Project Number:	**IST-2001-32603**
Project Title:	**6NET**
CEC Deliverable Number:	**32603/ULANC/DS/4.5.1/A1**
Contractual Date of Delivery to the CEC:	August 31st 2002
Actual Date of Delivery to the CEC:	October 31st 2002
Title of Deliverable:	Report on IETF Multihoming Solutions
Work package contributing to Deliverable:	WP4
Type of Deliverable*:	R
Deliverable Security Class**:	PU
Editors:	Martin Dunmore, Christopher Edwards
Contributors:	Njål T. Borch, Tim Chown, Martin Dunmore, Oliver Krämer, Pekka Savola

* Type: P - Prototype, R - Report, D - Demonstrator, O - Other

** Security Class: PU- Public, PP – Restricted to other programme participants (including the Commission), RE – Restricted to a group defined by the consortium (including the Commission), CO – Confidential, only for members of the consortium (including the Commission)

Abstract:

This document is an intermediate report on IPv6 multihoming solutions that have been proposed within the IETF. Although not within the IETF framework, some multihoming considerations within the BRAIN and MIND projects are also briefly described.

Keywords:

Multihoming, IPv6, Host, Site, Mobile IPv6, BGP, ISP.

Executive Summary

Multihoming is widely used in the IPv4 Internet today and is an essential service component for enterprises. With the introduction of IPv6 and the anticipated continued growth of the Internet, it is likely that multihoming will become increasingly common in the Internet.

However, the deployment of multihomed hosts and networks is not a straightforward task. Provider based addressing schemes (as used in CIDR and IPv6) mean that multihomed prefixes must currently be injected into the Internet's core routing tables in the 'default-free zone' (DFZ) for those networks to be reachable in all conditions. In addition, the concept of hosts changing their IP address during the lifetime of a connection is incompatible with the most popular transport protocols.

In recent years, there have been many proposals within the IETF to overcome the problems that are experienced when a site and/or host is multihomed. These proposals range from new transport protocols and mobility mechanisms to new routing techniques and geographic addressing schemes.

Yet, so far, no single multihoming solution will cater for all the requirements needed for both site and host based multihoming. It is most likely that several mechanisms will need to be used together to satisfy all the various requirements and parties concerned.

Table of Contents

1 Introduction 4
2 Overview of Multihoming 4
3 Host Multihoming Solutions 6
3.1 Host-Centric Multihoming 6
3.2 Mobile IPv6 for Multiple Interfaces 8
3.3 Per-flow Movement in MIPv6 8
3.4 BRAIN/MIND Multihoming Considerations 8
3.5 LIN6 9
3.6 Transport Solutions 10
3.6.1 Stream Control Transmission Protocol (SCTP). 10
3.6.2 Preserving active TCP sessions 11
3.6.3 End to End Multihoming 11
4 Site Multihoming Solutions 12
4.1 IPv6 Multihoming with Route Aggregation 12
4.2 Scalable Support for Multi-homed Multi-provider Connectivity 13
4.3 IPv6 Multihoming Support at Site Exit Routers 16
4.4 Multihomed Routing Domain Issues 16
4.4.1 Mutual Backup 16
4.4.2 Router Renumbering 17
4.5 Routing Support for IPv6 Multihoming 18
4.6 Routing Header Usage 19
4.7 MHAP 19
4.8 “GEO For Now” 20
5 Conclusions 21
References 22
Abbreviations 24

1 Introduction

Multihoming is widely used in the IPv4 Internet today and is an essential component of service for enterprises. In recognition of this, the built-in features of IPv6 make it easier for end-hosts and networks to be multihomed than in IPv4. This, combined with the expected continued growth of the Internet, means that it is likely that multihoming will become an increasingly common phenomenon.

Unfortunately, the deployment of multihomed hosts and networks is not a straightforward task. The IPv6 aggregatable global unicast address format [14] is based on the aggregation of provider address blocks rather than geographical ones. This means that, currently, multihomed prefixes need to be injected into the routing tables of the Internet's default-free zone (DFZ). This has resulted in a substantial increase in the size of BGP tables in the core of the Internet [34]. In addition, the possibility of changing destination and/or source IP addresses during the lifetime of a connection, is incompatible with the most widely used transport protocols like TCP and UDP.

Recently, there have been numerous proposals made within the IETF to allow IPv6 multihoming to prosper without incurring the associated scalability and transport problems. The following section gives a brief overview of what multihoming is, the motivations for being multihomed and the differences between host and site multihoming. Sections 3 and 4 then discuss some of the proposed solutions for host and site multihoming respectively.

2 Overview of Multihoming

There are typically two types of multihoming: host multihoming and site multihoming[1]. A multihomed host is a host with more than one global IP address assigned to it. These addresses could be from the same Internet Service Provider (ISP) or from different ISPs. Furthermore, a host may be multihomed with a single interface (multiple IP addresses on one interface) or may have one or more global IP addresses (from different subnets) on several interfaces.

Site multihoming refers to a site that has more than one connection to the public Internet through either the same or different Internet Service Providers (ISPs)

There are several reasons why a site or an end host may wish to be multihomed:

- *Fault resilience/redundancy*. If the link to one ISP fails, a different link can be used to carry the traffic.
- *Load Balancing*. To achieve higher throughput a host/site could spread its traffic load over two ISPs.
- *Service Value/Policy*. With some particular services (e.g. VoIP or VoD) certain ISPs may offer a cheaper price than others. Thus, a site/host may wish to allocate traffic types to different ISPs to save on traffic costs.

IPv6 is more suited than IPv4 for deploying multihomed hosts since it is more able to dynamically detect and inherit multiple addresses. When an IPv6 host connects to a network it receives router advertisement messages from all IPv6 routers attached to the LAN. Using stateless address autoconfiguration [15], an IPv6 host can automatically assign a globally unique IPv6 address for each different network prefix it receives through router advertisements. These advertisements may be received on one interface (thus having multiple addresses on single interface) or on several interfaces (one or more addresses per interface). Thus, an IPv6 host can automatically configure itself as a multihomed host when it boots up.

[1] There is also a third type of multihoming: ISP/operator multihoming, although this is a trivial case since address prefixes are large enough so that there are no restrictions on route advertisements in the DFZ.

To further understand how to implement site multihoming in IPv6, an IETF working group, 'multi6' [10], was established. However, activity in this group has been low with only one IPv6 multihoming requirements draft [32] and a draft on practices in IPv4 multihoming [33] having been submitted. A different multihoming group, called 'ipv6mh' [11], has recently been established. It is not affiliated to the IETF and considers all IPv6 multihoming issues within its charter (rather than just site multihoming).

3 Host Multihoming Solutions

Multihoming is commonly considered under the site multihoming aspect. In this scenario, an organisation's network is connected to the Internet via more than one Internet Service Provider (ISP), for example to provide resilience against link failures at an ISP, if Internet connectivity is critical, or to employ load balancing. End nodes are only affected in IPv6 if both address prefixes are advertised to the network, which means that more than one IPv6 address is associated with their (single) network interface.

As PCs (especially notebooks or PDAs) become more commonly equipped with multiple network interfaces, like Ethernet, WaveLAN, Bluetooth and UMTS, multihoming aspects gain importance from the terminal and end user viewpoint. For example, the user probably would like to control which interface is used, if both interfaces of his notebook are connected at the same time. This becomes especially important if there are differences in pricing between the different access methods.

