Pekka Savola

Examining Site Multihoming in Finnish Networks

CSC - Tieteellinen laskenta Oy CSC - Scientific Computing Ltd

HELSINKI UNIVERSITY OF TECHNOLOGY

ABSTRACT OF THE MASTER'S THESIS

Author:	Pekka Savola		
Name of the Thesis:	Examining Site Multihoming in Finnish Networks		
Date:	April 15, 2003	**Number of pages:**	81
Department:	Electrical and Comm. Engineering	**Professorship:**	S-38
Supervisor:	Raimo Kantola		
Instructor:	Jari Miettinen		

Site multihoming means end-sites connecting to multiple separate network service providers; currently, no IPv6 site multihoming mechanism has been widely accepted yet.

This thesis studies both IPv4 and IPv6 site multihoming mechanisms using literature, other similar studies, analysis of route advertisements at the Finnish exchange point FICIX, and queries on multihoming practices to major ISPs in Finland.

Currently in IPv4, there seem to be three-four main mechanisms which are used to achieve at least some of the multihoming benefits: obtaining their own address space and AS number and advertising those, advertising more specific routes with a different path, using multi-connecting and leveraging NAT.

In IPv6, the first two of IPv4 mechanisms which are considered architecturally unscalable have been operationally prevented for now and the fourth does not exist.

Based on a tentative roadmap introduced in this thesis, organizations are split into four categories: minimal, small, large and international; each have different multihoming requirements which can be met in different ways.

Focusing on immediate and short-term solutions, minimal organizations do not seem to require a solution, small ones could use multi-connecting, host-centric multihoming or multihoming at site exit routers, large ones the same or possibly separate provider independent address allocations, and international ones either provider independent allocations or be broken down to multiple large organizations and using techniques specified for large organizations.

It is apparent that only a limited amount of work is needed to enable sufficiently good multihoming mechanisms which should provide the required features. However, as the mechanisms are unarguably more difficult for the end-site, while taking the global Internet routing architecture better into account, it is unclear whether they might be adopted.

Keywords: site multihoming, multihoming, IPv6, multi6

TEKNILLINEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Pekka Savola	
Työn nimi:	Loppukäyttäjäverkkojen moniliittyminen suomalaisissa IP-verkoissa	
Päivämäärä:	15.4.2003	**Sivuja:** 81
Osasto:	Sähkö- ja tietoliikennetekniikka	**Professuuri:** S-38
Työn valvoja:	Raimo Kantola	
Työn ohjaaja:	Jari Miettinen	

Loppukäyttäjäverkkojen moniliittymisellä tarkoitetaan loppukäyttäjien muodostavien IP-verkkojen, kuten yritysverkon, liittymistä Internetiin useamman verkko-operaattorin kautta yhtäaikaisesti. Tällä hetkellä IPv6:lle tähän ei ole yleisesti hyväksyttyjä mekanismeja.

Diplomityössä tutkitaan loppukäyttäjäverkkojen moniliittymistä sekä IPv4 että IPv6 -protokollilla kirjallisuutta, muita vastaavia tutkimuksia, reititysmainostuksia FICIX-yhdysliikennepisteessä, sekä suomalaisille operaattoreille suunnattua kyselyä käyttäen.

IPv4:ssa on olemassa kolme-neljä menetelmää jolla saavutetaan ainakin osa halutuista hyödyistä: omien osoitteiden ja AS-numeron hankkiminen ja näiden mainostaminen, tarkemman reitin mainostaminen, monikytkeytyminen tai verkko-osoitemuunnoksen (NAT) käyttö.

IPv6:ssa kaksi ensimmäistä IPv4:n menetelmää ovat operatiivisesti estetty koska niitä pidetään arkkitehtuurillisesti skaalautumattomina, ja neljättä ei ole.

Diplomityössä esitellyn yleissuunnitelman avulla loppukäyttäjäverkot jaetaan neljään kategoriaan: minimaaliset, pienet, suuret ja kansainväliset; jokaisella näistä on erilaiset vaatimuksensa moniliittymiselle.

Keskittyen välittömiin ja lyhyen aikavälin ratkaisuihin, minimaaliset verkot eivät tarvitse ratkaisua, pienet voivat käyttää monikytkeytymistä, konekeskeistä moniliittymistä tai virtuaalitopologista moniliittymistä, suuret verkot samoja menetelmiä tai operaattoririippumattomia osoitteita ja kansainväliset joko riippumattomia osoitteita tai jakautumista useampiin suuriin verkkoihin ja näissä vastaavien menetelmien käyttämistä.

Näyttää selvältä, että tarvitaan vain rajallinen määrä työtä riittävän hyvien, vaatimuksia vastaavien mekanismien viimeistelemiseksi. Mekanismit ottavat koko Internetin reititysarkkitehtuurin paremmin huomioon, mutta ovat kuitenkin loppukäyttäjäverkolle hankalampia, joten on vielä epäselvää tullaanko niitä käyttämään.

Avainsanat: moniliittyminen, monikytkeytyminen, IPv6, multi6

Acknowledgments

This master's thesis has been done for CSC - Scientific Computing Ltd while working at the Funet network group.

I want to thank my instructor, Jari Miettinen, for providing excellent guidance and tirelessly going through different revisions of the thesis.

I would also like to thank my unofficial instructors, Juha Oinonen and Klaus Lindberg, who have been closely involved throughout the whole process and been able to offer good commentary and guidance.

I wish to thank my supervisor, Raimo Kantola, who has endured my unfocused ramblings in the early stages, and provided commentary and different perspective to the thesis.

I would also wish to thank my colleagues, past and present, and the work environment where I have been able to learn.

A special thanks also goes to my father for long discussions on every possible topic, and who has always been sparring with me: it has sparked my curiosity in many subjects and kept me going, always expanding my views.

My gratitude also goes to all the participants in the IETF with whom I've been able to exchange ideas with, especially those in the IPv6 site multihoming working group who have kept on going despite all odds.

Otaniemi, April 15, 2003

Pekka Savola

Contents

Abbreviations

Abbreviations

AS = Autonomous System
ASN = Autonomous System Number
BGP = Border Gateway Protocol
DNS = Domain Name System
FICIX = Finnish Communication and Internet Exchange
HIP = Host Identity Payload (and Protocol)
ICMP = Internet Control Message Protocol
IGP = Interior Gateway Protocol
IP = Internet Protocol
IRR = Internet Routing Registry
IS-IS = Intermediate System to Intermediate System
ISP = Internet Service Provider
LIN6 = Location Independent Addressing for IPv6
LIR = Local Internet Registry
NAPT = Network Address Port Translation
NAT = Network Address Translation
OSPF = Open Shortest Path First
PA = Provider Aggregatable/Allocated/Assigned
PI = Provider Independent
RIP = Route Information Protocol
RIR = Regional Internet Registry
RPSL = Routing Policy Specification Language
SCTP = Stream Control Transmission Protocol
TCP = Transmission Control Protocol
TE = Traffic Engineering
UDP = User Datagram Protocol
VPN = Virtual Private Network

Chapter 1

Introduction

Multihoming means connecting to multiple separate network service providers; site multihoming restricts this to end-sites which do not offer connectivity services themselves. These will be defined in more detail in the next chapter.

The problem is that the extent, reasons and mechanisms used for IPv4 site multihoming are not clear.

In consequence, there are no good established practices for other than a limited subset of multihoming in IPv6 either.

If how and why IPv4 site multihoming is done is not clear, it does not seem to be possible to be able to reach any good results when designing IPv6 site multihoming solutions. Therefore, the path from IPv4 to IPv6 site multihoming needs to be explored at both ends, not just in IPv6 when defining new practices.

Tt seems that the mechanisms used with IPv4 are unscalable for the Internet because they require sites are present in the global routing table. On the other hand, the number of sites is so huge that this multihoming model is unsustainable.

Long transition to the co-existence of IPv4 and IPv6 is underway and will become more important in the future. Currently, no IPv6 multihoming mechanism for end-sites has been widely accepted yet. Some feel that the lack of a viable multihoming mechanism will make potential IPv6 users more reluctant to start using it. Methods used with IPv4 have been deemed unscalable and prevented operationally – the new version of the protocol has been considered a possibility to develop a new concept for most site multihomers' needs.

The main objective of this thesis is to study IPv4 and IPv6 network multihoming mechanisms, and to find out which kind of network multihoming, in particular, site multihoming mechanisms are currently being used with IPv4.

Based on this study, it is hoped that one is able to better understand the motivations, mechanisms and actual requirements for multihoming as we transition to IPv6.

