniques described in Sections 3.1.1, 3.1.2, 3.1.3, and 3.1.4 respectively.

3.1.1 Network layer approach (*mobile IP and its extensions*)

Mobile IP version 4 and version 6 (Perkins, 2002; Johonson et al., 2004), or *MIPv4* and *MIPv6*, are the well-known approaches used in network layer. Briefly, the idea is that mobile nodes (MNs) have home IP addresses that act as their IDs. When the users move from one network to another network, they inform their home networks (home agents or HAs) of their new IP addresses (care-of-addresses or CoAs).

When a correspondent node (CN) wants to contact this mobile node, CN sends the packets to MNs' home address. The packet is intercepted by the HA and forwarded to the current CoA. The return packets follow the same path from the mobile node to HA and then to CN.

Optionally, route-optimisation functionality can be used by disclosing CoA to CN, allowing a direct communication between MN and CN (Johonson et al., 2004). In this case, there is an issue of ingress filtering because the foreign network routers may not allow the mobile node to use its home address in the source address field because it does not belong to the network.

Two optimisation techniques have been proposed: either use a reverse tunnelling technique (sending packets back toward the home network; however, it introduces the delay), or apply the IP-in-IP encapsulation technique with destination option (more header overhead).

There are several other optimisation approaches used to mitigate the route-to-home network delay and handoff latency, such as *HAWAII*, *cellular IP*, and *HMIP* (*Hierarchical MIP*) (Le et al., 2006). The idea is basically to deploy several HAs in a hierarchical manner at the edge routers.

With HMIP, the binding update is sent to the local HA resulting in delay optimisation. However, the need of synchronisation among HAs remains the key issue. Fast handovers for MIP (Koodli, 2005) allows mobile nodes to configure a new CoA before moving to new networks. When MNs attach to the new base station (BS), the

mobile nodes can communicate using its already-known new address. However, this requires a packet forwarding feature between the old and new BSs.

SAE has adopted the MIP concept, using *proxy-mobile IP* (*PMIP*) (Gundavelli et al., 2008). PMIP uses routers or proxy agents to act on behalf of the mobile nodes. With PMIP (network-based approach), the mobile nodes do not need to support the MIP feature (host-based approach). *Dual stack MIP*, or *DSMIP* (MIPv4 and MIPv6) (Soliman, 2007), concepts are also used by 3GPP for the purpose of backward interoperability between IPv4 and IPv6. DSMIP simply applies the IP encapsulation technique if the mobile node or the proxy agent does not support IPv6.

Consider user location privacy in a mobile wireless environment. When the mobile node is in the home network, a single IPv6 home address represents both node identity and locator. However, when the mobile node is outside the home network, Mobile IPv6 can be treated as an ID/locator split scheme because CoA is used as a locator. The node's home address does not change with its location and, therefore, serves as the node's identity. Notice that the home address provides an indication of the home location of the mobile node and may potentially violate the location privacy of the mobile node.

To clearly separate the function of the identity from that of the locator in Mobile IPv6, the concept of a virtual home address can be used. The IP address is divided into ID and locator spaces.

A location non-indicating virtual home address, called *virtual identifier* (*VID*), is used. VID is pre-defined and randomly assigned by the service provider. This VID is a 128-bit address format from the ID space used to represent the node's identity.

Note that VID is permanent and is not bound to any physical home networks and/or locations. In other words, VID is used even when the mobile node resides in the home network. Similar to Mobile IPv6, the IP-in-IP encapsulation technique is applied in that the mobile nodes update their CoAs when they are in different locations and/or networks.

3.1.2 Transport layer approach (*TCP modifications and alternate transport protocols*)

Since the connection disruption occurs at the transport layer, it may be also feasible and/or practical to focus on the mechanisms to maintain the connection at this layer.

In this section, we focus on TCP, not UDP, because UDP does not require a connection. We categorise the transport layer approach into two categories: TCP modifications and new or alternate transport protocols.

Several proposals have modified TCP states, such as *TCP migration* (Snoeren and Balakrishnan, 2000) that is a new TCP state is added so when the change of IP address occurs, the TCP connection will not break, but instead is kept alive until a new binding is made.

(*Mobile*) *STCP* or *steam control transmission protocol* (Stewart et al., 2000) and *datagram congestion control protocol* or *DCCP* (Kohler, et al., 2006) are examples of alternate protocols at the transport layer. Instead of a connection based on an IP address and TCP port number, (mobile) STCP uses the association concept, that is, each association allows more than one IP address to be mapped so the connection (or association) is alive as long as at least one IP address is available.

The idea of DCCP is to transfer the connection endpoint from one address to another address, and then the connection state is remapped to use the new address. DCCP is a message-oriented transport layer protocol that supports Explicit Congestion Notification (ECN). DCCP was not designed to provide reliable in-order delivery, and so is well-suited for steaming media, online gaming, and internet telephony applications.

To maintain the connection at the transport layer, another technique, called *TCP-freeze* (Goff et al., 2000; Liu and Singh, 2001), was proposed. In fact, the idea of a TCP freeze option is already built in the conventional TCP stack. Whenever a receiving end-host does not have enough buffer space left, it informs the sender to stop packet transmission by sending a zero window packet back to the sender. Then, the sender keeps probing if the receiving window is more than zero. Finally, the sender moves to the normal operation, when the window update packet with a non-zero window is sent back.

The main advantage of this technique is that no modification is required, and also the buffer space needed is reduced substantially. So-In et al. (2009) applied a TCP-freeze technique to a Mobile IP environment for preventing the short-term disconnection due to address changes.

For all the techniques we described above, either sender or receiver, or both need to be modified. We call these host-based approaches. However, these may be impractical because all nodes need to be modified.

MSOCK (Maltz and Bhagwat, 1998) is a redirection mechanism, to redirect the traffic through a split-connection proxy server, and that proxy maintains all connection states. We call this a network-based approach. The proxy can transparently splice the connection to an alternate server while it ensures the connection consistency, e.g., byte sent/received information.

The splice technique is used to simulate a single direct TCP connection although there are two separate TCP connections (mobile-to-proxy and proxy-to-server). During the client disconnection, the proxy keeps the server connection open and splices the new client connection when the client link is resumed, using a unique connection identifier. During the operation, the proxy acts as a relay server because it does not send the acknowledgement packet on behalf of the mobile node; however, it still maintains the mapping of the sequence space, i.e., a sequence number and an acknowledgement number.

3.1.3 Sub-layer approach

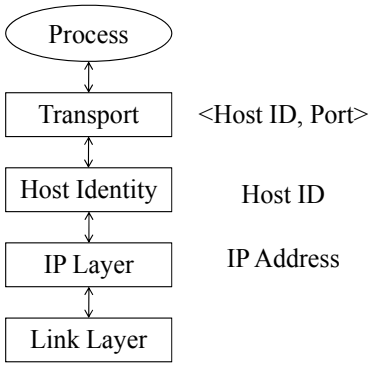
Normally, this approach introduces a sub-layer of indirection below the transport layer. The TCP connection is bound to a unique virtual address, or node identity. At this sub-layer, the virtual address is mapped to the actual or current IP address (or locator). The logical address, or identity, is named in several ways, such as the use of 32-bit user public key (Funato et al., 1997) combined with a random number, IP address, and TCP port number (Haas, 1997).

Host identify protocol (*HIP*) (Moskowitz et al., 2006; Koponen et al., 2005) introduced a new naming space to represent the host identity. Host identifier tag (HIT) is used to represent the host identity (HI). HIT is a 128-bit hash of the public key of the host. The idea is to simply bind the transport socket to HIT in this context. Then, HIT is translated to an IP address in the kernel. With this new naming scheme, the identity

represented as HIT is separated from the locator, or IP address. As a result, the identity persists regardless of the change of locators.

The mapping between a host name and HIT can be as follows: the IPv6 address can be derived from DNS as AAAA records (the first level of mapping). This mapping can also be stored in distributed hash table (DHT). However, to improve the mobility and to reduce the latency, instead of users receiving the destination IP address from DNS, the DNS resolution results in the IP address of the rendezvous servers, which are distributed hierarchically. Then, the rendezvous servers act as the mobility anchors in order to resolve ID to the current locators (the second level of mapping). Figure 1 shows the host identity architecture.