The following sections will give a short overview on existing terminal-focused multihoming approaches. Section 3.1 focuses on host related issues of site multihoming, which occur independent of the fact whether the host is equipped with a single or multiple interfaces. For example, issues like source address selection are also relevant for hosts with multiple interfaces, if the interfaces are connected via different ISPs and have different site prefixes. A possible scenario for this case is for example a home user having his laptop connected to one ISP via a cable modem and to another ISP via a GPRS or UMTS phone.

3.1 Host-Centric Multihoming

The "Host-Centric IPv6 Multihoming" Internet Draft [1] discusses problems and possible solutions of hosts receiving site prefixes in a multihomed network. If a site is connected via two ISPs, their border routers will advertise both prefixes to the site.

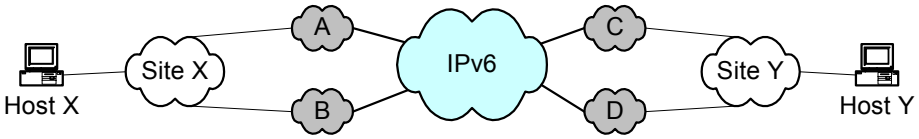

Figure 1 Site multihoming setup

Figure 1 shows an example multihoming scenario. Site X is connected via ISPs A and B to the IPv6 Internet, and site Y via ISPs C and D. Host X receives prefix advertisements from ISPs A and B, and generates the corresponding IPv6 addresses (say A:X and B:X) with A and B corresponding to the site prefixes and X being the host identifier. The same holds true for host Y and ISPs C and D: the addresses generated are C:Y and D:Y respectively.

If host X tries to communicate with host Y, it selects a destination address from the 2 available global unicast addresses of host Y (assuming that both are made available via DNS). Subsequently, host X selects a source address from the addresses associated with the outgoing interface. However, the selection of the source address is not as simple an issue as seems at first glance. Since most site exit routers now perform ingress filtering [37] as a way of preventing security attacks using spoofed source addresses, if host X selects the wrong source address its packets will be dropped by the site exit router. For example, if the address based on prefix A (A:X) is used as the source address, the interior routing of site X (which routes on the destination address) may route the packet through site X's exit router connected to ISP B. If the router connected to ISP B performs ingress filtering, packets with a source address based on prefix A will likely be dropped since

they do not match the expected prefix, which would be ISP B's prefix. This problem is known as the 'site exit issue'.

The following alternatives are proposed and evaluate to solve the site exit issue:

1. *Relaxing source address checks*

   Either the ISP could switch off the source address checks altogether, or the site exit routers may be configured in such a way that they accept all prefixes that are valid in the site, i.e. the exit router for ISP B accepts A's prefixes. The former approach requires a high level of trust between the ISP and the site, but has the advantage of not requiring any changes in the hosts or routers. If a list of authorised prefixes is used, the site exit routers need to be configured somehow to be aware of these address prefixes.

2. *Source address dependent routing*

   Packets originating in site X are passed through a source routing domain, to which all site exit routers are connected. These routers would choose a route based on the destination and source addresses of a packet, selecting the appropriate exit router for the given source address. Alternatively, the site exit routers could be connected via a mesh of tunnels. If the source address check on a packet reveals that the address prefix is not a valid one for the ISP the exit router is connected to, it forwards the packet through a tunnel to a more appropriate site exit router[1].

   As this approach would require significant changes in the router implementations, it would be only a long-term solution.

3. *Source address selection by the host*

   This possible solution relies on the host to determine the appropriate source address that is compatible with the site exit chosen by the routing protocol. Alternatively, the host could tunnel the packet to the correct site exit router. The former approach would require a form of "source address discovery", the latter an "exit router discovery" (for a given source address).

   Both approaches require significant changes in the IPv6 implementation in hosts, and additionally minor changes in routers (for generating the new ICMP packets for the "source address discovery" or "exit router discovery").

4. *Packet rewriting at the exit router*

   A site exit router receiving a packet with a "wrong" source address could take action to replace the address instead of dropping the packet, provided it recognises the address as a valid address from its site. If it would do so on any packet, it would prevent all ingress filtering. The packet rewriting could either mean adding a home address option, with the original source address in the option, and a newly generated "correct" address as the source address of the packet, or the site exit router could use IPv6-in-IPv6 encapsulation with the original destination address and a new conforming source address in the encapsulating header. Simply replacing the source address in the IPv6 header (without adding a home address option) is mentioned too, but is probably not a good solution due to problems with IPSec, for example. Besides, the router would need to know which alternative source address could be used for the host originating the packet.

Option 1 requires no changes to existing systems (if ingress filtering is switched off completely), or only minimal configuration at the site exit routers. All other options require more or less significant changes to existing host and/or router implementations. As a long-term solution, option 2 is favoured by the authors of [1].

[1] It is not detailed how this "more appropriate" router should be determined.

3.2 Mobile IPv6 for Multiple Interfaces

The "MIPv6 for Multiple Interfaces" (MMI) Internet Draft [2] considers MIPv6 [3] extensions required to accommodate multihomed mobile nodes, i.e. mobile devices equipped with multiple air interfaces, for example an 802.11 and a Bluetooth interface.

The basic idea consists of using more than one home address per mobile node (MN), one for each interface. If a flow is to be moved from its original interface 'I1' to the new interface 'I2' (which might only have become available after the connection was established), the MN sends a binding update with the address of I2 as the care-of address (CoA) and the address of I1 as home address. This results in all traffic sent to I1 being routed via I2.

Unlike the per-flow movement approach described in the next section, MMI only allows the movement of all flows to another interface. However, MMI enables the MN to react to connection problems by moving traffic to the second interface if one of the ISPs (assuming that the interfaces are receiving prefixes from different IPs) goes down. If the home address option is used on the outgoing traffic, ingress filtering in the site exit routers is not an issue as the CoA is used as the source address.

3.3 Per-flow Movement in MIPv6

The "Per-flow Movement" Internet Draft [4] specifies an extension to MIPv6 [3], which enables the incoming traffic of individual flows to be moved to another interface.

A flow can be uniquely identified by the 5-tuple (source address, source port, destination address, destination port, protocol type). When sending a binding update, a new sub-option can be included containing the necessary information to identify a flow. The correspondent node then stores a binding for this flow and sends packets belonging to this flow to the new CoA.

Alternate to the 5-tuple identification of flows, a version based on the flow label and source address can be used.

This approach provides a greater flexibility than MMI described in the previous section, since it supports interface selection on a per-flow basis. However, influencing interface selection for outgoing traffic is not considered, which prohibits redundancy for the outgoing links. Additional mechanisms, such as those described in section 3.1, would be required to achieve this redundancy and load balancing for outgoing traffic.

3.4 BRAIN/MIND Multihoming Considerations

Within the IST framework BRAIN (http://www.ist-brain.org) and its successor MIND (http://www.ist-mind.org), multihoming issues of mobile hosts in access networks have been studied. In [5], the Enhanced Socket Interface (ESI) is introduced, which extends the standard Unix socket interface with QoS capabilities and local management information.