The research is based on literature, other similar studies, analysis of route advertise-

ments at the Finnish exchange point FICIX, and queries on multihoming practices to major ISPs in Finland.

The secondary objective of this thesis is to create a model on different IPv4 multihoming types and how they're used, and how the proposed IPv6 site multihoming solutions apply to these practices; the research makes this possible.

Finally, an approach how to start untangling the IPv6 site multihoming complexity is presented, and future, short-term work items are identified.

The thesis is structured as follows.

In the second chapter, the basic terminology and scope of the work are defined and common motivators for doing multihoming are described. In particular, multihoming by a single node or whole ISPs are considered out of scope.

In the third chapter, the basics required to understand multihoming are described, and the data collection environment, collection procedure and processing are introduced.

The fourth chapter concentrates on the site multihoming: the scalability problem of all the end-sites multihoming with current mechanisms is elaborated, some generic constraints and issues to consider when designing site multihoming solutions presented, and an overview of both IPv4 and IPv6 multihoming mechanisms, as well as multi-connecting techniques are described.

The fifth chapter presents a categorization for current IPv4 multihoming-related route advertisements, describes the other means and results of collecting information on current multihoming practices, presents collected and processed route advertisement data, and in this light describes how multi-connecting is used in current networks and how current multihoming organizations can be divided into a few classifications, and how they fit there.

The sixth chapter presents an analysis of IPv6 site multihoming solutions, proposes a classification of organizations and mechanisms which would seem to apply to such classes, and finally examines the methods for choosing a multihoming mechanism and how the classifications and mechanisms apply.

The seventh chapter presents conclusions, related and future work.

Appendix A includes the query sent to major ISPs and an example ISP-specific appendix attached to the queries.

Chapter 2

Multihoming Background and Research Scope

In this chapter, the multihoming terminology is described, as well as the three major classes of multihoming. A common element in more than one of the classes is also noted. Based on these, the research scope is defined. Last, different motivations why multihoming is needed, which sets requirements for the multihoming mechanism, are given.

Next chapters will give background knowledge required to understand multihoming, insight to the data collection and processing to be done, site multihoming in particular, current IPv4 multihoming data analysis and the future of multihoming in IPv6.

2.1 Multihoming

Multihoming means connecting to multiple network service providers instead of just one.

Some use the term "multihoming" only when explicitly connecting to different network service providers. For multiple connections to a single operator, terms "multi-attaching" or "multi-connecting" are sometimes used [1].

This terminology is not well-established.

In this thesis, only connecting to different operators is considered multihoming.

However, as reasons for multi-connecting are often the same as with multihoming, multi-connecting will be briefly described in sections 4.5 and 5.4 and analyzed in section 6.1.8.

2.2 Different Types of Multihoming

The multihoming types can either be grouped based on the problem one is trying to solve, or based on a set of solutions.

As encouraged by [2], the separation is done based on the problem; this is considered a more open-minded alternative.

If the grouping was based on solutions, the second subset of site multihoming solutions would go under node multihoming.

2.2.1 Node Multihoming

In practice, node multihoming means obtaining network connectivity to a single node, typically a host, from multiple providers.

Only one set of solutions is known: the node must have multiple globally connected network interfaces. An example of this could be a very important server with multiple active Ethernet interfaces. These interfaces would use different IP addresses from different network connectivity providers.

2.2.2 Site Multihoming

Terminology

In short, site multihoming practically refers to enterprises connecting to more than one network service provider.

A longer definition is a bit more problematic; "site" has never been precisely defined, so the term "site multihoming" is also a bit vague; it can better be understood by describing what it is not rather than what it is.

Site is used to refer to the end-user – usually organization, but can also be a private person's network – which provides no connectivity to other sites; ie., is not a network service provider itself; a degenerate case of a site is a single node, which is described above.

Solutions

There are three main classes of solutions.

First, using only one IP address per node. Multihoming is achieved by making the address block reachable through more than one Internet Service Provider (ISP) using routing protocols. The result is similar to what happens with ISP multihoming, as described below.

Second, using multiple IP addresses per node, each block obtained from a different ISP where the site connects to, and nodes having only one network interface. Mul-

tiple addresses are assigned to the same network interface, and the default router of the node provides service for both of these addresses obtained from different ISP's. The node is not physically multihomed itself, as described in the node multihoming above; rather, some form of multihoming is provided by the network. An example of this is a subnet, the router of which has connections to two ISP's and is providing service to hosts using two different blocks of addresses.

Third, a combination of the two: nodes use only one address, but it is replaced by another address when leaving or entering the network using some address translation mechanism.

2.2.3 ISP Multihoming

The term ISP multihoming is used to describe the mechanisms which major network service providers use to connect to more than one upstream network connectivity provider.

ISP's are generally assumed to be independent, resilient against failures, providing network access services. They're expected to obtain their own address space, the portions of which will be assigned to their customers on need.

ISP's have the one widely accepted solution: using a routing protocol for advertising reachability through multiple upstream service providers to the whole Internet.

Note that very small ISP's are likely to be classified as sites, as they do not have the resources to justify the requirements for a full ISP multihoming solution.

2.2.4 Generic Features

One special set of solutions which may be applied to any mechanism employing multiple addresses is providing a mapping function between different kinds of addresses or parts of addresses using some mechanism.

Different mappings include separating "locator identifier" and "end-point identifier" addresses, or the first part of the address forming the locator and the second part the global identifier.

Examples of the former are Mobile IPv6 [3] and, to a greater extent, the Host Identity Protocol (HIP) [4], and of the latter Location Independent Addressing (LINA) [5]. These will be described in section 4.4.2.

2.3 The Research Scope

The scope of this thesis is to examine site multihoming solutions.

Therefore, node multihoming solutions, ie. the nodes having more than one globally connected interface, are out of scope – a single node is not considered a network.

Also, ISP multihoming, which will be noted for example in section 5.1.4, is out of scope. Major ISP's are expected to be able to be completely provider independent, and use routing protocols to multihome accordingly all the time; this is therefore an accepted practice and design feature compared to the site multihoming problem.

Site multihoming is a very extensive area, consisting of multiple different approaches to the problem.

The focus is on the first category of site multihoming, as that is what is being done today, and is the only thing that can be examined globally.

The second category solutions are based on the network providing a part of the multihoming service. This is not a common model today, but is likely to be more so in the future; therefore it will be analyzed. This is because it is not believed that the first category will be scalable and proper long-term approach for site multihoming.

The third category will be only briefly described when discussing solutions employing NAT or certain proposed IPv6 multihoming techniques.

Generic features with mechanisms using multiple addresses, as described above, are partially applicable to multihoming solutions. They will be noted only briefly in section 4.4.2.

2.4 Examining Motivations for Doing Multihoming

To understand why some organizations choose to multihome, or choose a specific solution for multihoming, some motivations must be examined.

Motivations are mostly based on [1]. IPv4 and IPv6 site multihoming mechanisms are described in sections 4.3 and 4.4; current trends in IPv4 are examined in section 5.3, and further analysis on IPv6 is done in section 6.3.

These motivations are observed and analyzed later, mainly in sections 5.5 and 6.2.

2.4.1 Independence

Technical reasons for independence are mostly covered under redundancy; independence here focuses on economic, political and administrative perspectives.

Multihoming, especially with your own addresses, a typical case as described in section 5.1.1, can bring the organization some degree of provider independence. If the organization can just change its ISP's with minimal effort, one undoubtedly has a better standing in the service agreement negotiations, e.g. being able to get a cheaper price and requested service agreement duration. ISP's could also just disappear, but in most cases services usually continue uninterrupted in the event of bankruptcy, mergers, etc. Also, especially some larger organizations consider it important to stand on their own, so that they can't be seen to depend on others.

An important factor for many, especially bigger sites, is also that the renumbering in

the event of a network service provider change is considered too difficult and time-consuming, and it is desirable to find mechanisms which do not require that; such feature is provided by independence.

In [1], this motivation was only briefly touched when discussing policy, so I've raised raised it as an important issue on its own.

2.4.2 Redundancy

Multihoming can provide a significant amount of protection for example for the following failure modes, some of them as described in [1]:

- physical link failures (e.g. fiber cuts due to construction works or problems in the lower-layer service provider such as Synchronous Digital Hierarchy (SDH) routing mistakes)
- other hardware failures (e.g. router hardware problems, in the line card or some other component)
- logical link failures (e.g. software failure in router link processing)
- some routing protocol failures (e.g. one misbehaving peer)
- human errors (e.g. configuration mistakes, but only to an extent)
- transit provider failures (any of the above in the upstream ISP's providing service to the organization, unless the transit in turn has adequate protection against such failures)

There is a small and very rare subset of software-related problems, ie. "bugs", which cannot be fully protected against. For example, if routers' routing tables and forwarding tables become badly de-synchronized, the routing protocols may still function without problems while the actual packet forwarding is disrupted.