Figure 1 Host identity architecture



To sum up, aside from an IP address and DNS naming space, HIP introduced a new secure naming space. This naming space represents the host identity bound to a new sub-layer below the transport layer. As a result, the connectivity can be maintained regardless of the change of IP addresses.

With the HIP concept, the users request the identity, not its locator, from DNS, and then the locators from the rendezvous server. To maintain user location privacy, a proxy can be used; that is, the user does the mapping resolution from the host name, or Fully Qualified Domain Name (FQDN), to ID, and then the proxy does the ID to locator resolution on behalf of the host. The proxy will rewrite or place the locators on the packets that can be removed at the destination proxy.

Secure i3 is a modification of i3 (Adkins et al., 2003), described in the beginning of Section 3. Secure i3 is used to provide the protection of a denial-of-service (DoS) attack by not allowing other users to see the IP address of the end host. In Secure i3, two types of triggers were supported: public (well-known services) and private (actual communication between sender and receivers).

Host identity indirection infrastructure (*Hi3*) (Nikander et al., 2004) was introduced to apply the concept of Secure-i3 and HIP together. The main weakness of Secure-i3 is that all traffic is directed through the overlay server, which increases the amount of network traffic. In addition, HIP does not address the issues of multicasting and anycasting. Hi3 allows IPsec-protected end to end flow (using HIP) and runs on the

indirection infrastructure (using i3) to route the HIP control packets. The actual IP address is hidden by the infrastructure.

3.1.4 Application layer approach

SIP (Rosenberg et al., 2002; Schulzrinne and Eddy, 2000) is the well-known approach for application layer mobility. It is designed for UDP applications in which there is no concept of a transport layer connection. In SAE, voice over IP (VoIP) applications will use SIP to establish voice calls over IP networks. Universal resource identifier (URI) is used to represent the user identity (ID). This URI consists of ID + SIP domain (realm); for example, *sip: name.lastname@mydomain.com* and *sip: 3149352113@telephone.com.*

The session is bound to SIP URI, not IP addresses. Real-time applications, such as IPTV, also use SIP. Whenever users move, the users send the new binding update to a SIP server to renew the mapping, i.e., URI to IP address, and the communication continues without disconnection. SIP URI is used to represent the user identity over the entire session duration.

Note that SAE needs to deal with all kinds of applications including TCP-based applications, such as file transfer protocol (FTP) and World Wide Web (WWW). When users move, and the IP address changes, TCP connections are terminated. Therefore, a *SIP-eye* (Vakil et al., 2001) extension has been introduced to mitigate this problem. The idea of the SIP-eye extension is similar to mobile IP. When a user moves, the user informs the CN about the change of IP addresses with a SIP INFO message. The inbound IP-TCP packet is encapsulated inside the new IP packet.

With this encapsulation, the TCP connection is still maintained. However, a SIP-eye agent needs to be installed in all hosts (consider a host-based approach). This approach, like other encapsulation approaches, leads to an increase of header overhead.

3.2 Routing scalability

Routing scalability is one of the key issues for FWNs due to an exponential growth of the size of the IP routing table (Meyer et al., 2007). As explained earlier, if a site uses PA addresses, it has to renumber all its hosts when it changes providers.

However, if it uses PI addresses, these addresses are not aggregatable, and thus results in an increase of the routing table (Devarapalli et al., 2005; Launois and Bagnulo, 2006). With an ID-locator overlay, it is possible to use PI addresses as IDs and PA addresses as locators. With this approach, only PA addresses are used in the core, and the routing scalability issue is resolved. Here, the key consideration is where the ID-locator translation is implemented – in the host or in the edge router.

In general, on one hand, the goals of Mobile IPv6 and its extensions and HIP are to solve the mobility issue. Shim6 is for multihoming. Consider HIP in particular, in case the global naming space is introduced, PA addresses can be used for routing purposes, and thus results in mitigating the routing scalability issue.

On the other hand, LISP, six/one, and enhanced MILSA' goals (see also Section 3.3.1) are to resolve the routing scalability issue. Normally, the concepts of address rewriting and/or RLOC are used to mitigate the routing scalability issue.

3.3 Multihoming

Multihoming is used primarily for fault-tolerance, load sharing and balancing, and traffic engineering. The multihoming feature will be more common in FWNs. In general, having more than one networking interface and/or more than one egress path implicitly states that one or more IP addresses exist within one of more IP networks.

As we described at the beginning of Section 3, it is possible to solve both mobility and multihoming problems by ID/locator split. However, most ID/locator split proposals have concentrated on mobility while a few (e.g., shim6) concentrate on multihoming. In ID/locator split, the host name is first mapped to the host ID, and this ID is used to create the TCP connection. ID is translated to one of many locators (IP addresses for multiple interfaces) depending upon the multihoming policies and requirements.

Recent proposals and extensions for supporting multihoming can be categorised into three groups: network-based (LISP) (Farinacci et al., 2007; Meyer, 2008), host-based (Mobile IPv6, HIP, and Shim6) (Johonson et al., 2008; Moskowitz et al., 2006; Koponen et al., 2005; Nordmark and Bagnulo, 2007), and the combination or hybrid (six/one, HRA, and MILSA) (Vogt, 2007; Xu and Guo, 2008, Pan et al., 2008, 2009) approaches.

3.3.1 Network-based approach (*LISP and LISP-ALT*)

3.3.1.1 LISP (Farinacci et al., 2007; Meyer, 2008)

Introduced by Cisco Systems, LISP is designed primarily to mitigate the routing scalability issue. However, it can also provide support for multihoming. Two main concepts are: routing locators (RLOCs) and endpoint identifiers (EIDs). RLOC describes the device attachment to the network, and EID identifies the device itself. Thus, EIDs/RLOCs are used to implement the ID/locator split scheme. RLOCs are allocated hierarchically or aggregately using PA addresses. EIDs are allocated for use within the organisational boundary.

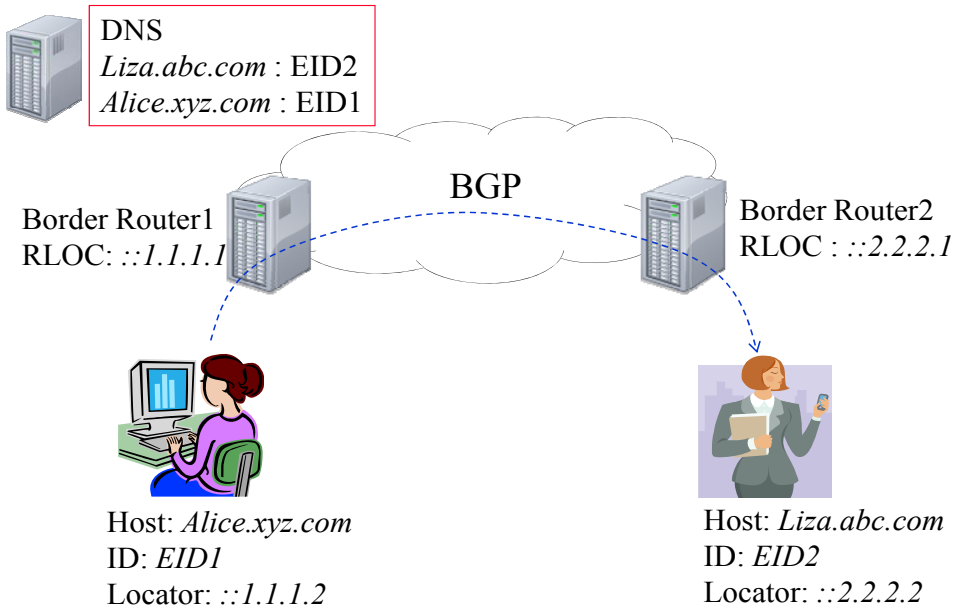
The two main ideas of LISP are the mapping and encapsulation, as well as the address rewriting. The mapping and encapsulation mechanism is illustrated in Figure 2. The source *alice.xyz.com* (within the domain *xyz.com*) with an identifier *EID1* queries the domain name server (DNS) to resolve the destination *liza.abc.com* (within the domain *abc.com*) with an identifier *EID2*. Then, *alice.xyz.com* forwards packets to its border router (gateway).