In MIND, multihoming of mobile nodes can occur in two main scenarios:

- A node has one air interface, but is connected to the access network via an "ad-hoc fringe" consisting of other mobile nodes extending the range of the access network. One of these might be multihomed to 2 access networks, thus the mobile node can attach to 2 different access networks.
- The mobile node has 2 air interfaces, being attached to the same or different access networks. It is also possible that the node is attached to an access network directly via one interface and attached via an ad-hoc fringe to another access network.

Especially in the second case, user configuration is desirable as different costs may be associated which each interface. Thus, additional support in the ESI for multihomed mobile terminals will include user/application configurable interface selection for traffic flows. This will allow the user to specify preferences on which kind of traffic should be routed via which interface, if more than one interface is available and active.

Traffic classes can be identified either on a per-flow basis, where flows are identified by the 5-tuple (source address, source port, destination address, destination port, protocol type), or with larger granularity on a traffic class basis, which is based on the destination port and protocol type. The latter approach allows for the possibility to route traffic through a specific interface according to traffic type. For example, voice traffic could be routed through the interface offering the best QoS, while other, less demanding, traffic could be routed through a less capable interface. It has also the advantage of being easier to configure, as the user does not need to configure each flow.

3.5 LIN6

LIN6 [6], [7] is a protocol to support host mobility and multihoming in IPv6. LIN6 introduces two new types of IPv6 addresses, the *LIN6 generalized ID* and the *LIN6 address.*

The LIN6 generalized ID is 128 bits and consists of a 64-bit LIN6 prefix (3ffe:501:1830:1999) and a 64-bit LIN6-ID. The LIN6-ID is the global unique identifier of a mobile node and uses the EUI-64 format. The LIN6 generalized ID is only used in the transport and upper layers and is assigned on a per-node basis. This address will not change even when a mobile node moves between networks.

The LIN6 address is 128 bits and consists of a 64-bit network prefix and a 64-bit LIN6-ID (the same LIN6-ID as for the LIN6 generalized ID). The 64-bit network prefix is the actual network prefix that a mobile node is currently located at. The LIN6 address is used in the network and appears in the IPv6 header. It is not passed to the transport layer. This address is assigned on a per-interface basis and will change value when a mobile node changes to another network (but only the network prefix will change).

Within the LIN6 protocol, the LIN6 generalized ID is used by the transport and application layers to designate the identity of a mobile node since this address remains constant. The LIN6 address is used by the network layer to designate both the location and the identity of a mobile node as it moves between networks. The correlation between the LIN6 generalized ID and the LIN6 address is held by one or more mapping agents. For more details on the LIN6 protocol, the reader should refer to [7].

Figure 2 illustrates multihoming support using LIN6. Host B is a multihomed host, having 2 network interfaces. Hosts A and B have the LIN6 generalized IDs ‘3ffe:501:1830:1999:LIN6_ID_A’ and ‘3ffe:501:1830:1999:LIN6_ID_B’ respectively. Host A has the LIN6 address ‘Prefix_A:LIN6_ID_A’ and host B has two LIN6 addresses: ‘Prefix_B1:LIN6_ID_B’ and ‘Prefix_B2:LIN6_ID_B’.

If host A wishes to establish a TCP connection to host B it uses the LIN6 generalized IDs such that the TCP association is between 3ffe:501:1830:1999:LIN6_ID_A and 3ffe:501:1830:1999:LIN6_ID_B. In LIN6, the mapping agents will resolve these addresses to the current location of the nodes (i.e. their LIN6 addresses) for transmission at the network layer.

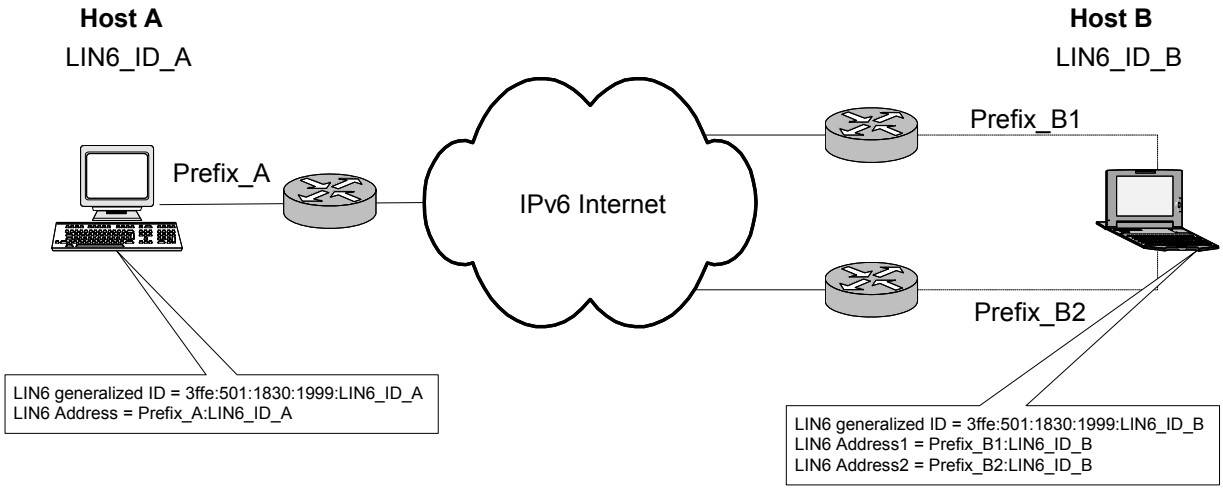

Figure 2 Multihoming in LIN6

Assume that host A learns both 'Prefix_B1' and 'Prefix_B2' as the network prefixes of host B and chooses 'Prefix_B1'. The src/dest addresses in the IPv6 header will now be Prefix_A:LIN6_ID_A and Prefix_B1:LIN6_ID_B respectively. Supposing host B's connection to network prefix 'Prefix_B1' breaks. Host A will receive ICMP unreachable errors and can then decide to use 'Prefix_B2' as the destination address. However, the original TCP connection is not broken because the TCP association is between the LIN6 generalized IDs and not the LIN6 addresses.

One can imagine how this multihoming solution could be extended so that host A could use both host B's prefixes simultaneously for different flows.

3.6 Transport Solutions

Although there are suitable host multihoming solutions at the IP layer, changing an IPv6 address during the lifetime of a connection has implications for transport layer protocols such as UDP and TCP. Quite simply, the transport layer will not recognise a received packet with a different source/destination address as belonging to the same transport connection before the address was changed. Therefore, either the IPv6 address change must be hidden from the transport layer, or support for address changes must be added to the typical transport protocols.

3.6.1 Stream Control Transmission Protocol (SCTP).

The Stream Control Transmission Protocol (SCTP) [8], is a reliable transport protocol operating on top of IPv4/IPv6 that provides network-level fault tolerance by supporting host multihoming at either end of the connection.

If a client host is multihomed it informs the corresponding server host about all of its IPv6 addresses in the INIT chunk's address parameters. The corresponding server host will then list all of its IPv6 address in the INIT-ACK chunk. An instance of SCTP regards each IPv6 address of its peer as a 'transmission path' towards it. Each SCTP instance chooses one of these addresses as the primary transmission path, upon which data exchange will normally occur.