Redundancy also includes the concept of convergence time; that is, how long it takes to work around a detected failure, to switch to the redundant mode. Some mechanisms provide better convergence characteristics than others; if fast convergence is important, the set of possible mechanisms providing redundancy may be more limited.

2.4.3 Load Sharing

Some consider it important to be able to distribute the traffic between several links or several different ISP's. Reasoning behind this might be that any one transit provider might not have enough capacity to all the destinations the organization wishes to be reachable to.

The model how networks are built, maintained, and extended is also a factor. Networks that evolve all the time are more likely to require mechanisms for load sharing, especially for transitional periods. This kind of requirement is typical with ISP's, but rather rare with end-sites.

Outbound load sharing is very easy to accomplish, even with almost all mechanisms, but inbound load sharing is a significantly more difficult issue, because one has to somehow distribute the desired policy to source nodes in the Internet.

Load sharing is particularly important in large regions, like the U.S., where network operators typically wish to exchange traffic locally if possible, but some operators may not have significant presence in all areas.

2.4.4 Performance

In case some transit providers are experiencing performance difficulties to some destinations, it may be useful to be able to redistribute traffic on other providers.

Performance can include several different metrics like delay, jitter, raw capacity, reliability in the form of a Service Level Agreement (SLA), and others.

If there are several special usage scenarios, like high bandwidth requirements for some traffic, and low delay/jitter for some other, sometimes such traffic could be better distributed to several providers, for example based on policy reasons, below.

2.4.5 Policy

It is sometimes desirable to be able to distribute the traffic based on some policy of non-technical nature, e.g. cost, acceptable use – especially with joint research/commercial organizations, commercial reasons or others.

Chapter 3

Background, Data Collection and Processing

In the previous chapter, terminology, scoping and motivations were introduced. This chapter describes required basic background that's needed for understanding the concepts; the data collection environment is also described. In the next chapters, the focus is looking deeper into site multihoming, and doing the actual analysis of multihoming practices in IPv4 and IPv6.

First, literature background is described: the building blocks used in multihoming. Then, the data collection environment and the procedures are introduced. Last, the data collection and processing are described.

3.1 Multihoming Building Blocks

This section gives an overview of current building blocks used for multihoming purposes. These include Border Gateway Protocol (BGP) details to a sufficient degree, route aggregation, IP address allocation and assignment, routing policy management and databases, IP address formats and Network Address Translation (NAT).

3.1.1 Border Gateway Protocol

Border Gateway Protocol (BGP) [6] is a very important routing protocol, integral to the routing between administrative domains. Its operation and the way it works must be understood prior to being able to realize the implications on multihoming.

Therefore, BGP will be covered at length. First, fundamentals and protocol details introduce the basic operation; then path attributes and best path selection describe two important details of the operation. Last, a few operational procedures are described.

Fundamentals

BGP, based on a path vector algorithm, is the routing protocol that is being used between different Autonomous Systems (AS's); it has been the single exterior routing protocol, excluding static routing for very simple interconnections, for over a decade.

The term Autonomous System is very commonplace and is used to refer to the set of routers under a single administrative body; a domain, which has a coherent routing plan and has a consistent view of how traffic is handled. These AS's are connected to other AS's with an exterior routing protocol like BGP and internally use some interior routing protocol(s) such as OSPF, IS-IS, or RIP. The generic term for any interior routing protocol is Interior Gateway Protocol (IGP).

BGP sessions are established using the TCP protocol with neighboring AS's, typically one or more BGP-speaking border routers. Over these sessions, network reachability information is advertised. The advertised network reachability is a subset of one used by the advertising BGP speaker; only routes it itself uses are eligible.

How, and what, network reachability is advertised is a routing policy decision. That is how exterior routing protocols generally differ from IGP's; instead of optimizing the shortest path within a domain, BGP is mainly a tool which can be used manage how routing goes on a higher level.

The use of the reliable transport protocol TCP ensures that the BGP protocol itself does not need to explicitly implement retransmissions, acknowledgments or any such features that are typically necessary for IGP's. In addition, any authentication scheme applicable to TCP is also applicable to BGP.

Protocol Details

When a BGP session is established, first Open and Confirm messages are exchanged to set up connection parameters. Then the entire BGP routing table – which is typically a subset of the routing table, controlled by policy – is sent. Subsequent updates using Update messages are incremental; BGP speakers must store the current version of BGP routing tables for the duration of the connection. KeepAlive messages are sent periodically to make sure the connection is still up and functional. Notification messages are sent on special occasions or when responding to error conditions. If an error condition occurs, a notification is sent and the connection is closed and must be established again after the error has been fixed.

BGP speaker does not have to be a router, necessarily. An example of these non-routing BGP speakers are so-called route servers which can be used for example for collecting and analyzing routing policies.

As mentioned before, the view of an AS must be consistent. This is especially important in systems which provide transit service for other AS's and have multiple BGP speakers. Internal routes may be confirmed to be consistent using IGP; external routes, e.g. those for which transit is provided to, can be synchronized by

forming BGP sessions between all the internal BGP speakers. Often, the networks are designed so that IGP includes only router loopback and point-to-point interface addresses, and BGP everything else. Typically some routers are chosen to act as entry/exit points for specific outside destinations.

The above implies an important distinction in BGP: the logical separation of "internal BGP" and "external BGP". External BGP typically has a per-neighbor BGP routing policy. Internal BGP is always fully meshed between the routers in an AS, and uses typically the same routing policy; this basic tenet can be modified by using e.g. BGP Route Reflectors, Confederations, or a design based on private AS numbers.

Path Attributes

In addition to the IP address prefix and prefix length, the BGP update messages also contain other data such as path attributes. Path attributes convey vital information, so they are covered here in detail. Path attributes are encoded in Type, Length, Value (TLV) -notation so they're easily extensible.

Attribute type consists of attribute flags (1 byte) and an attribute type code (1 byte). The former is used to encode more generic features of the attribute, the latter the specific attribute type.

The most important attribute flags are, from the most significant to least significant bit:

- whether the attribute is optional or well-known
- whether the attribute is transitive or non-transitive
- whether the information in the attribute is partial or complete
- whether the attribute length is one or two octets, sometimes called an "extended length" attribute

The last 4 bits haven't been specified. Well-known attributes must always be transitive. Only optional transitive attributes can be partial.

Well-known attributes must be recognized and implemented by every BGP speaker, and being transitive, must always be passed on to other BGP speakers. An additional category for well-known attributes are "mandatory" meaning it must be sent in every Update message or "discretionary" which are only sent when needed.

Attribute type codes are defined as follows:

1. ORIGIN; a well-known, mandatory attribute which communicates the origin of the path information from the originating AS. It can be either "IGP", "EGP", or "Incomplete".
2. AS_PATH; a well-known, mandatory attribute which is composed of AS path ("path vector") the route has been passed through.

3. NEXT_HOP; a well-known, mandatory attribute which defines the IP address of the router which should be used to reach the advertised prefix.

4. MULTI_EXIT_DISC (multi-exit discriminator, MED); an optional, non-transitive attribute which can be used to influence the neighbor AS's decision making process when there are multiple exit points to the same AS.

5. LOCAL_PREF (local preference); a well-known attribute which can be used to inform internal BGP speakers of the local AS's preferences towards the multiple instances of the same prefix. It is not conveyed on external BGP sessions between different AS's.

6. ATOMIC_AGGREGATE; a well-known discretionary attribute, which is used to inform that automatic aggregation has occurred and at least some AS_PATH information was excluded from the Update message.

7. AGGREGATOR; an optional, transitive attribute, used to convey the BGP identifier ("IP address") and AS-number of the BGP speaker which has performed automatic aggregation.

More path attributes have since then been specified; an important and commonly used one is COMMUNITIES [7]; an optional, transitive attribute. It can be used to mark any route for special handling. There are three special well-known communities: NO_EXPORT, NO_ADVERTISE and NO_EXPORT_SUBCONFED. The first restricts the advertisement of the route to one AS or BGP confederation; the second forbids re-advertisement even to the local internal BGP peers; the third forbids the re-advertisement of the route to any BGP speaker outside of the local AS, whether in confederation or not.