Note that the assumption that EIDs are locally routable is required. The border router of *xyz.com* maps *EID2* to the corresponding RLOC, usually the address of the border router of the destination *abc.com* domain. Next, the *xyz.com* border router encapsulates the packets with a new header, i.e., a set of a RLOC pair (::1.1.1.1/::2.2.2.1), and forwards the packets through the core IP network. The destination border router decapsulates the packets, and then forwards these to the final destination, *liza.abc.com*. With this approach, *Alice* does not need to know the *Liza*'s locator, ::2.2.2.2.

In this scheme, there are two mapping levels: from the host name to EID and then from EID to its RLOC. The first mapping is stored at a legacy DNS. The second mapping can be queried in two different ways. First, the ingress border router, the so-called LISP ingress tunnel router (ITR), queries RLOCs from the authoritative egress tunnel router (ETR) by sending the data probe, and then that ETR responds with a map-reply message.

Second, ITR can send the map-request message with the particular EID-to-RLOC map to the authoritative ETR.

Figure 2 LISP encapsulation/decapsulation example (see online version for colours)



Host name	*Source locator*	*Destination locator*	*Source EID*	*Destination EID*
alice.xyz.com			EID1	EID2
Border Router 1	::1.1.1.1	::2.2.2.1	EID1	EID2
Border Router 2	::1.1.1.1	::2.2.2.1	EID1	EID2
liza.abc.com			EID1	EID2

Aside from the mapping/encapsulation technique, the other approach, address rewriting, is to allow the border routers to rewrite the address in the IP packet. This approach does not require an additional header. As shown in Table 1, the 128-bit IPv6 address is divided in two parts, and the most significant 64 bits are used as the RLOC and the lower 64 bits as the EID.

Table 1 Address rewriting example

	Source address (*64b.64b*)	*Destination address* (*64b.64b*)
alice.xyz.com	N/A.EID1	RLOC2.EID2
Border Router 1	RLOC1.EID1	RLOC2.EID2
Border Router 2	RLOC1.EID1	RLOC2.EID2
Liza.abc.com	RLOC1.EID1	N/A.EID2

In this table, the scenario is similar to the mapping/encapsulation technique described in Figure 2 in that *alice.xyz.com* sends packets to *liza.abc.com*. First, *Alice* queries the *Liza's* identifier, *RLOC2.EID2*, and then it forwards the packets out with an unspecified source RLOC. When the packets reach the border router, *Border Router 1*, it fills in *RLOC1*, and then forwards the packets to the destination. Whenever the packets reach the destination, *Border Router 2*, *RLOC2* is removed, and the packets are forwarded to the destination, *liza.abc.com*. At this point, no encapsulation is required, but the fully specified destination address, *RLOC2.EID2*, is known by *alice.xyz.com*.

3.3.1.2 LISP-alternative-topology (LISP-ALT) (Farinacci et al., 2007; Meyer, 2008)

This enhanced version was proposed to minimise the change of software and hardware for deployment purposes. The key idea is to build an overlay network using generic routing encapsulation (GRE) and have border gateway protocol (BGP) run on top of that. ALT routing information base (RIB) consists of the EID prefixes and the associated next hop. The LISP-ALT routers use the external BGP protocol to talk to each other to propagate the EID prefix information. Notice that there are no changes to BGP and GRE.

To summarise, LISP applies the ID/locator split concept as the EID/RLOC separation. LISP supports mobility and multihoming. RLOC can be deployed hierarchically and aggregately. However, there is no discussion on how to construct the secure ID. A combination of some special values and MAC addresses can be used as an example of EID. Again, this mapping/query is done by a DNS resolution process. The EID to RLOCs resolution process is scalable and is based on a BGP routing look-up process.

In addition, to support fast endpoint mobility, Mobile IPv6 is recommended based on the use of EID. To enable LISP support full mobility without Mobile IPv6 integration, LISP may use a secure ID like using HIT in HIP as EID, and then use only the lower 64 bits as EID and keep the top 64 bits as RLOCs. The mapping of EID to RLOCs can be stored at a hierarchical DHT or may be exchanged like a traditional LISP-ALT scheme.

3.3.2 Host-based approach (*mobile IP, HIP, and Shim6*)

3.3.2.1 Mobile IP and its extensions (Perkins, 2002; Johonson et al., 2004)

In general, Mobile IPv6 and its extensions can provide full mobility. The permanent home address (HoA) acts as the host/node identity; the CoA serves as the node locator. Traditionally, Mobile IP cannot support multihoming because only a single CoA, or locator, is bound to each host.

However, Ahlund and Zaslavsky (2003) and Wakikawa et al. (2009) proposed the multihoming feature added to Mobile IPv6. Briefly, the idea is to allow multiple CoAs to register at the HA and the CN. In this extension, there is no discussion on how to maintain the operation during roaming. Therefore, especially in case of a multihomed device or a user with different connected networks, the cooperation among the foreign agents for all roamed networks is required.

Figure 3 HA chaining example (see online version for colours)

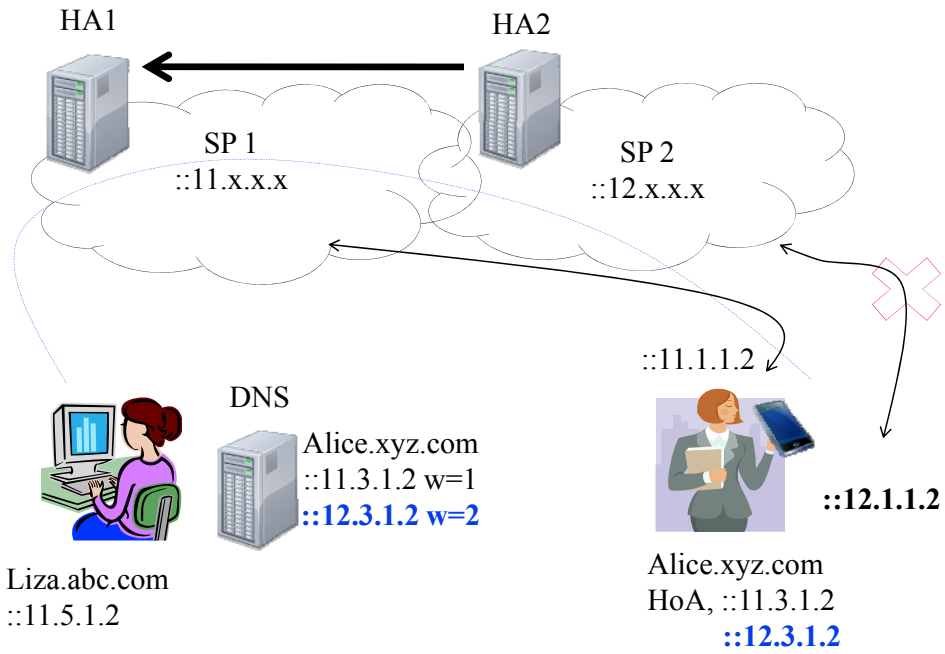


Figure 3 illustrates the HA chaining concept. The basic idea is as follows: *Lisa.abc.com* contacts *Alice.xyz.com* with her two HoAs ::11.3.1.2 and ::12.3.1.2. Alice prefers the ::12.3.1.2 interface, i.e., higher weight, as the primary connection, or identity. Suppose the connection of this path toward the network ::12.*x.x.x* is broken, or Alice wants to switch her service provider or path perhaps due to the service charge constraint. In this scenario, the cooperation between two service providers is required. For example, HA_2 needs to forward the packets directed to Alice toward HA_1 until Alice informs the domain name server (DNS) to change her preferred path so that Lisa can contact Alice toward ::11.3.1.2 instead.

3.3.2.2 Level-3 Shim for IPv6 (Shim6) (Nordmark and Bagnulo, 2007)

This is also a host-based approach using the ID/locator split concept; there is no modification in the network infrastructure. Originally, Shim6 was introduced to solve the site multihoming problem by inserting a sub-layer below the transport layer. This layer is used to hide the change of IP addresses from transport protocols. Shim6 selects one of the IPv6 addresses (locators) as the node identity, called upper layer identifier (ULID).