Each end of the SCTP connection monitors all the transmission paths to its peer by sending HEARTBEAT chunks on every path that is not being used to exchange data chunks. The peer SCTP hosts acknowledge each HEARTBEAT chunk with a HEARTBEAT-ACK chunk.

Each transmission path is assigned a state: either active or inactive. A path is active if it has recently been used and has received the corresponding ACK for the SCTP datagram that was transmitted on it. If these

ACKS repeatedly fail, the path is marked as inactive and a notification is sent to the application layer of the host.

If the primary path becomes inactive, the sending host may automatically choose a new primary path or the user may instruct the local SCTP instance to use a new primary path.

3.6.2 Preserving active TCP sessions.

A way of allowing hosts to change their IP addresses during the lifetime of a TCP connection is presented in "Preserving Active TCP Sessions on Multihomed IPv6 Networks" [36]. This solution recognises that each participant in a TCP session may potentially have several global IPv6 addresses. Furthermore, this set of valid addresses remains relatively stable over a period of time. Therefore, it seems logical that, prior to the establishment of a TCP connection, the valid set of IPv6 addresses that could be used during the lifetime of the connection, should be exchanged by the participants. Thus, assuming that the *set* of addresses valid for a host does not change during the TCP connection, the connection should be able to accept any IPv6 address change by either participating host, provided that address appears in the agreed set..

The specific proposal in [36] is to include a 'PREFIXES' IPv6 option with the SYN and SYN/ACK packets during the 3-way handshake of a TCP connection establishment. Each participant acknowledges the PREFIXES with a PREFIXES_ACK containing the prefixes that it received.

3.6.3 End to End Multihoming

The notion of end to end multihoming as described in [9] is supported by transport (TCP or UDP) or application layers. Since multihoming support is implemented in the end systems no routing protocol changes are needed in the network. Instead end to end multihoming requires modifications to APIs and applications on the end systems.

End to end multihoming is a simple approach whereby multiple addresses are assigned to an interface and the choice of addresses to use is delegated to the application or transport layer. The authors of [9] propose to change TCP so that applications can pass multiple addresses to the transport layer and that all possible addresses are transmitted to the TCP peer via a TCP option. It is proposed to use DNS lookups for applications to learn of all the possible addresses for a destination. It is not clear how a host learns of all the addresses it may use as a source address (perhaps by router advertisements) nor how to best select a source address based on the destination (or vice versa) Changes must also be made so that the transport layer or application can detect and react to a loss of connection. Although details of how this may be achieved are not presented in [9].

4 Site Multihoming Solutions

The multi6 WG charter defines a multihomed site thus:

A multihomed site is a site that has more than one connection to the public internet with those connections through either the same or different ISPs. Sites choose to multihome for several reasons, especially to improve fault tolerance, perform load balancing, etc.

Note that some of the solutions discussed in this section can arguably be labelled as being host based multihoming solutions. This is certainly true for techniques that impact upon source address selection within the hosts. However, since they are all presented with the pretext that the hosts are multihomed because their site is multihomed they are included here.

4.1 IPv6 Multihoming with Route Aggregation

The Internet Draft "IPv6 Multihoming with Route Aggregation" [24], describes a routing scheme that supports IPv6 site multihoming. A site that is multihomed to two or more ISPs will designate one of its ISPs as its primary ISP and receive IPv6 address assignment from the primary ISP's address block. In the example in Figure 3, ISP A is chosen as the primary ISP for the multihomed site and assigns prefix A from its address block to the multihomed site.

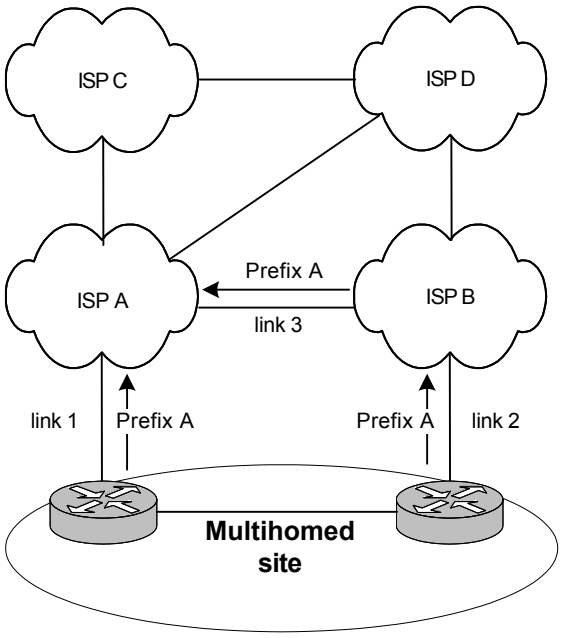

Figure 3 IPv6 Multihoming with Route Aggregation

The multihomed site will advertise prefix A to all of its connected ISPs, in this case ISP A and ISP B. ISP B will propagate the advertisement of prefix A to ISP A, but to no other ISP. ISP A will advertise its own aggregation into the Internet. Thus, traffic destined for prefix A will be routed to ISP A. ISP A will then forward the traffic either directly to the site (link 1) or to ISP B (link 3) according to the specific routing policy implemented. Load balancing may be achieved by having ISP A forward traffic for prefix A on both

link 1 and link 3 (differentiated by, for example, traffic type). Similarly, traffic originating from the multihomed site may be sent on either link 1 or link 2 according to the site's requirements.

If the site chooses to multihome for redundancy purposes, all of its traffic will be routed via its primary ISP, ISP A, should the connection to ISP B (link 2) fail. Conversely, should the connection to ISP A (link 1) be broken, all of the site's traffic will be routed via ISP B. This can be achieved because ISP A will know from the route advertisements it receives from ISP B that traffic for prefix A is reachable via ISP B.

Note that with this scheme, the specific routing prefix associated with the multihomed site is only known by the directly connected ISPs (A and B in the example) and is never propagated to the rest of the Internet. Thus, it has good scaling properties for the global Internet since the DFZ will not contain multihomed site prefixes. However, this scheme does rely on the primary ISP having links to the secondary ISP(s). If this is not the case, the site prefix would need to be advertised along the path between the primary and secondary ISPs, which could conceivably include the DFZ of the Internet.

A further drawback with this scheme is that the primary ISP is a single point of failure. Since the secondary ISPs do not inject the specific prefix of the multihomed site into the DFZ, the site will become unreachable should its primary ISP fail (that is, the primary ISP itself, rather than the link to the primary ISP). Also, the scheme does not cater for the multihomed site having multiple address prefixes assigned by different ISPs.

4.2 Scalable Support for Multi-homed Multi-provider Connectivity

RFC 2260 [30], describes several strategies for achieving scalable site multihoming. Although the document is not specific to IPv6, the strategies it describes are applicable to IPv6 as well as IPv4. The multihoming strategies in RFC 2260 are aimed at removing (or minimising) the need for routes to be held in the DFZ of the Internet and to minimise the amount of coordination needed between ISPs.