So, to summarize using a fabricated example in one BGP implementation where all the described attributes are present:

```
62.71.0.0/16 (3 entries, 1 announced)
     Nexthop: 193.64.136.33
     MED: 1000
     Localpref: 200
     AS path: 1759 5515 I <Atomic>
     Aggregator: 5515 194.89.196.253
     Communities: 790:51 790:61 790:335
```

So, every route may contain:

1. the IP address prefix and the prefix length (62.71.0.0/16) (required)

2. the AS path of the route (1759 5515) (required)

3. the next-hop address; where this route was learned from (193.64.136.33) (required)

4. how the route was originated – IGP, EGP or unknown (I) (required)
5. whether multi-exit discriminator (MED) attribute had been added by the neighbor (1000)
6. whether local-preference has been configured in the local AS (200)
7. whether the route was automatically aggregated en route (<Atomic>)
8. ..and if so, the aggregator's AS number and IP address (5515 194.89.196.253)
9. whether any Communities have been included in the route (790:51 790:61 790:335)

Best Path Selection

BGP best path selection algorithm has an important role in BGP multihoming, so it is briefly introduced here. Several implementations offer configuration options to modify the algorithm, but those are not listed here. Also, some extensions like route reflectors and confederations modify the list slightly but those are not listed either for simplicity.

It should be noted that the best path selection algorithm only chooses among prefixes of equal length: it is a common practice to effectively affect the forwarding decision by using more or less specific routes.

Another thing to note is that in a system where multiple routing protocols are used, the implementations set "distances" for each routing protocol. The route with a lower distance is considered better than the same route of a higher distance. For example, IGP's have lower distances than BGP.

For a route to be considered eligible in the path selection process, its next-hop must always be valid: in practice, this often means it is present in IGP. Also, if a route received from an external peer already includes the AS of the external peer, signifying a certain routing loop, it is silently ignored.

The list of equally good paths are compared two-at-a-time until a tie-breaker is found, in order:

1. prefer the path with the highest LOCAL_PREF.
2. prefer the path with the shortest AS_PATH.
3. prefer the path with lowest origin type: IGP $<$ EGP $<$ INCOMPLETE.
4. prefer the path with lowest multi-exit discriminator, only applicable in comparisons between two paths from the same neighboring AS, otherwise ignored. If missing, MED is considered to be 0.
5. prefer the path learned using external BGP to those learned using internal BGP.

6. prefer the path with the lowest IGP metric for the NEXT_HOP.
7. prefer the path with the lowest BGP identifier ("IP address").
8. prefer the path whose neighbor IP address where the route was learned is the lowest.

Some Operational Procedures

There are a lot of operational procedures and practices to the use of BGP, but two of them require a special mention as they are used later in the thesis.

First, there are basically two types of AS numbers: public and private, as with IPv4 addresses. Public addresses are applied for from the registries, and private ones can be used at will. Indeed, in many cases, private AS numbers suffice very well, for example in multi-connecting scenarios; the only thing you have to be careful of is to remove the private numbers from the AS-path if the routes are propagated outside of the area where private AS numbers are used.

Second, so called "AS-path prepending" is the process of artificially lengthening the AS-path by repeating an AS in the path multiple times as a single chain of the same AS; this is typically done to affect the best path selection process.

3.1.2 Route Aggregation

In IPv4, the routing table could be, in the worst case, $\sum_{i=0}^{32} 2^i$ (about $8.6 * 10^9$) entries big. However, 2^{32} would also be sufficient. Calculation with IPv6 would give even more overwhelming results.

So, it is clear that the number of entries in the routing table must be severely restricted.

This is done by advertising network prefixes rather than host addresses; a prefix signifies multiple addresses in the blocks of 2^{32-n} addresses, where n is the prefix length, and 32 used with IPv4. The notation is "prefix/prefix length", also known as the Classless Interdomain Routing (CIDR) notation.

For example, consider an ISP which has an address block of 193.166.0.0/16. Assume it has 100 customers, each of which have been given an address block of a /24, like 193.166.0.0/24, 193.166.1.0/24, etc. Now the ISP could advertise all of these to the rest of the Internet, but it really should perform route aggregation: as all of these customers are reachable through the same ISP, it is enough to advertise only the superblock, 193.166.0.0/16: the internal routing tables of the ISP have to include all the entries, though. In this way, Internet routing can be made a "divide and conquer" -problem, as some complexity can be split off from the required data set of the global routing table.

When someone advertises additional routes, like /24's here, in addition to the nec-

essary one(s), the extra ones are commonly called "more specific routes"; the superblock(s) may also then be called a "less specific route" or a "root prefix".

Despite common interests to keep Internet clean of redundant routes, a lot of completely unnecessary more specific routes are still being advertised. Usually this is caused by configuration mistakes or lack of knowledge on how to aggregate routes.

However, there are some cases where the use of more specifics may be desirable at least for some purposes, like multihoming or traffic engineering. This is described in detail later, in section 4.3.

3.1.3 IP Address Formats

IPv4

Originally, IPv4 addressing was classful: the network and host part of an address was determined by looking at the address itself.

Since then, the addressing has become classless: addresses, wherever they come from, are paired with prefix length which tells how many bits from the beginning of the address correspond to the network part. This is called Classless Interdomain Routing (CIDR) [8]. No relevant technologies exist anymore which would not work with variable-length prefixes.

IPv6

In the beginning of IPv6 specification effort, address hierarchy was based on several levels of aggregation [9]. The proposal is shown below.

3	13	8	24	16	64 bits
FP	TLA ID	RES	NLA ID	SLA ID	Interface ID

Here,

- FP = Format Prefix, 001 for unicast addresses
- TLA ID = Top-Level Aggregation Identifier
- RES = Reserved for future use
- NLA ID = Next-Level Aggregation Identifier
- SLA ID = Site-Level Aggregation Identifier
- Interface ID = Interface Identifier

However, this proposal has since then been rejected; it was realized that the address format is not the right place to enforce soft limits usually caused by non-technical

reasons. In particular, this model restricted the number of "upstream-provider - independent ISPs" to 8192; this was a noble and idealistic effort to solve routing table growth problems, but disconnected from the operational realities.

Currently the address format is much more flexible and pragmatic:

n bits	m bits	128-n-m bits
global routing prefix	subnet ID	interface ID

Here, subnet ID is recommended to be 16 bits and interface ID 64 bits, leaving first 48 bits for global routing prefix – but this is not enforced in the address architecture.

3.1.4 IP Address Allocation and Assignment

The Procedure

When considering which prefixes should be routable in the Internet, one should take a look at how IP addresses are distributed. The process is the same with IPv4 and IPv6.

Internet Architecture Board (IAB) delegated all the currently used address space to Internet Assigned Numbers Authority (IANA). Practically, IANA is the head of the address allocation chain. It allocates addresses to Regional Internet Registries (RIR) such as RIPE NCC (Europe), ARIN (North America), APNIC (Asia-Pacific), LACNIC (Latin America); these allocations are usually done when needed, e.g. once, twice a year. A typical IPv4 allocation size from IANA is one /8 at a time.

RIR's further sub-allocate their address blocks to Local Internet Registries (LIR) or in APNIC, to a country-level registry and then LIR's. LIR's are typically big ISPs or enterprises operating in a country or even internationally. There are member fees to attain LIR membership. The sizes of IPv4 allocations have typically ranged from /16 to /20, even less. IPv6 allocation size has been about /32 out of 128 bits.

When addresses have been used up to the sufficient degree – one such yardstick being the H-ratio [10] – another allocation can be made: in IPv4, the focus has been more on the preservation of the address space and conservative allocations than aggregation – that is, giving enough addresses at once. The IPv6 policy, as there are more addresses available, is more focused on aggregation. That is, the goal of IPv6 policies is that an ISPs addresses could be advertised with just one big route, not as dozens or hundreds, or even more, smaller routes as with IPv4.

LIR's assign blocks of addresses to their customers on request. The customers are typically either enterprises or small ISPs. Home users are not typically assigned addresses, rather, the assignment is done to their ISP, and they're only allowed to use the addresses.

Note that address allocation to end-users is always called "assignment", not allocation.

Implications

Address allocation is independent of routability in the sense that even if you get a block of addresses from a RIR, no stance whatsoever is taken whether it is actually routable and what one might have to pay for that service.

Nevertheless, it is useful to consider address allocation in more detail. Clearly, address allocations obtained directly from RIR's seem to have some form of implied provider-independence in them. However, this encourages those parties that do not offer IP address services to other organizations, typically enterprises, to join as LIR's, to obtain these portable IP addresses.