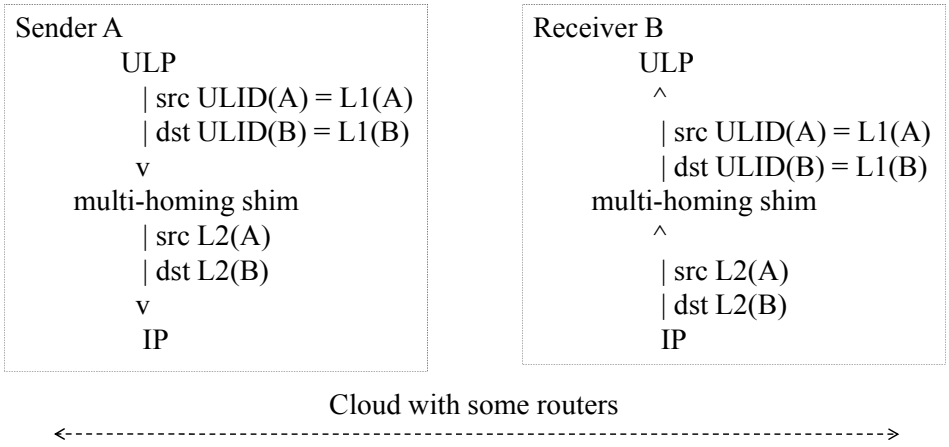
Unlike HIP, Shim6 does not introduce a new naming space. The redirection occurs between the identity and its locators (both from Shim6 current addresses). This identity remains the same for the period of the session or connection. With this redirection concept, traffic engineering, such as load sharing and/or load balancing, are feasible, but the mobile host needs more information on the routing path. Site multihoming can be achieved, if and only if, the update of DNS is fast. To mitigate routing scalability, the locator addresses should be assigned aggregately.

There are some problems with Shim6. For example, by using one of the locators, IPv6 addresses, as the node ID, or ULID, this selected ID will continue to be used without any changes for a long time. ULID may be reassigned to other sites while it is currently used by the mobile host. As we described previously, when the host and/or network are renumbered, the identity and locator mapping depends solely on the update of DNS.

Consider user location privacy, similar to Mobile IPv6, Shim6 can partially support it, if and only if, the mobile nodes move and still maintain the old identities. So, in this case, the CN does not know where the mobile nodes are. However, when the mobile nodes update their identities in DNS, user location privacy is no longer maintained.

Figure 4 shows a diagram of the mapping structure when the locators are changed. *Sender A* transmits packets to *Receiver B*. Upper layer protocol (ULP) is used to map ULID and its locators (L). *Lx*(*Y*) denotes the locator number *x* of host *Y*. This ULID persists over the session period. In this example, source *ULID*(*A*) is mapped to *L1*(*A*), and destination *ULID*(*B*) is mapped to *L1*(*B*). At *Sender A*, the current source locator is *L2*(*A*), and destination locator is *L2*(*B*).

Figure 4 Mapping with changed locators



In Shim6, the concept of a unique context identifier is also used to uniquely represent the session. The minimum combination for one context tag consists of a peer ULID, a local ULID, and a local context tag. The context tag is in the Payload extension headers: 47 bits number plus one bit to differentiate Shim6 signalling from the Shim6 header. Shim6 also supports a context forking process, that is, it allows more than one current location pair for each context.

To summarise, Shim6 was introduced to allow multiple locators (IPv6 addresses). One of these is chosen to represent the host identity. By using IPv6 address format as ID, no new naming space is required, but the ID is not secure. In addition, multihoming is supported (by allowing multiple locators); mobility is partially supported. By using both ID and locator from the same IP address space, the address duplication may occur. One possibility to overcome this issue is to combine Shim6 and Mobile IPv6 to resolve the full end-host mobility.

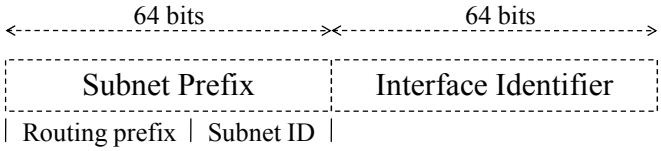
3.3.3 Hybrid approach (*Six/One, HRA and MILSA*)

3.3.3.1 A solution for routing and addressing in IPv6 (Six/One) (Vogt, 2007)

This is a combination of Shim6 and address rewriting techniques. Six/One uses the idea of Shim6, that is, to allow multiple locators (a set of IPv6 addresses from service providers), the so-called *active addresses*. Then, Six/One uses one of the locators as the node identity, the so-called *primary address*. The primary address is unchanged over the session period. A group of IP addresses is called an *address bunch*.

Unlike Shim6, the host address differs in the high-order bits, the so-called routing prefix. In addition, Six/One uses an address rewriting technique to change these routing prefix bits. Figure 5 shows the IPv6 address structure used for six/one. The original source address, filled by the mobile host, is only a suggestion as to where its packets should be routed, but this can be replaced by the network provider. The top 64 bits or subnet prefix consists of both routing prefix and subnet ID. When re-homing and re-writing occur, the subnet ID remains the same; only the routing prefix is modified.

Figure 5 IPv6 address structure



Since six/one allows the address rewriting so as to allow the packets to go via a different provider, a host must be aware of its own bunch addresses and its correspondent's bunch addresses. This can be done in a per-communication-session context. This context, created during a session establishment, uniquely represents the session or communication.

Table 2 shows an example of a mobile host's contexts on session establishment. In this example, the mobile host chooses address *RP1a:SID1::IID1* as its primary address. *RP1a* is the routing prefix. *SID1* is the subnet. *IID1* is the interface identifier. There are two addresses in the bunch with different routing prefixes: *RP1a* and *RP1b*. The CN address is *RP2a:SID2::IID2* as the primary address. *RP2a*, *SID2*, and *IID2* are the routing prefix, subnet, and interface ID, respectively.

Table 2 Host's contexts on session establishment

Mobile host	*Local context*	*Remote context*
Context ID	CID1	(unknown)
Primary address	RP1a:SID1:IID1	RP2a:SID2:IID2
Active address	RP1a:SID1:IID1	RP2a:SID2:IID2
Address bunch	RP1a:SID1:IID1	RP2a:SID2:IID2
	RP1b:SID1:IID1	(rest unknown)

In this technique, the mobile host does not know the remote context. At this step, the primary and remote addresses are used as the source and destination addresses. The first packet carries the IPv6 destination option extension with context setup option so that the context identifier is set up.

Table 3 shows CN's contexts. Once the CN receives the session establishment from the mobile host, it may update the list of address bunches (in this example, with two more routing prefixes: *RP2b* and *RP2c*). Then, it sets up the context identifier, *CID2*. This example shows the address rewriting from the routing prefix *RP1a* to *RP2b* by the service providers.

Table 3 Correspondent host's contexts on session establishment

Correspondent node	*Local context*	*Remote context*
Context ID	CID2	CID1
Primary address	RP2a:SID2:IID2	RP1a:SID1:IID1
Active address	RP2a:SID2:IID2	RP1b:SID1:IID1
Address bunch	RP2a:SID2:IID2	RP1a:SID1:IID1
	RP2b:SID2:IID2	RP1b:SID1:IID1
	RP2c:SID2:IID2	

To summarise, Six/One uses one of the locators, IP addresses, as the node identity (similar to Shim6) and allows address rewriting at the top 64 bits of the IPv6 address space. Similar to Shim6, mobility and multihoming are supported. A new naming space is no longer required, but the node identity is not a secure ID. Pros and cons of six/one are derived from Shim6. Six/One does not support user location privacy because the active address is sent, and then a context identifier is used instead. To resolve this issue, a technique like proxy-assisted address rewriting can be used to remove the active address before forwarding packets to the CN.