The fundamentals of the RFC 2260 method are illustrated in Figure 4 and Figure 5. A multihomed site is connected to ISPs A and B via border routers A and B respectively. Under normal conditions, i.e. the connections to both ISPs are active, each border router advertises, to its connected ISP, the prefix that was allocated by that ISP. Thus, in Figure 4, border router A advertises prefix A to ISP A's border router. Similarly, border router B advertises prefix B to ISP B's border router. Since the prefixes advertised by each of the site's border routers are aggregated by the respective ISP, no additional routes are injected into the default-free zone of the Internet. However, if the link to one ISP fails, the site will not be reachable on that respective prefix.

To get around this problem, RFC 2260 proposes that when a site's border router detects link failure between one or more of the site's other border routers and their respective ISPs, it should advertise the prefixes allocated by these ISPs. As an example, in Figure 5, when the link between the site's border router B and ISP B fails, border router A begins to advertise reachability of prefix B to ISP A. One way to achieve this would be to have an iBGP peering between the site's border routers. Thus, the absence of prefix B being advertised by border router B (due to the link failure) can trigger border router A to advertise prefix B to ISP A.

In order for traffic destined to prefix B to be able to reach its destination, ISP A would need to inject this prefix into the DFZ of the Internet. Although injecting site prefixes into the DFZ is undesirable, RFC 2260 argues that this method (known as *auto-route injection*), only results in site prefixes being injected into the DFZ temporarily whilst a connection between the multihomed site and one of its ISPs is down. Thus, the expected average number of site prefixes injected into the DFZ at any given time would be a small percentage of the total number of multihomed prefixes in the entire Internet.

This is true for sites choosing to multihome for reasons of redundancy. In the examples given, there is no reason for prefix A to be reachable via ISP B (or prefix B to be reachable via ISP A) unless there is a link failure. However, if the site chooses to multihome for other reasons (e.g. for traffic engineering purposes) there may well be a requirement for prefixes A and B to be reachable via both ISPs. In this case, both

prefixes would need to be injected into the DFZ to be reachable via the non-aggregating ISP (i.e. the ISP not owning the prefix).

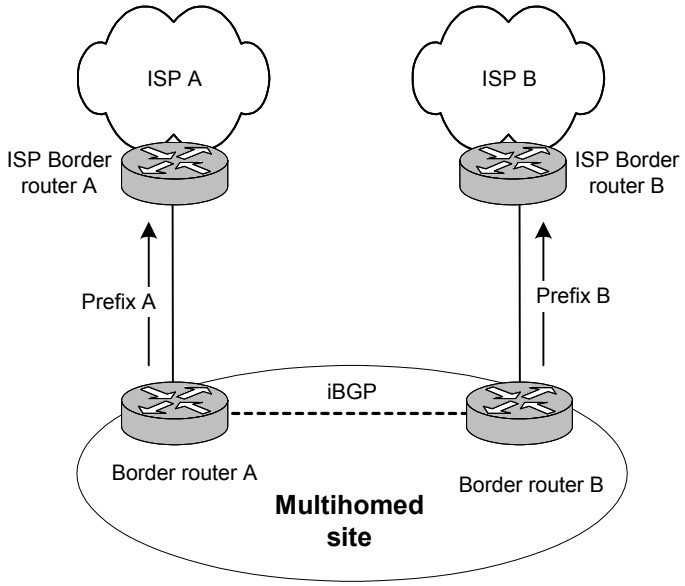

Figure 4 Non-empty intersection

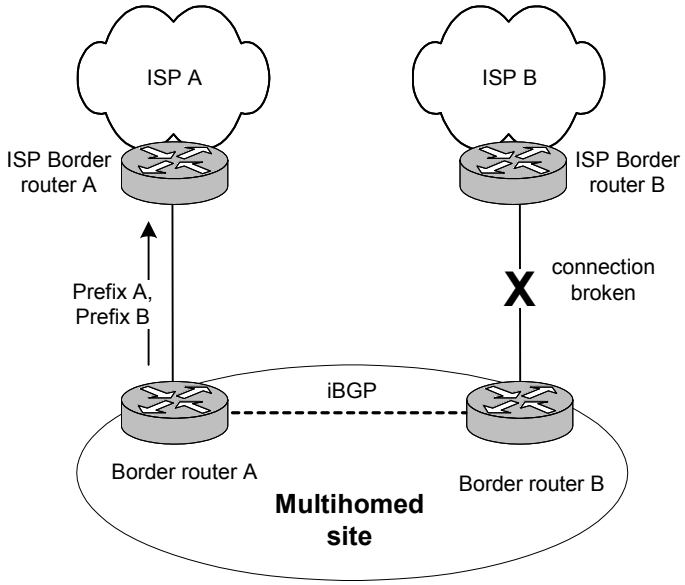

Figure 5 Empty intersection

To achieve greater scalability, RFC 2260 suggests an improvement that eliminates the need for any multihomed site prefixes to be injected into the Internet's DFZ. Quite simply, each of a site's border routers

maintains BGP peerings with all the ISPs to which the site is connected. Thus, as illustrated in Figure 6, in addition to direct BGP peerings with their directly-connected ISPs, border routers A and B maintain non-direct BGP peerings with ISPs B and A respectively.

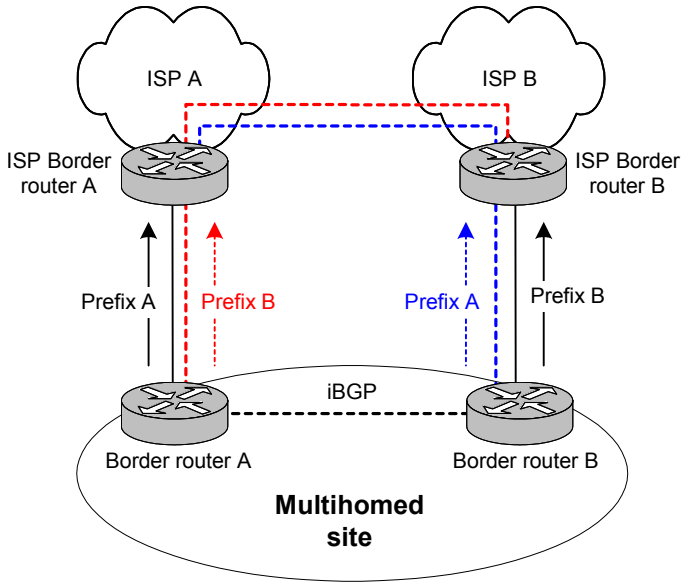

Figure 6 Non-direct peerings

The site's border router A would advertise prefix A on its direct eBGP peering with ISP A, and would advertise prefix B on its non-direct eBGP peering with ISP B. Similarly, the site's border router B would advertise prefix B on its direct eBGP peering with ISP B, and would advertise prefix A on its non-direct eBGP peering with ISP A.

Within every border router (both at the site and at the ISPs), routes received over direct peerings have greater preference to routes received over non-direct peerings. Thus, in normal circumstances, traffic destined for prefix A is routed through ISP A and traffic destined for prefix B is routed via ISP B. If, for example, the link between the site's border router B and ISP B were to fail, ISP B would then use its less-preferred route, obtained via its non-direct peering. Forwarding along a route received from a non-direct peering is performed via tunnelling. Thus, ISP B's border router would encapsulate traffic destined for prefix B and tunnel it to the site's border router A, which would then decapsulate the traffic and send it to border router B.