On the other hand, sub-allocations or assignments made to e.g. enterprises from LIR's do not have this property: if enterprise changes ISPs, switching from one LIR to another, it's much more understandable if renumbering IP addresses would be required.

Address space portability and routability will be a major issue, and is an important factor on which mechanisms are applicable, as will be seen.

3.1.5 RPSL and Internet Route Registry

It has been deemed important to be able to specify one's routing policy using a well-defined language; for this purpose Routing Policy Specification Language (RPSL) [11] has been developed.

The use of RPSL and publishing one's policy gives more control to the operators to act on others' policies. RPSL can also be used to add safeguards to protect from out-of-policy advertisements from neighbor sites, for example configuration mistakes which could otherwise disrupt communications.

The policy can be published in one or more Internet Route Registries (IRR) [12]. Examples of these route registries are the Routing Assets Database (RADB) [13] and the RIPE database [14]. The policy can be retrieved and processed manually or using automated tools like IRRToolSet [15]. Automated tools can even generate routing policy configuration directly usable on many router platforms.

When obtaining an IP address block from a registry, the block is recorded in the database and included in a routing policy. It is a common practice to upkeep at least some form of policy in the registry.

In theory, all the prefixes advertised by anyone should be available in the registry and thus could be useful when examining multihoming; in practice, there are gaps and misinformation in the registries which make it a bit unreliable, at least as an exclusive source of information.

3.1.6 Network Address Translation

Network Address Translation (NAT) [16] is an old mechanism that was invented when the exhaustion of IPv4 addresses seemed to be inevitable.

There are two operational modes to NAT: basic NAT and Network Address Port Translation (NAPT).

In basic NAT, IP addresses are mapped from one address to another, in one-to-one fashion. This is typically done at some edge router(s).

In NAPT, multiple network addresses and their TCP/UDP ports are mapped to one or more "external" IP addresses and TCP/UDP ports. A NAPT device could be interpreted to act as a multiplexer for the external address(es), for example to circumvent issues with address allocation and assignment, as described in section 3.1.4. One-to-one mapping is generally not possible.

Together, these mechanisms provide a way to e.g. interconnect private networks with private addresses [17] to the global Internet using a public IP address. This mechanism is referred as "traditional NAT" or simply as NAT.

Several ways how NAT's could help in multihoming scenarios have been proposed; these are described later, in section 4.3.2.

3.2 Data Collection Environment

Having introduced the basic multihoming components, it's now possible to describe the environment where data collection, used later in this thesis is done.

In short, route advertisements in peerings over FICIX internet exchange are being monitored at the Funet connection.

First, some background is given on Funet and FICIX – where the collection is performed. Then, the particular environment is described at a bit more length. Last, special characteristics of the environment are described; that is, why the chosen place was ideal for this kind of data collection.

Actual data collected and processed is described in the next sections; analysis is done in section 5.3.

3.2.1 Background on Funet

Finnish University Network (Funet) [18] was founded in 1984 to interconnect Finnish universities and develop the research network. Connectivity to abroad became available in 1985, and direct connectivity to NSFnet, the predecessor of Internet, was established through NORDUnet in 1988.

At the end of 2002, the Funet network consisted of about 80 research organizations, all universities and most similar higher level education institutions but also some

non-commercial research organizations.

Network connectivity to the Internet is still provided by the Nordic research network, NORDUnet.

3.2.2 Background on FICIX

Finnish Communication and Internet Exchange (FICIX) is currently the only internet exchange in Finland [19]. It was founded in 1993 and now includes two separate exchanges for redundancy; there were about 15 members at the end of 2002.

The organization is a non-profit association, and the exchange has no direct carriers or co-location space. Full members are required to provide a significant amount of access service in Finland and must peer with everyone at least to the extent of Finnish networks free of charge. Peering or transit of other networks is up to bilateral arrangements.

FICIX is and has always been a "layer 2" exchange; peering sessions are established to the peer routers, not for example any neutral routers containing all the routes ("route collectors").

FICIX is the place where all Finnish Internet traffic is exchanged. Other than Finnish networks are currently also advertised; these had previously been forbidden. These are typically from Nordic regions, by operators who also have presence there.

3.2.3 Overview of the Environment

Route advertisements are monitored from all peers, both in FICIX1 and FICIX2. FICIX members typically have a similar routing policy for both. Due to recent changes in equipment of FICIX1, only FICIX2 will be analyzed. The Funet connections to FICIX are shown in figure 3.1. Funet has typically two connections to other FICIX members, across FICIX1 and FICIX2. Physical links from two separate Funet routers go to the FICIX infrastructure; the physical topology is star-like. However, BGP sessions are established directly to the neighbor organizations. The Funet routers which connect to FICIX also have connectivity to the rest of the Internet, via NORDUnet, and naturally the rest of the Funet network.

The peers, 15 at that time, at FICIX2 in the analysis are, sorted by the AS number:

- 719 (Elisa Solutions Oy)
- 790 (EUnet Finland)
- 1342 (Fujitsu Invia Oy)
- 1759 (Sonera Carrier Networks Ltd)
- 2686 (AT&T Global Services Oy)

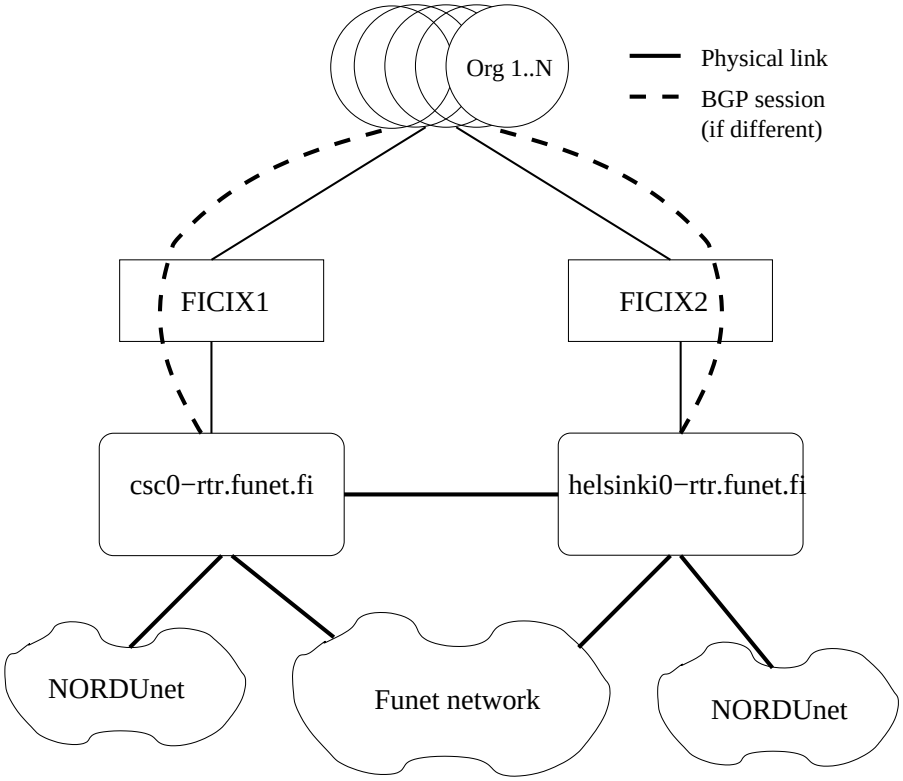

Figure 3.1: Funet and FICIX environment

- 3246 (Song Networks Oy)
- 6667 (Jippii Group)
- 6743 (Equant Finland Oy)
- 8434 (Utfors Oy)
- 9060 (BT Ignite Nordics Ltd)
- 16023 (Netsonic)
- 16086 (Finnet Group)
- 20542 (Helsinki Television Company)
- 20569 (RSL Com Finland)
- 24751 (Multi.fi)

3.2.4 Special Characteristics

By the definition of BGP, only best paths can be advertised to peers. Therefore, to be able to get a full view of what's being advertised in FICIX, you have to be directly connected to all FICIX members. For example, collection of routing data from what a FICIX member's customer might see, as a full routing table, would be close to useless.

In consequence, when monitoring the FICIX route advertisements, one can see the best paths of every other FICIX member, not only the best paths of everything in FICIX. There is a significant difference as there is an overlap between route advertisements by different FICIX members.

Therefore, this specific environment is best suited for observing route advertisements of Finnish networks.

3.3 Data Collection

In the previous section, the data collection environment was described. Now it's time to introduce what kind of data is collected and how; this is a rather simple operation.

An advertised prefix may contain a lot of attributes, some of which are required, as described in section 3.1.1.