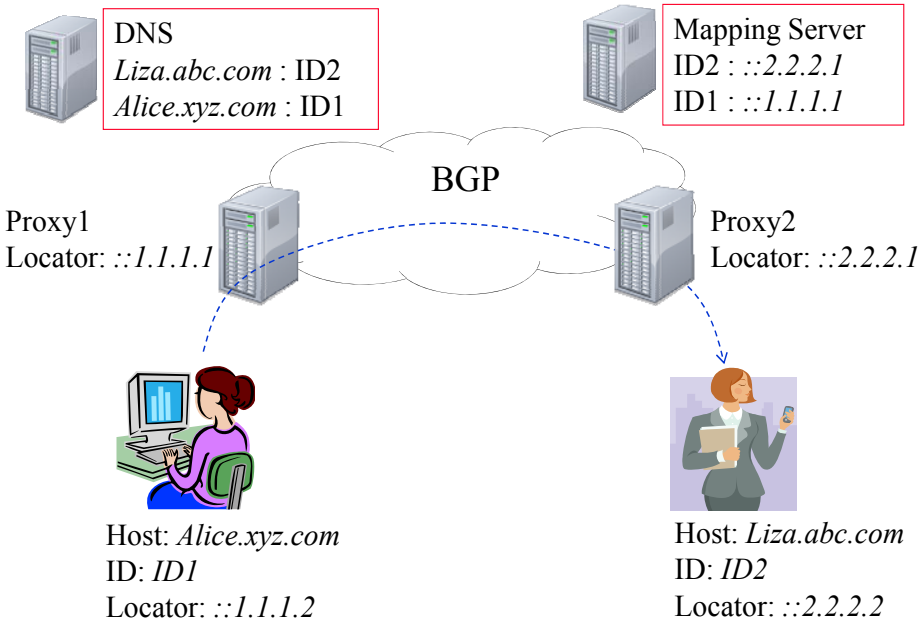
3.3.3.2 HRA (Xu and Guo, 2008)

The design principle of HRA is to mitigate the routing scalability issue. HRA uses the HIP concept to provide a secure identity (ID). Unlike HIP, HRA's IDs combine a hash value of the host identity (HIT) with an administrative domain ID.

This administrative ID (AD ID) is labelled hierarchically. Similar to HIP, there are two levels of mapping: from host name to ID and from ID to locator. The first mapping step is stored at DNS. The second mapping, from HIT to locator domain (LD) and Locators, is stored at distributed hash table (DHT) and/or hierarchical DHT systems (Gares-Erice et al., 2003).

Notice that LD was introduced to represent the independence of address spaces from the IP address scheme. LD consists of a set of locators. The locator is not required to be an IP address. In other words, HRA offers one more level of an ID/locator mapping process by introducing LDs; LD is like an IP subnet, and the locator is like a MAC address.

Figure 6 User location privacy example (see online version for colours)



	Source locator	*Destination locator*	*Source ID*	*Destination ID*
alice.xyz.com	::1.1.1.2	::1.1.1.1	ID1	ID2
Proxy1	::1.1.1.1	::2.2.2.1	ID1	ID2
Proxy2	::1.1.1.1	::2.2.2.1	ID1	ID2
liza.abc.com	N/A	N/A	ID1	ID2

In addition, HRA recommends the idea of an inter-LD routing protocol. Due to the hierarchical manner of LD, this routing mechanism improves the stability of the traditional inter-domain routing. The inter-LD routing protocol can use the BGP extension to exchange LD reach-ability information. The traditional prefix-based routing is still used within each LD.

To sum up, HRA includes a hierarchical part to its secure ID. It supports both mobility and multihoming. To mitigate the scalability issue, again a hierarchical ID is used, and a hierarchical DHT is recommended. When HRA's IDs are deployed, the PA address must be used. There is no discussion on user location privacy. However, the concept of address rewriting at the proxy and/or the border router can be applied. We illustrate the modification in Figure 6.

This extension is similar to LISP, described in the previous section. In general, this modification is as follows: *alice.xyz.com* queries the *liza.abc.com* identifier, or *ID2*, from DNS. Then, the mobile node forwards the packets to *Proxy1*. *Proxy1* looks up the destination locator, i.e., IP address in this scenario, of *Proxy2* and rewrites the locator address.

When the packets reach *Proxy1*, *Proxy2* removes the locator and forwards those packets to *liza.abc.com*. Note that there is always a trade-off between the user location privacy and the complexity of a proxy mapping mechanism. Aside from avoiding an

additional proxy mapping step, the mobile host can also directly look up the destination locator and choose a path. However, the user location privacy cannot be maintained.

3.3.3.3 MILSA (Pan et al., 2008)

This technique is also one of the ID/locator separation proposals and uses a Hierarchical URI-like Identifier (HUI). This HUI consists of two parts: flat and hierarchical parts. The flat portion represents the object identity uniquely within a particular administrative domain.

Similar to LISP, the identity to locator mapping resolution is done by the hierarchical border gateway routers, called realm-zone bridging servers (RZBS). Note that realm is a hierarchical group of objects that logically belong within the same administrative domain. Zone is a topologically aggregated physical network. The mapping concept is similar to that is found within a dynamic hash table overlay network. However, most parts of RZBS infrastructure are preconfigured. Similar to HIP and Shim6, a new mapping sub-layer is inserted beneath the transport layer.

Enhanced MILSA (Pan et al., 2009): an extension to MILSA, enhanced MILSA specifies details of how to form two parts of the HUI structure. The first flat part is the hash of a user public key, which is similar to HIT in HIP. The second part, or hierarchical part, is somewhat similar to current domain names; however, it strictly defines the organisational affiliation. In addition, HUI is used at the transport layer not the application layer. Similar to SIP, the hierarchical part is human-readable. FQDN (host name) is mapped to HUI which are then mapped to locators.

To sum up, (enhanced) MILSA introduced the concept of a combination of human readable ID and cryptographic ID as a single node identity. Similar to other ID/locator split proposals, mobility and multihoming are supported because the connection will be bound to the identifier instead of the IP address (locator).

3.4 Multihoming multi-interface selection

In this section, we briefly describe the issue of multi-interface selection in cooperation with a multihoming feature. All preceding techniques we described mainly focus on how to preserve or maintain a connection by introducing the node identity. Then, this identity can be dynamically mapped to the locations.

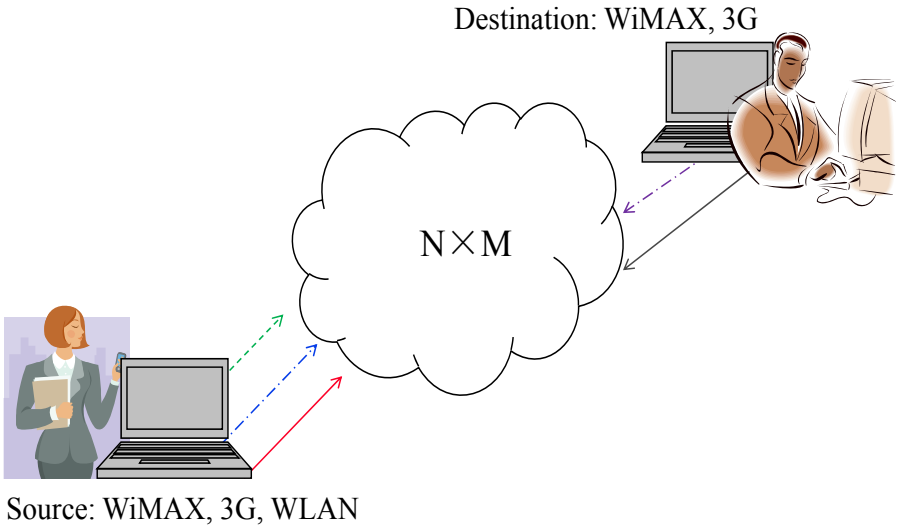
However, none of those proposals describe how to select the networking interface. In general, the problem of selecting a networking interface has been investigated in the concept of always best connected (Gustafsson and Jonsson, 2003). The basic idea is to find the best single active interface given the interface and network constraints, such as bandwidth, power consumption, and accessing technology.

Consider a mobile wireless environment. Recently, a *flow binding option* (Soliman et al., 2009) was introduced in Mobile IPv6. The idea is to map a particular flow to a particular interface. Note that this extension is based on the use of *multiple CoA registrations* (Wakikawa et al., 2009) which we briefly described earlier. To meet user requirements and QoS control parameters, a mechanism similar to the policy-oriented model is required for Mobile IP (network layer approach).

However, a limitation of single best active interface selection is that the multihoming feature cannot be fully utilised with only one interface. For example, suppose a device consists of both 3G and WiMAX interfaces. Also, suppose the device is power-line operated and paid for by the flat rate fee; therefore, using two interfaces simultaneously will achieve twice as much throughput as available with just one interface.

Figure 7 shows a simple configuration for the end-to-end multihoming. In this set up, there are three different interfaces at the source: WiMAX, 3G, and WLAN; and only two at the destination: WiMAX and 3G. There are $3 \times 2 = 6$ possible paths between these users. The path characteristics, e.g., the path throughput, congestion level, loss probability, end-to-end delay, and so on, may be different from one path to the other.