Using this technique means that site prefixes are never injected into the DFZ of the Internet. It also has the advantage that all of the site's prefixes are reachable via any of its connected ISPs. However, this solution is still geared towards redundancy and does not facilitate traffic engineering requirements (non-direct routes are only used when the direct routes are not available). Furthermore, when non-direct routes are used, poor traffic performance may be experienced due to the sub-optimal nature of the tunnelled routes. Sub-optimal routing may also occur during normal circumstances. Imagine another site directly connected to ISP A, and sending traffic to a host containing prefix B of the multihomed site. Normal operation would see this traffic being routed to ISP B from ISP A (possibly traversing numerous hops) whereas the optimal route would be via border router A of the multihomed site. One way around this would be to allow ISP A to receive advertisements for prefix B from border router A, and distribute this to its customer sites (but not to its transit peers).

4.3 IPv6 Multihoming Support at Site Exit Routers

RFC 3178 [31], describes multihoming support for IPv6 site exit routers. This document is simply a more detailed description, for IPv6, of the method described in RFC 2260 that uses non-direct peerings. This method can be implemented for IPv6 by using BGP4+. The non-direct peerings themselves would be implemented via IPv6-over-IPv6 tunnels. However, in some circumstances (such as when IPv6 connectivity is achieved via IPv6-over-IPv4 tunnels) it may be more appropriate to use IPv6-over-IPv4 tunnels.

Since IPv6 allows hosts to have multiple addresses, hosts within the multihomed site may be assigned addresses for each ISP the site is connected to. In this case, some sort of framework for source address selection (such as that described in [26]) needs to be in place. The problem of ingress filtering by ISP routers can be resolved using a variety of techniques as discussed in section 3.1.

4.4 Multihomed Routing Domain Issues

This Internet Draft [13] discusses some of the issues for multihomed routing domains that use the aggregatable addressing and routing scheme [14], and proposes some solutions.

4.4.1 Mutual Backup

The 'mutual backup' scheme is proposed to address the transparency issue. The fact that its customer is multihomed to a different ISP should be transparent to the provider. However, in order for the site's prefixes to be reachable through both providers, each provider must advertise reachability of both prefixes. Obviously, this breaks the transparency requirement and has poor scaling properties since the prefixes are injected into the DFZ.

The mutual backup scheme supports a scenario where two providers have an agreement to provide backup for each other.

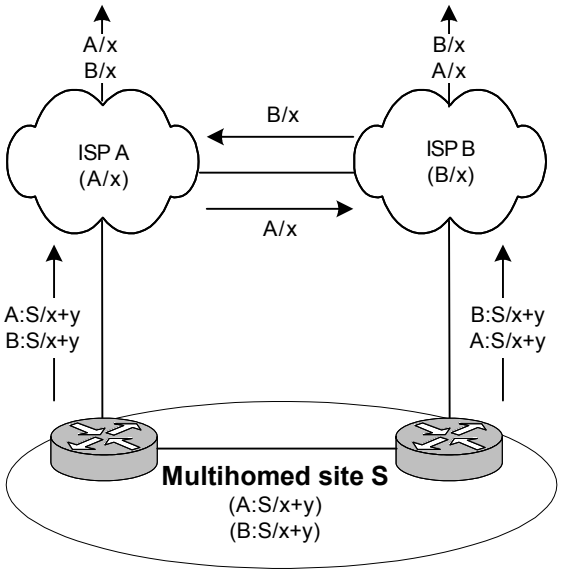

Figure 7 Mutual Backup

Traffic from the Internet destined for A:S/x+y will be routed to ISP A. At the level of ISP A and B (or above), the route advertisement for A/x from ISP A is preferred to that of ISP B since it is direct (one element in the path from A compared to two elements from B). Conversely, traffic destined for B:S/x+y will be routed to ISP B. Should the path through ISP A fail, all of the traffic for the multihomed site S (A:S/x+y and B:S/x+y) will go through ISP B.

Although this scheme works, and does not inject long prefixes into the DFZ, it does rely on ISPs A and B having a direct connection to each other and consenting to the backup agreement.

4.4.2 Router Renumbering

The Internet Draft proposes to use router renumbering [16] to force hosts inside the multihomed site to chooses the correct source address when making new connections.

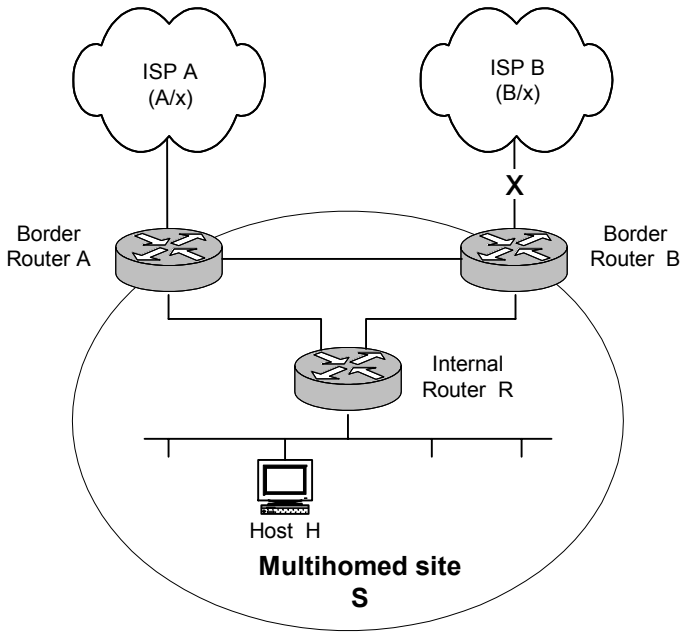

Figure 8 Router Renumbering

Referring to Figure 8, host H will have two global addresses, A:S:H and B:S:H, which it learns from the router advertisements sent by the site's internal router R. Host H is thus able to choose either A:S:H or B:S:H as its source address when making new connections. However, if the site's connection to ISP B were to fail (as shown in Figure 8), the host H should be made aware of this and thus form new connections using A:S:H as the source address. The proposal is for the site's border router (in this case border router B) to notify the site's internal routers of the broken path via the router renumbering protocol when it is detected. When the site's internal routers (in this case router R) issue their router advertisements, the 'preferred lifetime' of all the prefixes associated with the broken path are set to zero. The effect is to deprecate the associated source address so that the hosts do not use them to form new connections (and will thus choose alternative source addresses).

Of course, this will not cater for existing connections that are using the deprecated source address when link failure occurs. Mechanisms to deal with existing connections (e.g. mobility mechanisms) were discussed in section 3.

4.5 Routing Support for IPv6 Multihoming

Following on from the router renumbering technique just described, an alternative method for hosts to choose their source address, in light of network conditions, is presented in [17]. In this Internet Draft the author describes the possibility of a host within a multihomed site unwittingly selecting a source address that is unreachable on its return path due to link failure on the path associated with the source address. The solution fundamentally involves each ISP (more specifically, each AS) advertising its own prefix downwards, through lower-level aggregators towards the border router of the multihomed site. Furthermore, each prefix received from upper-level aggregators are also propagated downwards towards the border router of the multihomed site. The example in Figure 9 illustrates this procedure.