To summarize, all the possible attributes – ignoring more recent and less used new ones – are listed in a fabricated example below:

```
62.71.0.0/16 (3 entries, 1 announced)
     Nexthop: 193.64.136.33
     MED: 1000
     AS path: 1759 5515 I <Atomic>
     Aggregator: 5515 194.89.196.253
     Communities: 790:51 790:61 790:335
```

Data, all the received route advertisements in the most extensive format, is collected by a script which is run on the two routers handling FICIX traffic once a week. The collection is done prior to any policy configuration, such as rejecting some routes or modifying the attributes. The data is stored in the routers' hard drive in a compressed format. From there, the data is manually copied on a workstation for analysis.

Route advertisements from the Internet, from NORDUnet, are also stored for reference, but those are not further processed or analyzed.

3.4 Data Processing

Now, as the environment and the collection process is clear, the processing of data is described in a lot of detail. The processing includes some minor formatting, reducing the amount of data by cutting out irrelevant parts, and describing which kind of analysis on AS numbers and prefixes is done with the results.

The actual results of the processing are described later, in section 5.3.

3.4.1 First Processing

The route advertisement data is processed with a short UNIX shell script which does a few tasks on its own and calls other analyzing scripts as needed.

First, all the FICIX route advertisements are uncompressed and concatenated to a single file; if Internet route advertisements, from NORDUnet, are still present, they are ignored.

Next, the concatenated route file is slightly modified: in the specific router software version running on the routers, there exists an output formatting bug, and a few changes are automatically applied to the file, changing it to the format it should really look like.

3.4.2 Cleaning Up the Data

Now, the extensive route advertisement data can be processed; the output is tailed down a bit by leaving out the specific communities used (leaving the number of communities instead), removing the Aggregator IP address and the information about which routing protocol the route originated from. So, the example below is summarized to the following:

```
Prefix             Nexthop         MED   Aggr? Communities AS path
62.71.0.0/16      |193.64.136.33  |1000 |5515 |3          |1759 5515
```

At this stage, general information about prefixes is also printed out:

- the total number of prefixes,
- the number and percentage of prefixes with community attributes,
- the number and percentage of prefixes with the atomic aggregate attribute, and
- the number and percentage of prefixes with the MED attribute.

Now, after the data has been stored in a relatively concise format, it is analyzed in steps.

3.4.3 Analyzing the AS Paths

First, AS paths are analyzed with a short and simple script written in the Perl programming language. AS numbers in advertised routes are classified in three categories:

- neighbors, ie. FICIX members, at the beginning of AS paths,
- transits, ie. AS's which are seen at the middle of AS paths, and
- origins, ie AS's which are seen at the end of AS paths.

For each of these, the number of such AS's, and AS numbers themselves, are listed.

After the initial analysis, these results are further analyzed, in three aspects:

- AS's which are both neighbors and are being transited,
- AS's which do not originate anything, and
- AS's which perform AS path prepending.

For each of these, as above, the number of such AS's, and AS numbers themselves, are listed.

The first item is of particular interest; practically that means that the AS in question is doing ISP multihoming inside FICIX, as described in section 5.1.4.

3.4.4 Analyzing Prefixes

Then, a more complex script, also written in Perl, analyzes the prefix information.

The script prints out prefixes and statistics based on four main classifications:

- clearly multihomed prefixes (section 5.1.1)
- identical prefixes from different paths and origin (section 5.1.2)
- more specific routes from different origin (section 5.1.2)
- more specific routes from the same origin (section 5.1.3)

In addition, some other data is also printed out:

- a general hierarchy of more and less specific routes,
- single routes with AS-path prepending, and
- routes which were aggregated by a private AS number.

The statistics printed out include:

- the number of prefixes of different length and their respective percentages,
- the number and the list of uniquely multihomed AS numbers, and
- the percentage of which types are included in the more specifics.

Also, there are a lot of other numbers and percentages listed out; these include:

- unique prefixes advertised via more than one path,
- extra prefixes related to the above, i.e. the number of prefixes that are more than dual-homed,
- more specific routes,
- routes which are related to both more and less specific routes,
- less specific routes,
- more specific routes with a different origin, and
- more specific routes with the same path and origin.

Chapter 4

Site Multihoming

In the previous chapters, scope, motivations, background, data collection, and processing were introduced. This chapter describes issues specific to site multihoming, to give perspective before going into analyzing current IPv4 multihoming practices or analyzing IPv6 multihoming mechanisms.

This thesis focuses on site multihoming, as part of network multihoming. This is due to the other part, ISP multihoming, being a very simple operational practice, contrary to issues related to site multihoming, and is therefore out of scope.

First, the generic scalability problem of site multihoming is presented. Then, some constraints and factors for site multihoming are described. Next, the two typical site multihoming techniques with IPv4 are presented. Then, a large number of proposed IPv6 multihoming methods are introduced. Last, multi-connecting is described.

4.1 The Generic Scalability Problem

In this section, the potential scalability problem if every site was present in the global routing table is described. This is done to further justify that current mechanisms do not scale up for end-sites of all size: some other methods will also be required.

At the moment, the majority of sites do connect to the Internet without multihoming mechanisms; here we consider the theoretical future case of all sites being provider independent.

First, some metrics and background are presented. Then, a few simple mathematic formulas based on the metrics are derived. Last, these formulas are tested with some example values to see which kind of results they yield.

4.1.1 Background and Metrics

At the end of 2002, the global routing table was roughly 115 000 entries in size.

If all end-sites wanted to multihome like ISPs, having their routes present in the global routing table, this could have devastating results. Unfortunately, even though BGP scalability has been analyzed and written about, no papers seem to have considered these long-term dynamics.

When considering the scalability problems of the global routing table, analysis can be done with three metrics [20] for routers:

- processing power
- memory
- network capacity for transmitting BGP messages

All of the above metrics can be evaluated using two criteria: the absolute number of prefixes, and the changes in the prefixes. The latter is particularly interesting as instabilities mixed with a large number of prefixes are assumed to cause a large number of changes.

It's assumed that network capacity is not a major factor in the current backbone networks, even though updates can be very big especially with a lot of changes, and the convergence could be enhanced by making the updates transmit faster.

4.1.2 Mathematical Estimates

The number of routes in the global routing table at any one time is:

$$N_{routes} = \sum_{i=1}^{i=C} \sum_{j=1}^{j=N_i} p_{i,j}$$

Where C is the number of countries, N_i the number of end-sites in the country i and $p_{i,j}$ the number of prefixes in end-site j in country i.

Now, by substituting the first sum by estimate c for countries and the second sum with estimate n for end-sites in a country with p prefixes each, we get:

$$N_{routes} \approx c * n * p$$

This simplified estimate will be used in subsequent formulas.

In particular, note that the number of operators the end-site is multihomed to is not seen in the equation. This is due to the BGP best path selection process: only one identical prefix is observable at one time.

The number of end-sites has also been semi-artificially separated per country to make it easier to make estimates; Internet is assumed to be quite a heterogeneous environment, consisting of different types of networks, but, here, it is simplified to a homogeneous model.

Now, let's consider another metric, the change rate; as above, the following variables are defined:

- n end-sites with p prefixes each in c countries
- M the number of ISPs operating in a country
- M_t the number of transit ISPs operating globally
- P_n failure probability of an end-site, per day
- P_M failure probability of an ISP, per day
- P_{M_t} failure probability of a transit ISP, per day

Now, the amount of changes to the global routing table seems to be roughly:

$$\begin{aligned} N_{changes/d} &= 2*c*p*n*P_n + 2*p*n*\frac{1}{M}*P_M + 2*c*p*n*\frac{1}{M_t}*P_{M_t} \\ &= 2*c*p*n*(P_n + \frac{1}{c*M}*P_M + \frac{1}{M_t}*P_{M_t}) \end{aligned}$$

The equation consists of three terms: failures affecting the end-site, failures affecting the end-site's ISP and failures affecting the transit ISP. The 2 in the equations is caused by the fact that both the withdrawal, whether implicit or explicit, and re-advertisement must be processed.

In above, a number of simplifying assumptions have been made; it is assumed that more than one ISP failing at once is uncommon enough and insignificant as a second-order term. It is also assumed that the end-sites in a country are equally distributed to M ISPs; if, in fact, the number of significant ISPs is smaller, their reliability becomes even more important. The same applies to M_t: ISPs are assumed to be equally distributed to their international upstream transit ISPs.

In here, the "failure" of a transit, ISP, or end-site is defined as any event that will trigger a change in the global route advertisements. Typically this is a result of switching to another upstream service provider, causing a change in the AS path attribute, at the very least. Typically, multi-connecting, as described in section 4.5, does not cause such changes: these changes are typically only local to the connecting ISP.