Figure 7 Example of end-to-end multihoming (N × M = 3 × 2) (see online version for colours)



A policy-based QoS and user requirement model incorporating the constraints, such as power consumption, air-time charges, and completion time, can be formulated to make use of the best *N* active interfaces. Each application can use just one of these *N* interfaces or can spread its traffic over some or all of *N* interfaces, and thus achieve the throughput aggregation.

In fact, there have been several proposals to resolve the throughput aggregation problem. We can consider those by a layer perspective, such as the session, transport, network, and link layers. For example, *session layer mobility management (SLM)* (Landfeldt et al., 1999) was introduced to achieve the mobility at the session layer. SLM allows multiple transport connections for each application, which results in higher throughput.

At the transport layer, most approaches are based on creating multiple sub-flows or virtual connections which are bound to a traditional TCP connection. Each sub-flow is per TCP end-to-end connection path. A wrapper at the transport layer, called *pTCP* (Hsieh and Sivakumar, 2002), was used to allow multiple virtual connections to

aggregate the total throughput. Each virtual connection has its own sequence space, congestion control, and setup/teardown.

Similarly, *mTCP* (Zhang et al., 2004) also makes use of multi-path TCPs; however, it uses only a single sequence space, and requires that the acknowledgement packets be sent along the same forwarding path. Costin Raiciu and his colleagues (Raiciu et al., 2009) derived the congestion windows to balance multiple paths and also contributed a multi-path bandwidth aggregation mechanism given a packet loss constraint. Recently, Ford et al. (2010) proposed an architecture guideline for multipath TCP development.

A simple link scheduling algorithm, e.g., a *deficit round robin (DRR)* and a *weighted fair queue (WFQ)* (Adiseshu et al., 1996), was proposed to balance the per-packet transmission. A modified version of earliest deadline first (EDF), called *earliest delivery path first (EDPF)* (Chebrolu and Rao, 2006), was introduced not only to achieve bandwidth aggregation but also to ensure packets meet the playback deadline by scheduling packets based on their estimated delivery time.

Note that each proposal we described above has its pros and cons. For instance, a lower level modification leaves upper layers unaware of the aggregation and multiple connections; however, it lacks flow and QoS (application-based) information. A higher level modification does not change the protocol stack; however, there is no explicit mechanism to select a particular networking interface.

4 Discussion

Table 4 summarises the mobility techniques categorised in terms of layer mobility. In general, Mobile IP and its extensions can fully support the host mobility, but require a trade-off of the encapsulation header and permanent home address. Many techniques have been proposed to resolve the mobility problem at the transport layer; however, modifications of the TCP stacks are required, which leads to the deployability issue.

Table 4 Mobility summary

Layer mobility	*Examples*	*Pros*	*Cons*
Network mobility	Mobile IP and its extensions	Support TCP/application unaware of mobility.	Require an additional encapsulation header and a permanent home address.
Transport mobility	TCP migration, Freeze TCP, MSOCK, SCTP, and DCCP	Focus on the disconnection operation at the transport layer.	Need a modification of TCP states except freeze TCP.
Sub-layer mobility	HIP, i3, and secure i3	Support the secure identity using indirection layer.	Require a modification of protocol stacks.
Application mobility	SIP and SIP eye extension	Provide full host mobility support.	Require changing applications, e.g., all services run over SIP.

HIP, i3, and secure i3 introduced the new secure naming system which binds the host name to a secure ID, and then to locators. This concept can fully resolve the mobility issue; however, the disadvantage is to modify the protocol stack and the introduction of a

new naming system. For application mobility, SIP and its extensions support the host mobility, but this helps only the applications that use SIP signalling.

Tables 5 and 6 show the comparisons and summary of the pros and cons of each proposal. We summarise the features of each proposal and classify it in terms of mobility, multihoming, user location privacy, scalability, routing scalability, secure identity, and deployability. In brief, all proposals support full mobility except Shim6 and its derivative, e.g., six/one, due to the address duplication issue.

Multihoming is supported by all schemes. Little discussion regarding user location privacy exists within these proposals, except in LISP with the address rewriting technique. However, as we previously recommended, this issue can be resolved by introducing a proxy.

Three main mapping schemes: DNS, DHT, and a hierarchical DHT, are used with the scalability trade-off. Consider the ID structure: HIP, HRA, and (enhanced) MILSA introduce new naming spaces in a secure manner; however, lead to the deployability issue vs. the names derived from legacy IP addresses.

To summarise, in general these competing proposals can be categorised as follows:

1 they are either host-based or network-based or the combination (hybrid) approaches

2 they are either introducing a new naming space or deriving/using the legacy IP address naming space

3 they are either using the secure ID/flat label or the hierarchical ID like DNS.

Each side has its own pros and cons. Table 6 also shows pros and cons of each technique. In brief, all proposals can support mobility and multihoming except the original Mobile IP. Proposals which require a new naming space can be modified to support user location privacy with the help of a proxy.

Table 5 Proposal comparisons

Proposals	*Mobility*	*Multihoming*	*User location privacy*	*Scalability*	*Routing scalability*	*Security (ID)*
Mobile IP v6 + its extensions	Yes	Yes with multiple CoA registration	Partly (only outside home network)	Based on DNS	N/A	N/A
HIP	Yes	Yes	Yes	N/A	Require PA	Secure ID
Shim6	Partly	Yes	N/A	Based on DNS	N/A	N/A
LISP	Yes	Yes	Yes	Based on DNS	Require PA	N/A
Six/one	Partly	Yes	N/A	Based on DNS	Require PA	N/A
HRA	Yes	Yes	N/A	H-DHT	Require PA	Hierarchical secure ID
Enhanced MILSA	Yes	Yes	N/A	Modification of H-DHT	Require PA	Hierarchical secure ID

Table 6 Pros/cons summary

Proposals	*Infrastructure*	*Pros*	*Cons*
Mobile IP v6 + extensions	Host	Easy to deploy, No new naming space, No additional mapping systems.	Require a permanent home address. No user location privacy within home network. Identity is not secure. Host cannot select the path.
HIP	Host	Secure ID (public key); flat label.	Require new naming space and a new mapping system. Host cannot select the path.
LISP	Network	No host modification.	Host cannot select the path.
Shim6	Host	Host can choose its path (requires information about path). Does not introduce a new name space.	Duplication of IP address as an identity. Identity is not secure. No user location privacy.
Six/One	Hybrid (Host + Network)	Allow host and network to choose the path.	Duplication of IP address as an identity. Identity is not secure. No user location privacy.
HRA	Hybrid (Host + Network)	Hierarchical secure ID.	Require a new naming space and a new mapping system. Host cannot select the path.
Enhanced MILSA	Hybrid (Host + Network)	Hierarchical secure ID.	Require a new naming space and a new mapping system. Host cannot select path.

Consider scalability, a hierarchical system like DNS strongly supports the scalability vs. flat ID. Whether the secure ID or flat ID is better is still being debated. If a new naming space is introduced, the PA address is easier to deploy because the identity will be bound to the new naming system. With a derived IP address naming space, suppose the host name can be used as the node identity, the PA address can be also deployed as a hostname-identity system, and thus results in mitigating the routing scalability problem.

Consider routing scalability in particular, the question is how to distribute the IP reachability information in a compact manner? The ideal solution is to enforce all IP-based networks to use the PA addresses; however, that creates problems when the organisations change their service providers. As a result, the idea of the ID/locator split concept was introduced to allow organisations to use the identity, which may be a new naming space. Whether introducing a new naming scheme is feasible is still another question.

Consider mobility and multihoming, whether the mobile host should be changed or the network or both (hybrid) still remains an outstanding question. Many proposals have attempted to solve these issues. There are pros and cons on each side.

In FWNs, all IP-based applications running on mobile devices can use either UDP or TCP, or other future transport protocols. All options require supporting mobility, multihoming, etc. One consideration is the concept of evolution, not revolution, introduced by 3GPP. Therefore, the re-structuring of the networks needs to be made smoothly. SIP is a well-known application layer mobility concept. 3GPP also adopted

proxy mobile IP and dual stack mobile IP for interoperation with non-3GPP networks. The question remains as to whether this combination is good enough.