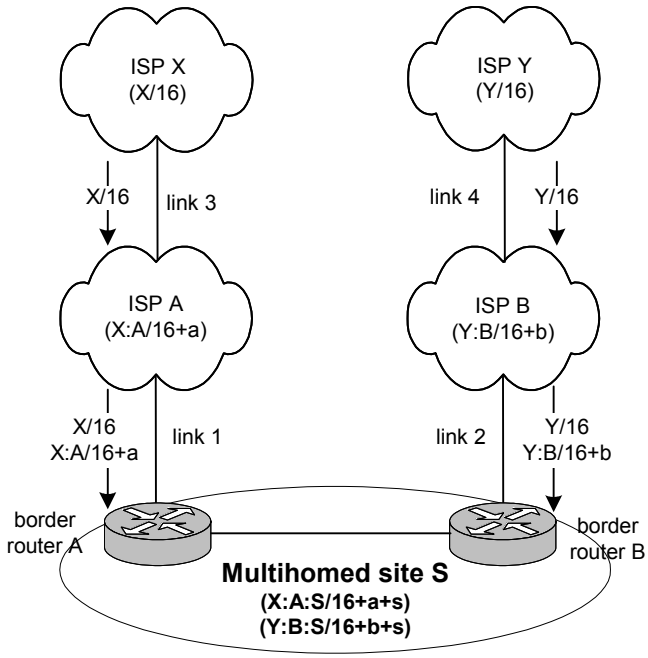

Figure 9 Routing Support for IPv6 Multihoming

The site S is multihomed to ISPs A and B, who in turn connect to their upper level aggregators, ISPs X and Y respectively. The routing information is propagated as follows:

- ISP X advertises the prefix X/16 to ISP A
- ISP A propagates X/16 and advertises X:A/16+a to site S
- ISP Y advertises the prefix Y/16 to ISP B
- ISP B propagates Y/16 and advertises Y:B/16+b to site S.

Thus, a site's border router can detect failures in the upstream path by the absence of the relevant prefix. For example, a failure of link 3 would result in the absence of the X/16 prefix advertisement to the site's border router A. Normally, border router A would have no indication of this and hosts within the site could continue to choose source addresses derived from X. If the prefix advertisements are distributed within the site, internal routers will know that the Internet is unreachable through ISP A. This information can be disseminated to the hosts (e.g. by extending the prefix option semantics in router advertisements) and used to select the correct source address for outgoing connections. Furthermore, the hosts can still select source

addresses based on a faulted prefix, if the destination address matches a prefix longer than the faulted prefix, but which is still active. As an example, if link 3 fails, a host in the site can still choose a source address based on X if the destination address matches the X:A/16+a prefix. In other words, the destination address is 'below' the link failure.

4.6 Routing Header Usage

A technique whereby a multihomed site can force routing via a specific ISP is suggested in [35] and [12]. This technique consists of inserting a routing header containing an intermediate anycast address that identifies the routers of the intended ISP. Thus, the packets containing such a routing header will be explicitly routed to the selected ISP. However, this method has the drawback that the ISP must process the routing headers; a possible expensive overhead especially if many sites were to employ this technique.

An alternative to this would be to use site exit router selection. Instead of the routing header containing the anycast address for an ISP, it would contain an address specifying a site-exit router that connects to the desired ISP. If only one site exit router is connected to the desired ISP, the address in the routing header would be a site-local address of one of the router interfaces. If the desired ISP is served by multiple site exit routers, the address in the routing header would be an anycast address. This method has the advantage that the routing header is processed by the site exit routers and not by the ISP routers, thus demonstrating greater scaling properties. It also means that the ISP's co-operation in implementing all routers anycast does not have to be sought by the customer site.

4.7 MHAP

The Multi-Homing Aliasing Protocol (MHAP) [18], is a solution for site multihoming that does not affect the routing tables of the DFZ and does not use tunnels. The fundamental concept is that multihomed traffic (defined as traffic whose destination address is in the multihomed address space) is aliased and transported across the Internet as singlehomed traffc.

MHAP makes use of two types of address space, *singlehomed* and *multihomed*. Singlehomed addresses are any globally aggregatable IPv6 address except those that are reserved for multihomed addresses. Multihomed addresses may be either from reserved blocks of globally aggregatable IPv6 addresses (2345::/16 and 3FFE:FFF::/32 are used in the draft), and are know as *MHAP prefixes*, or from blocks of geographic PI (Provider Independent) addresses (the draft uses 2346::/16 and 3FFE:FD00::/24), and are known as *geo-PI prefixes*. MHAP prefixes are allocated to organisations and are portable, whereas geo-PI prefixes are allocated to a geographical area and are not portable.

The multihomed address space exists only at the edges. End-to-end multihomed traffic is carried over the core of the Internet using the aggregated singlehomed address space and the mapping between multihomed and singlehomed addresses is held in the *MHAP alisasing table*. Thus, an end-host will use a multihomed address as the destination address for its peer. When an outgoing IPv6 packet reaches the first MHAP router, the multihomed destination address is replaced with one of the possible singlehomed addresses owned by the associated host. The packet is transported across the core of the Internet using the singlehomed destination address. When the packet reaches the destination site, another MHAP-capable router replaces the singlehomed destination address with the original multihomed destination address, thus preserving the end-to-end semantics.

Since, the multihomed address space exists only at the network edge, MHAP has the advantage that no additional routes are injected into DFZ routing tables. The MHAP Internet Draft is still very much work in progress. Consequently, its feasibility of deployment and relative strengths and weaknesses need further study to be determined. There are obviously some security and performance implications for allowing destination addresses to be altered in transit, although the same problems exist for NAT services yet they are widely deployed.

4.8 "GEO For Now"

The Internet Draft "Provider-Internal Aggregation based on Geography to Support Multihoming in IPv6" [19], (referred to as "Geo for Now") proposes new operational practices that will allow networks to handle a much larger global routing table, so multihoming in IPv6 can be made possible within a very short time frame. In essence, Geo for Now proposes to relax the condition that the full global routing table must be present in every router in the DFZ. Instead, the global routing table is split into several parts, or geographical zones using the geographical addressing scheme described in [20]. Each router in the DFZ announces an aggregate for the part/zone of the global routing table that it serves. Since each router in the DFZ would only need to manage its own part/zone, a global routing table of arbitrary size could be efficiently managed by the DFZ as a whole. In other words, injecting specific routes for multihomed networks into the DFZ will be much less of a problem than it is today. It is claimed in [19] that growth to several million multihomed networks should be possible without incurring too much of a performance hit in the DFZ.

Although this is a technically feasible short-term solution, there are doubts as to whether this would ever be realistically deployed as there are no economic incentives for Internet exchanges to support this model.

5 Conclusions

The discussion in section 3 shows that multihoming support in (mobile) terminals needs enhancements both in the terminals and also support by routers, depending on the approach taken. In terminals, the following areas of work can be identified:

- Enhancements to support site multihoming, e.g. multi-site aware address selection. Depending on the approaches above, this is also needed for terminals with only one interface.
- For nodes equipped with multiple interfaces, extensions to support re-routing traffic through another interface, and
- configuration support for users to configure interface preferences based on traffic classes.

Thus, approaches such as the one mentioned in Brain/MIND in the multihomed terminal case might be required for a future multihomed scenario.