As another way to try to be able to estimate the effects of a failure, the number of changes which will happen in the case of the inevitable failure is analyzed: not considering the probability factor at all; this is:

$$N_{changes/failure} = \begin{cases} 2 & \text{end-site} \\ \frac{2*p*n}{M} & \text{ISP} \\ \frac{2*p*c*n}{M_t} & \text{transit ISP} \end{cases}$$

To conclude, routers should be capable of handling significantly more than $N_{changes/d}$ changes per day, N_{routes} routing table entries and $N_{changes/failure}$ rapid changes when inevitably a failure occurs, the half of them happening typically simultaneously.

4.1.3 Numerical Estimates and Analysis

Now, let us consider the variables above, and assign some values to them. The values are a result of very rough estimates, and if the work in the thesis would base on them, they would have to be justified in a much more scientific manner. However, as an independent data point, this approach to gain at least some perspective seems sufficient.

- $n = 5000$ end-sites with $p = 5$ prefixes each in $c = 100$ countries
- $M = 10$ ISPs operating in a country
- $M_t = 100$ transit ISPs operating globally
- $P_n = 0.02$, failure probability of end-sites, per day
- $P_M = 0.01$, failure probability of the ISP, per day
- $P_{M_t} = 0.005$, failure probability of a transit ISP, per day

These values seem rather conservative – the reality, at least in the near future, is probably different.

To take a few examples, 5000 end-sites per 100 most-developed countries is a very conservative number; but that will be examined. It is expected that every 50 days an end-site suffers a failure for one reason or another; ISP every 100 days and transit ISPs every 200 days. For example, a software or hardware upgrade of a router is one possible source of such failures.

With these we have:

$$N_{routes} = c * p * n = 100 * 5 * 5000 = 2500000$$

and:

$$\begin{aligned} N_{changes/d} &= 2 * c * p * n * (P_n + \frac{1}{c * M} * P_M + \frac{1}{M_t} * P_{M_t}) \\ &= 5000000 * (0.02 + 0.00001 + 0.00005) \\ &= 100300 \end{aligned}$$

So, with these conservative values, we'd get 2.5 million routes in the routing table, with 0.1 million critical changes requiring computation every day. A notable thing in the amount of changes is the strong emphasis of the end-sites in the final number.

Now, let us consider the number of changes in the inevitable event of failure with the above values:

$$N_{changes/failure} = \begin{cases} 2 & \text{end-site} \\ 5000 & \text{ISP} \\ 50000 & \text{transit ISP} \end{cases}$$

To get a more qualitative grasp of the number of end-sites per country, two different statistics have been examined:

- in Finland, with population of about 5 million, there were 266 enterprises of over 500 employees and about 2800 enterprises of at least 50 employees in 2000 (53 and 560 per million people, respectively) [21]
- in the USA, with population of about 280 million, there were 17 153 enterprises of over 500 employees and about 1000000 enterprises of at least 20 employees in 2000, with an estimate I've had to make for 50 employees being around 650000 (61, 3570, and 2300 per million people, respectively) [22]

So, if only enterprises with at least 500 employees would need a multihoming solution, this would result in at least in the order of 100,000 end-sites even if developing countries were not counted in. On the other hand, if we assume that enterprises with at least 50 employees would also require one, this would result in at least in the order of several million end-sites; so the estimates seem at least roughly reasonable.

4.2 Constraints in IPv4 and IPv6

In this section, some generic constraints and factors to be considered for a solution are presented; these aspects include prefix length filtering, address space size, aggregation, handling of multiple addresses, and network address translation.

4.2.1 Prefix Length Filters

ISPs typically restrict which kind of prefixes they accept from their BGP neighbors.

A typical approach is to enable a BGP policy, a "filter", which discards all advertisements which are considered to be "too specific". For example, one such policy would be to ignore all IPv4 advertisements of the length longer than /24.

An additional, possible approach is to generate automatically prefix lists from route registries, as described in 3.1.5, and accept only these routes.

In IPv6, there are no protocol-specific architectural restrictions on the advertisement length. However, the common policy adopted by practically everyone, at the end of 2002, was to disallow any more specific advertisements: those that could not have been allocated within RIR allocation borders, as noted in section 3.1.4: the limit, in 2002, was /35.

As even large IPv6 end-sites can very well cope with a /48, but not with an IPv4 /24, one can deduce that in practice this leads to allowing only ISPs to announce their IPv6 address aggregates: all end-site advertisements are discarded. This is the intended outcome.

In consequence, IPv4, but even more so IPv6, policies set restrictions on multihoming models which could be applied. Of course, given community consensus, such policies could be changed if needed.

4.2.2 Scalability Problems with More Addresses

Even though the IPv4 address range has enough prefixes to exceed the capacity of current routers if unfiltered, as noted in section 3.1.2, the situation with IPv6 is even more grave.

As the address space grows, the possibility for advertising even more addresses also grows; for example, a misconfiguration could easily result in advertising an enormous amount of more specific routes – a lot more of them than with IPv4.

With the growth of the address space, the possibilities for intentional advertisements may also feel enticing. For example, if every IPv6 /48 site, like all enterprises and home users, could have addresses advertised to the global routing table, the routing table would grow enormously, beyond the capacity of the routers to handle it.

Therefore, prefix length filters will become even more important in the future, and multihoming mechanisms should be designed to cope with that.

4.2.3 Aggregation

With IPv4 address allocation and assignment policies, as noted in section 3.1.4, the focus has been more on address space preservation than aggregation. In consequence, the prefixes advertised by an ISP can not typically be highly aggregated, for example to one route entry. Also, as large enterprises may require a high number of addresses, even larger than many ISPs, it is not easy to distinguish ISPs and end-sites by syntactically examining an advertised prefix.

With IPv6 policies, currently the ISPs get a relatively large block of addresses, typically of about the same size each. Almost every site gets the same amount of addresses, one /48: this is enough, as it enables 2^{16} subnets with 2^{64} nodes each.

Therefore, it is much easier to distinguish ISPs and sites in IPv6: the latter is one /48, no matter the size of the site. This results in the ability to require the aggregation at the ISP level using prefix length filters, only allowing the ISP aggregates, except by explicit bilateral agreement.

4.2.4 Handling Multiple Addresses

From the start, IPv6 has been designed to be able to have multiple addresses on an interface; of course, this is also possible on most IPv4 nodes today, but it is not a common practice.

Indeed, many IPv6 site multihoming solutions are based on the fact that different

addresses from multiple providers are configured on each node, typically using the stateless address autoconfiguration. This way, one can hope to try to solve the site multihoming problem in the sites, not in the routing system.

Such site multihoming mechanisms could be applied to IPv4 as well, but this has been considered unnecessary, as other mechanisms are also possible.

4.2.5 Network Address Translation

In IPv4, NAT, as described in section 3.1.6, is widely used, despite its drawbacks.

In IPv6, there has been a very strong desire to avoid NAT as it breaks several key assumptions of the Internet architecture. Therefore, such an approach for IPv6 multihoming is not possible or desirable. At most, one could consider entirely different approaches to the renumbering problem NAT can be used to solve. One related approach is introduced briefly.

4.3 Overview of Different Mechanisms for Site Multihoming in IPv4

Having described the general scalability problem and constraints, now it's time to consider site multihoming mechanisms for IPv4; IPv6 and multi-connecting will be described in the next sections.

Many different multihoming mechanisms, like those described with IPv6 in section 4.4, are also applicable to IPv4. However, it seems such mechanisms aren't commonplace; therefore, only the two most relevant techniques currently in use with IPv4 are described.

4.3.1 Site Multihoming with BGP

Site multihoming with BGP consists of a few main tasks:

- obtaining an address space,
- obtaining an autonomous system number, or negotiating the use of a private one with ISPs,
- obtaining physical connectivity to multiple ISPs,
- connecting ISPs to separate routers at the site, and
- establishing BGP sessions to these ISPs and advertising the address space.

One way to obtain the address space is to apply for it directly for this specific, provider independent purpose. In practise, this typically means joining a RIR as a LIR.

The other way, which is frowned upon and considered a bad practice, is to take a part of the existing address space of one ISP. This often happens when a site has been connected to only one ISP before, and then wishes to simultaneously connect to another; otherwise, the networks would have to be, typically, completely renumbered.

BGP requires an AS number; it is much less of a problem to apply for from a RIR than getting address space, but the use of private AS numbers is also possible. If private AS numbers are used, one has to negotiate which one to pick with the ISP(s) so that nobody else of ISP's customers is using that AS.