Consider the network layer. Mobile IPv6 (MIPv6) provides full features for mobility, and its extensions, such as HMIPv6 and Fast handover for IPv6, can mitigate the delay/latency problem. Similar to MIPv4 with the multihoming feature, MIPv6 can support multihoming by simply allowing the multiple CoA registration.

In addition, the multi-interface selection problem should be considered. Mobile users should be able to choose paths or egress and ingress exits, with the help of service providers, e.g., providing path characteristics. These require a modification of DNS.

To support guaranteed QoS globally, the cooperation amongst different service providers is also required, e.g., to forward the packets among corresponding service providers' agents. This may require a different type of IP-in-IP tunnelling and/or encapsulation technique. This encapsulation should last until the end of the session or until the new binding is updated at the domain name server and/or the home networks.

With Proxy Mobile IPv6, the proxy agent can provide the mobile IP feature on behalf of the mobile node. This can mitigate the deployability issue. Mobile IPv6 uses the conventional DNS hierarchical naming space so the scalability issue is mitigated, but MIPv6 identity or IP address lacks security as provided by some of the other ID/locator split schemes (secure IDs).

Consider the sub-layer. Several techniques have introduced the identity, or ID, concept separated from the locator. This obviously raises the concerns of deployability. The introduction of a new naming space and a new mapping scheme from the host name to identity, and then on to locators requires a modification of DNS and/or rendezvous servers.

Flat or the combination of hierarchical and flat portions of the node identity can lead to a trade-off of scalability vs. security. In contrast to Mobile IP and its extensions, user location privacy can be supported due to the identity concept if the address rewriting technique at the border router, or proxy agent, is deployed. Mobility and multihoming are supported inherently.

Consider the application layer. SIP is commonly used for application layer mobility over UDP. A SIP eye extension using an encapsulation concept can be used to support TCP-based applications. This encapsulation concept is similar to SIP over Mobile IPv6 adopted by 3GPP. To support multihoming, multiple IP registrations can be added to the SIP server.

There are also some proposals that combine several techniques we described; for example, Tschofenig et al. (2007), So et al. (2005) and Henderson (2004) proposed the SIP and HIP combinations to resolve both mobility and multihoming; however, since SIP was originally based on UDP, whether adding HIP is beneficial is a lingering question.

The main advantage of SIP and HIP is the support of a secure identity. However, whether this main advantage can overcome the trade-off of new naming spaces remains in questions. There are also some techniques such as Atkinson et al. (2008) applied the concept of dynamic DNS (Thomsone et al., 1997), that is, the identity introduced can be resolved from DNS, so whenever the mobile nodes move, they update the DNS record directly, again, assuming that the update is fast.

5 Conclusions

In FWNs, the network will be a convergence of diverse wired and wireless technologies, such as Ethernet, WLANs, WiMAX, 2G/3G/4G, and so on. The future applications, such as voice, video, TV broadcasting, online games, medical applications, etc., require varying levels of QoSs. In addition, many features are required to be supported in FWNs: mobility, multihoming, privacy, scalability, deployability, etc.

In FWNs, the network should be able to support billions of mobile networking devices with many networking interfaces. To communicate diversely, the IP is a potential networking protocol for use in FWNs. However, the current IP architecture faces many well-known problems, and those will prevent FWNs from achieving the features mentioned above. One of the difficulties of the IP address architecture is an overloading of the identity and the location functionality in IP addresses.

As a result, in this paper, we focus on the ID/locator split concept in the IP architecture. In general, we discuss the issues required for FWNs, and then we survey recent proposals based on mobility and multihoming. We categorised the mobility based on the layer perspective, such as NEMO, transport mobility, and application mobility.

For multihoming, we grouped the recently proposed techniques into host-based, network-based, and the hybrid approaches. We also briefly discussed the issue of multi-interface selection in a multihoming environment.

For routing scalability, the PA addresses need to be deployed; however, this introduces the re-homing issue. Therefore, the introduction of identity and location split was discussed. For deployability, there is always a debate over which side, i.e., host or network, should be modified, and whether the introduction of new mechanisms is backward compatible.

References

Adiseshu, H., Parlkar, G. and Varghese, G. (1996) 'A reliable and scalable striping protocol', *ACM SIGCOMM Computer Communication Review*, Vol. 26, No. 4, pp.131–141.

Adkins, D., Lakshminarayanan, K., Perrig, A. and Stoica, I. (2003) *Towards a More Functional and Secure Network Infrastructure*, University of California, Berkley Technical Report, UCB/CSD-03-124, available at http://i3.cs.berkeley.edu/publications/papers/csd-03-1242.pdf.

Ahlund, C. and Zaslavsky, A. (2003) 'Multihoming with MOBILE IP', *Lecture Notes in Computer Science*, Springer, Vol. 2720, pp. 235-243.

Atkinson, R., Bhatti, S. and Hailes, S. (2008) 'Mobility through naming: impact on DNS', *Proceedings of International Workshop on Mobility in the Evolving Internet Architecture*, pp.7–12.

Campbell, A.T., Gomez, J., Sanghyo, K., Chieh-Yih, W., Turanyi, Z.R. and Valko, A.G. (2002) 'Comparison of IP micromobility protocols', *IEEE Wireless Communication Magazine*, Vol. 9, No. 1, pp.72–82.

Chebrolu, K. and Rao, R.R. (2006) 'Bandwidth aggregation for real-time applications in heterogeneous wireless networks', *IEEE Transactions on Mobile Computing*, Vol. 5, No. 4, pp.388–403.

DARPA Internet Program (1981a) *Internet Protocol*, RFC 791.

DARPA Internet Program (1981b) *Transmission Control Protocol*, RFC 793.

Deering, S. and Hinden, R. (1998) *Internet Protocol Version 6* (*IPv6*) *Specification*, RFC 2460.

Devarapalli, V., Wakikawa, R., Petrescu, A. and Thubert, P. (2005) *Network Mobility* (*NEMO*) *Basic Support Protocol*, RFC 3693.

Eddy, W.M. (2004) 'At what layer does mobility belong?', *IEEE Communication Magazine*, Vol. 42, No. 10, pp.155–159.

Farinacci, D., Fuller, V., Oran, D. and Meyer, D. (2007) *Locator/ID Separation Protocol* (*LISP*), Internet-Draft, draft-farinacci-LISP-03.

Ford, A., Raiciu, C., Barre, S. and Iyengar, J. (2010) *Architectural Guidelines for Multipath TCP Development*, Internet-Draft, draft- draft-ietf-mptcp-architecture-00.txt.

Funato, D., Yasuda, K. and Tokuda, H. (1997) 'TCP-R: TCP mobility support for continuous operation', *Proceedings of International Conference on Network Protocols* (*ICNP 1997*), pp.229–236.

Gares-Erice, L., Biersack, E., Felber, P., Ross, K. and Urvoy-Keller, G. (2003) 'Hierarchical peer-to-peer systems', *Proceedings of ACM/IFIP International Conference Parallel and Distributed Computing*, pp.643–657.

Goff, T., Moronski, J. and Phatak, D.S. (2000) 'Freeze-TCP: A true end-to-end TCP enhancement mechanism for mobile environments', *Proceedings of IEEE International Conference on Computer Communication* (*INFOCOM 2000*), pp.1537–1545.

Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K. and Patil, B. (2008) *Proxy Mobile IPv6*, RFC 5213.

Gustafsson, E. and Jonsson, A. (2003) 'Always best connected', *IEEE Wireless Communication*, Vol. 10, No. 1, pp.49–55.

Haas, Z.J. (1997) 'Mobile-TCP: an asymmetric transport protocol design for mobile systems', *Proceedings of IEEE International Conference on Communication* (*ICC 1997*), pp.1054–1058.

Henderson, T. (2004) *Can SIP use HIP?*, HIP workshop.

Hsieh, H. and Sivakumar, R. (2002) 'A transport layer approach for achieving aggregate bandwidths on multi-homed mobile hosts', *Proceedings of International Conference on Mobile Computing and Networking*, pp.83–94.

IEEE Std. 802.21-2008 (2009) *IEEE Standard for Local and Metropolitan Area Networks Part 21: Media Independent Handover Services*, IEEE.

ITU-T Y.2015 (2007) *General Requirements for ID/Locator Separation in NGN*, Y.2015, 18pp.

Jain, R. (2006) 'Internet 3.0: ten problems with current internet architecture and solutions for the next generation', *Proceedings of IEEE Military Communication Conference* (*MILCOM 2006*), pp.1–9.