Regarding site multihoming there are several problems

- Route filtering is a problem i.e. ISPs that filter routes based on the length of the address prefixes
- Inter-ISP connection and/or co-operation is required for many solutions to work

What is apparent is that, so far, no single multihoming solution will cater for all the requirements needed for both site and host based multihoming. It is most likely that several mechanisms will need to be used together to satisfy the various requirements. For example, the router renumbering mechanism can be used in conjunction with multihoming support at site exit routers. When the connection to one ISP fails, existing connections can be re-routed in tunnels whilst the router renumbering informs hosts to use an alternative source address to make new connections. In an alternative but similar combination, mobility or transport-level mechanisms could be used to cater for the existing connections while the router renumbering again caters for new connections.

References

[1] C. Huitema, R. Draves, “Host-Centric IPv6 Multihoming”, IETF Internet-Draft draft-huitema-multi6-hosts-01.txt, June 2002 (work in progress).

[2] N. Montavont, T. Noel, M. Kassi-Lahlou, ”MIPv6 for Multiple Interfaces”, IETF Internet-Draft draft-montavont-mobileip-mmi-00.txt, July 2002 (work in progress).

[3] D. B. Johnson, C. E. Perkins, J. Arkko, “Mobility Support in IPv6”, IETF Internet-Draft draft-ietf-mobileip-ipv6-18.txt, June 2002 (work in progress).

[4] H. Soliman, K. El-Malki, C. Castelluccia, “Per-flow movement in MIPv6”, IETF Internet-Draft draft-soliman-mobileip-flow-move-01.txt, November 2001 (work in progress).

[5] IST 1999-10054 Project BRAIN, Deliverable D2.2, March 2001, http://www.ist-brain.org

[6] LIN6 homepage, http://www.lin6.net/

[7] F. Teraoka, M. Ishiyama, M. Kunshi, A. Shionozaki, “LIN6: A Solution to Mobility and Multi-Homing in IPv6”, IETF Internet Draft draft-teraoka-ipng-lin6-01.txt, August 2001.

[8] R. Stewart, et al, “Stream Control Transmission Protocol”, IETF RC 2960, October 2000.

[9] M. Ohta, “The Architecture of End to End Multihoming”, IETF Internet Draft draft-ohta-e2e-multihoming-02.txt, July 2001.

[10] Site Multihoming in IPv6 (multi6), http://www.ietf.org/html.charters/multi6-charter.html

[11] IPv6 Multihoming Solutions (ipv6mh), http://arneill-py.sacramento.ca.us/ipv6mh/

[12] M. Bagnulo, A. Garcia-Martinez, A. Azcorra, D. Larrabeiti, “Survey on Proposed IPv6 Multi-Homing Network Level Mechanisms”, IETF Internet Draft draft-bagnulo-multi6-survey6-00.txt, July 2001.

[13] F. Dupont, “Multihomed Routing Domain Issues for IPv6 Aggregatable Scheme”, IETF Internet Draft draft-ietf-ipngwg-multi-isp-00.txt, September 1999.

[14] R. Hinden, M. O’Dell, S. Deering. “An IPv6 Aggregatable Global Unicast Address Format”, IETF RFC 2374, July 1998.

[15] S. Thomson, T. Narten, “Stateless Address Autoconfiguration”, IETF RFC 2462, December 1998.

[16] M. Crawford, “Router Renumbering for IPv6”, IETF RFC 2894, August 2000.

[17] N. Bragg. “Routnig Support for IPv6 Multi-homing”, IETF Internet Draft draft-bragg-ipv6-multihoming-00.txt, November 2000.

[18] M. Py, “Multi Homing Aliasing Protocol (MHAP)”, IETF Internet Draft draft-py-mhap-01a.txt, April 2002.

[19] I. Van Beijnum, “Provider-Internal Aggregation based on Geography to Support Multihoming in IPv6”, IETF Internet Draft draft--van-beijnum-multi6-isp-int-aggr-00.txt, October 2002.

[20] I. Van Beijnum, M. Py, “A Geographically Aggregatable Provider Independent Address Space to Support Multihoming in IPv6”, Work in Progress.

[21] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, V. Paxson, “Stream Control Transmission Protocol”, IETF RFC 2960, October 2000.

[22] M. O’Dell, “GSE – An Alternative Addressing Architecture for IPv6”, IETF Internet Draft draft-ipng-gseaddr-00.txt, February 1997.

[23] T. Hain, “Application and Use of the IPv6 Provider Independent Global Unicast Address Format”, IETF Internet Draft draft-hain-ipv6-pi-addr-use-02.txt, March 2002.

[24] J. Yu, “IPv6 Multihoming with Route Aggregation”, IETF Internet Draft draft-ietf-ipngwg-ipv6multihome-with-aggr-01.txt, August 2000.

[25] M. Blanchet, “A Flexible Method for Managing the Assignment of Bits of an IPv6 Address Block”, IETF Internet Draft draft-ietf-ipv6-ipaddressassign-03.txt, August 2002.

[26] R. Draves, “Default Address Selection for IPv6”, IETF Internet Draft draft-ietf-ipv6-default-addr-select-09.txt, August 2002.

[27] L. Coene, “Multihoming Issues in the Stream Control Transmission Protocol”, IETF Internet Draft draft-coene-sctp-multihome-03.txt, February 2002.

[28] L. Coene, “Multirouting”, IETF Internet Draft draft-coene-multi-00.txt, February 2002.

[29] F. Dupont, “The Host Identity Payload Protocol: toward a secure solution to mobility and multi-homing”, May 2002.

[30] T. Bates, Y. Rekhter, “Scalable Support for Multihomed Multi-provider Connectivity”, IETF RFC 2260, January 1998.

[31] J. Hagino,H. Snyder, “IPv6 Multihoming Support at Site Exit Routers”, IETF RFC 3178, October 2001.

[32] B. Black, V. Gill, J. Abley, “Requirements for IPv6 Multihoming Architectures”, IETF Internet Draft draft-ietf-multi6-multihoming-requirements-03.txt, May 2002.

[33] J. Abley, B. Black, V. Gill, “IPv4 Multihoming Motivation, Practices and Limitations”, IETF Internet Draft draft-ietf-multi6-v4-multihoming-00.txt, June 2001.

[34] G. Huston, “Analyzing the Internet’s BGP Routing Table”, Internet Protocol Journal, Volume 4 Number 1, March 2001.

[35] D. Johnson, S. Deering, “Reserved IPv6 Subnet Anycast Addresses”, IETF RFC 2526, March 1999.

[36] P. Tattam, “Preserving Active TCP Sessions on Multihomed IPv6 Networks”, September 1999.

[37] P. Ferguson, D. Senie, “Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing”, IETF RFC 2827, May 2000.

Abbreviations

AS	Autonomous System
BGP	Border Gateway Protocol
eBGP	Exterior Border Gateway Protocol
CIDR	Classless Inter-Domain Routing
iBGP	Interior Border Gateway Protocol
DFZ	Default-Free Zone
IGP	Interior Gateway Protocol
IPSec	IP Security
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISP	Internet Service Provider
PA	Provider Aggregatable
PDA	Personal Digital Assistant
PI	Provider Independent
SCTP	Stream Control Transmission Protocol
TCP	Transport Control Protocol
UDP	User Datagram Protocol
VoD	Video on Demand
VoIP	Voice over IP