Jeffrey, G., Andrews, J., Arunabha-Ghosh, A. and Muhamed, R. (2007) *Fundamentals of WiMAX Understanding Broadband Wireless Networking*, Prentice Hall, 496pp.

Johonson, D., Perkins, C. and Arkko, J. (2004) *Mobility Support in IPv6*, RFC 3775.

Kohler, E., Handley, M. and Floyd, S. (2006) 'Designing DCCP: congestion control without reliability', *Proceedings of ACM SIGCOMM*, pp.27–38.

Koodli, R. (2005) *Fast Handovers for Mobile IPv6*, RFC 4068.

Koponen, T., Gurtov, A. and Nikander, P. (2005) 'Application mobility with HIP', *Proceedings of NDSS Wireless and Mobile Security Workshop*.

Landfeldt, B., Larsson, T., Ismailov, Y. and Seneviratne, A. (1999) 'SLM, a framework for session layer mobility management', *Proceedings of International Conference on Computer Communication and Networks* (*ICCCN 1999*), pp.452–456.

Launois, C.D. and Bagnulo, M. (2006) 'The paths toward IPv6 multihoming', *IEEE Communication Surveys & Tutorials*, Vol. 8, No. 2, pp.38–51.

Le, D., Fu, X. and Hogrefe, D. (2006) 'A review of mobility support paradigms for the internet', *IEEE Communication Surveys & Tutorials*, Vol. 8, No. 1, pp. 38-51.

Liu, J. and Singh, S. (2001) 'ATCP: TCP for mobile ad hoc networks', *IEEE Journal on Selected Areas in Communication*, Vol. 19, No. 7, pp.1300–1315.

Maltz, D.A. and Bhagwat, P. (1998) 'MSOCKS: an architecture for transport layer mobility', *Proceedings of IEEE International Conference on Computer Communication* (*INFOCOM 1998*), pp.1037–1045.

Meyer, D. (2008) 'The locator identifier separation protocol (LISP)', *Cisco Systems: The Internet Protocol Journal*, Vol. 11, No. 1.

Meyer, D., Zhang, L. and Fall, K. (2007) *Report from the IAB Workshop on Routing and Addressing*, RFC 4984.

Moskowitz, R., Nikander, P. and Jokela, P. (2006) *Host Identity Protocol* (*HIP*) *Architecture*, RFC 4423.

Nikander, P., Arkko, J. and Ohlman, B. (2004) 'Host Identity Indirection Infrastructure (Hi3)', *Proceedings of the 2nd Swedish National Computer Networking Workshop*.

Nordmark, E. and Bagnulo, M. (2007) *Shim6: Level 3 Multihoming Shim Protocol for IPv6*, Internet-Draft, draft-ietf-shim6-proto-09.

Pan, J., Paul, S., Jain, R. and Bowman, M. (2008) 'MILSA: a mobility and multihoming supporting identifier locator split architecture for naming in the next generation internet', *Proceedings of IEEE Global Communication Conference* (*GLOBECOM 2008*), pp.1–6.

Pan, J., Paul, S., Jain, R., Bowman, M. and Chen, S. (2009) 'Enhanced MILSA architecture for naming, addressing, routing and security issues in the next generation internet', *Proceedings of IEEE Global Communication Conference* (*GLOBECOM 2009*), pp.1–6.

Paul, S., Pan, J. and Jain, R. (2009) 'A survey of naming systems: classification and analysis of the current schemes using a new naming reference model', WUSTL Technical Report, Available at http://www.cse.wustl.edu/~jain/papers/naming.htm.

Perkins, C. (2002) *IP Mobility Support for IPv4*, RFC 3220.

Raiciu, C., Wischik, D. and Handley, M. (2009) 'Practical congestion control for multipath transport protocols', University Collage of London Technical Report, available at http://nrg.cs.ucl.ac.uk/mptcp/mptcp-techreport.pdf.

Raychaudhuri, D. and Gerla, M. (2005) *New Architectures and Disruptive Technologies for the Future Internet: the Wireless, Mobile and Sensor Network Perspective*, NSF Wireless Mobile Planning Group Workshop, 49pp.

Rosenberg, J., Scheulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and Schooler, E. (2002) *SIP: Session Initiation Protocol*, RFC 3261.

Schulzrinne, H. and Eddy, W.M. (2000) 'Application-layer mobility using SIP', *Proceedings of IEEE Service Portability and Virtual Customer Environments*, pp.29–36.

Snoeren, A.C. and Balakrishnan, H. (2000) *TCP Connection Migration*, Internet-Draft, draft-snoeren-tcp-migrate-00.txt.

So, J., Wang, J. and Jones, D. (2005) 'SHIP mobility management hybrid SIP-HIP scheme', *Proceedings of the 6th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pp.226–230.

So-In, C., Jain, R. and Dommety, G. (2009) 'PETS: persistent TCP using Simple Freeze', P*roceedings of the 1st International Conference on Future Information Networks* (*ICFIN 2009*), pp.97–102.

Soliman, H. (2007) *Mobile IPv6 Support for Dual Stack Hosts and Routers (DSMIPv6)*, Internet-Draft, draft-ietf-mip6-nemo-v4traversal-05.txt.

Soliman, H., Castelluccia, C., Elmalki, K. and Bellier, L. (2008) *Hierarchical Mobile IPv6* (*HMIPv6*) *Mobility Management*, RFC 5380.

Soliman, H., Tsirtsis, G., Montavont, N., Giaretta, G. and Kuladinithi, K. (2009) *Flow Bindings in Mobile IPv6 and NEMO Basic Support*, Internet-Draft, draft-ietf-mext-flow-binding-03.txt.

Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L. and Paxson, V. (2000) *Stream Control Transmission Protocol*, RFC 2960.

Stoica, I., Adkins, D., Zhuang, S., Shenker, S. and Surana, S. (2004) 'Internet Indirection Infrastructure', *IEEE/ACM Transactions on Networking*, Vol. 12, No. 2, pp. 205-218.

Sultan, F., Srinivasan, K., Iyer, D. and Iftode, L. (2002) 'Migratory TCP: highly available internet services using connection migration', *Proceedings of IEEE International Conference Distributed Computing System* (*ICDCS 2002*), pp. 17-26.

The 3rd Generation Partnership Project (2007a) *The 3GPP Technical Specification Group Service and System Aspects; General Packet Radio Service* (*GPRS*) *enhancements for Evolved Universal Terrestrial Radio Access Network* (*E-UTRAN*) *Access*, 3GGPP TS 23.401 V8.0.0, 167pp.

The 3rd Generation Partnership Project (2007b) *The 3GPP Technical Specification Group Service and System Aspects; Architecture enhancements for non-3GPP accesses*, 3GPP TS 23.402 V8.0.0, 131pp.

Thomsone, S., Rekhter, Y. and Bound, J. (1997) *Dynamic Updates in the Domain Name System*, RFC 2136.

Tschofenig, H., Ott, J., Schulzrinne, H., Henderson, T. and Camarillo, G. (2007) *Interaction between SIP and HIP*, Internet-Draft, draft-tschofenig-hiprg-host-identities-05.txt.

Vakil, F., Dutta, A., Chen, J-C., Tauil, M., Baba, S., Nakajima, N., Shobatake, Y. and Schulzrinne, H. (2001) *Supporting Mobility for TCP with SIP*, Internet-Draft, draft-itsumo-sip-mobility-tcp-00.txt.

Vogt, C. (2007) *Six/One: A Solution for Routing and Addressing in IPv6*, Internet-Draft, draft-vogt-rrg-six-one-01.txt.

Wakikawa, R., Devarapalli, V., Tsirtsis, G., Ernst, T. and Nagami, K. (2009) *Multiple Care-of Addresses Registration*, Internet-Draft, draft-ietf-monami6-multiplecoa-14.txt.

Xu, X. and Guo, D. (2008) 'Hierarchical routing architecture', *Proceedings of the 4th Euro-NGI Conference on Next Generation Internetworks*, pp.92–99.

Zhang, M., Lai, J., Krishnamurthy, A., Peterson, L. and Wang, R. (2004) 'A transport layer approach for improving end-to-end performance and robustness using redundant paths', *Proceedings of the USENIX 2004 Annual Technical Conference*, pp.8–18.