To increase redundancy and minimize single points of failure, physical connectivity must be obtained to both ISPs. These should be connected to multiple routers at the site: one router failing must not bring all the connectivity to a halt.

Finally, BGP sessions must be established on the routers to the multiple ISPs. This requires some configuration; the most important part is to advertise the address space, obtained as described above, to the ISPs, and from there to the Internet.

There are some slightly varying methods used in different BGP techniques, as described in sections 5.1.1 and 5.1.2.

In short, if the address space is truly provider independent, so that there is no overlapping, the address prefix is advertised identically to all of the multiple ISPs, and from there it will be advertised to the Internet. This leads to inbound traffic being load-balanced between these ISPs; this case is shown in section 5.1.1.

However, if the address space is only part of an existing prefix, this advertisement will typically not be re-advertised by the ISP which owns the larger part of the address space; in this case, almost all the inbound traffic will come through the secondary ISP, where the address space was not derived from. As described in section 3.1.1, this is due to the best path algorithm always choosing the more specific route; this case is shown in section 5.1.2.

4.3.2 Site Multihoming with NAT

Network Address Translation, as briefly described in section 3.1.6, has been proposed to help in multihoming in several ways.

Cisco IOS provides some functions [23] but the particular solution is relatively complex to set up. The solution expands on a slightly simpler approach [24] which does not require NAT, and which will also be described with IPv6 solutions in section 4.4.4.

Some other products exist as well that claim to solve the multihoming problem easily and without BGP [25, 26].

These mechanisms typically depend on performing NAT to several public IP addresses obtained from different ISPs and modifying DNS lookup requests and replies; this is a rather complex operation. Then load-balancing and some form of resiliency can be made to work using those addresses.

The main problem with these kind of DNS and/or NAT mechanisms is that if one path fails, existing long-lived TCP connections will break, typically after a timeout.

"Intercepting" and continuing the connections is not possible: in the case of NAT, the sessions will stay alive up to the NAT device, but an outgoing public IP address cannot be replaced without breaking the session, as the IP addresses are specific to the ISP.

Therefore, these kind of mechanisms are heavily geared towards very short-lived connections, e.g. Web surfing/serving where the loss of existing connections is not a major problem. This is a subset of usage scenarios for connection-oriented transport protocols: for example, long-lasting remote terminal sessions or the transfer of huge files might be severely disrupted if connectivity breaks.

Of course, with solutions that are based on NAT, one should always remember that more complex application protocols require explicit support for remapping the addresses and/or ports.

4.4 Overview of Different Mechanisms for Site Multihoming in IPv6

In this section, the most interesting IPv6 network multihoming mechanisms are described; [27] lists some more.

Some very research-oriented or raw ideas, such as the requirements for the new Internet routing protocol being able to handle multihoming [28] are not covered here, either.

Note that "site multihoming with BGP" is not currently possible with IPv6, due to the operational restrictions on the advertised prefix lengths; NAT is also not possible.

Mechanisms are only presented here, and will be analyzed in detail in section 6.1.

4.4.1 Transport Solutions

Transport solutions focus on introducing some features in the most prominent transport-layer protocols, such as TCP.

TCP Modifications

A TCP connection is always established between two end-points, using one IP address for each, even if the nodes would have multiple addresses.

If the network connectivity using the addresses between the end-points is lost, the connection will be rendered unusable for the duration of the loss, or times out and is terminated if the loss lasts long enough.

For the purposes of multihoming with multiple addresses per node, and renumbering purposes, some extensions for TCP have been proposed.

One such proposal [29] gives additional options to be used when initiating the connection with the three-way handshake: PREFIXES option would be sent with TCP SYN and TCP SYN/ACK, and PREFIXES_ACK in response alongside with TCP SYN/ACK and ACK. In this way, when setting up the connection, both end-points would agree on the alternative addresses which could be used to deliver packets which should be considered to be part of the TCP connection.

SCTP

Stream Control Transmission Protocol (SCTP) [30] is a new, reliable connection-oriented protocol. The main differences from TCP are that it is possible to accept non-ordered delivery of packets in a stream, and that every SCTP association can have multiple end-points.

When the connection is established, multiple addresses can be listed as end-points for the node. During the connection, each of these possible alternative paths are probed with heartbeat packets, to see that they're operational. So, if one network path fails, SCTP can switch to another end-to-end path, if set up, when it notices the problem.

This enables the existing connections to be kept alive even when the addresses primarily used change. Some issues of multihoming with SCTP are discussed in [31].

4.4.2 Identifier and Locator Separation

IP addresses include two entirely separate functions: the address or the locator of the node, "how to get there", and the identifier of the end-point, "the name". The semantics of IP address have been overloaded to include both.

The former is used to forward packets in routers, mostly, and the latter in the protocols like TCP and UDP to identify the end-points of connections or packets.

It is understandable that these have been bundled in one: in simple nodes with only one stable address, these are the one and the same; indeed, this was the initial Internet architecture. However, when a node has multiple addresses with different network connectivity properties, this becomes an extremely challenging problem. The protocols in a node should be able to process packets belonging to a connection, regardless of how they reached the destination, and vice versa.

The problem has been extensively researched, for example by Internet Research Task Force Namespace Research Group [32], but the problem still remains on the table. Solutions have been presented, of course, but the separation would require a radical change in thinking, and securing the mapping between locators and identifiers is a very challenging problem.

Transport solutions, above, try to tackle this issue on per-protocol basis, while the following three approaches do it for all the protocols.

HIP

Host Identity Payload and Protocol (HIP) [4] is a proposal to separate locators and identifiers. The proposal is to create a virtual "Host Identity" -layer between the internetwork and transport layers.

The idea is that all packets themselves contain the locator, as before, but the end-nodes use the addresses of the Host Identities at the transport layer. In consequence, the addresses used for packets can change freely without disrupting the sessions which are bound to Host Identities.

The fundamental design feature behind HIP is security; the protocol includes a four-way handshake, where the parties negotiate keying material needed for enabling full use of end-to-end encryption. The Host Identity is proposed to be a hash of a public key signature.

The mapping between Host Identities and IP addresses can be done using DNS.

LIN6

Location Independent Addressing for IPv6 (LIN6) [33] separates the locator and the identifier by splitting the 128 bit address into two, the first part containing the locator, and the last part the identifier.

The proposal works so that when the transport layer (and higher) protocols use addresses, the first 64 bits of an address are substituted by a static, pre-defined value, "LIN6 prefix", forming the "LIN6 generalized ID". The last 64 bits consist of a globally unique "LIN6-ID", which is typically derived from the MAC address of an Ethernet network adapter. The routing and forwarding of packets use the regular representation of the address.

In this way, a mapping function, also similar to DNS, is used for nodes to be able to find out the locator address corresponding to an address consisting of the static LIN6 prefix and the global "LIN6-ID".

As the addresses used by transport protocols do not have any knowledge of the network location, nodes can move and be multihomed without problems – when network connectivity to one address is lost, it is possible to seamlessly start using the other, with no effect to the transport protocols.

Mobile IPv6

Mobile IPv6 [3] introduces the concepts of Home Address and Care-of Address; a mobile node is always expected to be reachable at its Home Address, even though

Care-of Addresses used may change at any time. The main goal of Mobile IP is to provide non-breaking connections when moving between different network segments.

Sometimes, Mobile IPv6 or some of its features, typically an extended Binding Update, has been proposed as one solution for changing IP addresses. As connections are tied to the Home Address, Care-of Addresses which are used as locators can change at will. Indeed, Mobile IPv6 provides a rudimentary case of identifier and locator separation.

4.4.3 Host-Centric IPv6 Multihoming

The model of host-centric IPv6 multihoming was first briefly noted in [34], and has been presented again in a full form in [35], and is shown in figure 4.1.

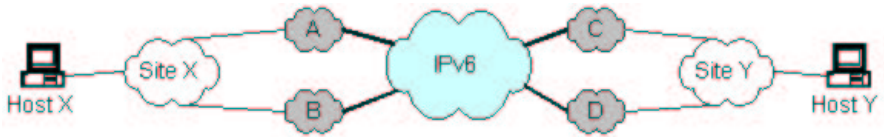

Figure 4.1: Host-centric IPv6 multihoming [27]

The idea of this framework proposal is that sites connect to and obtain IP address prefixes from multiple ISPs, here A, B and C, D for sites X and Y, respectively. Each node has multiple addresses, and all addresses are configured in the DNS and other relevant registries and configuration databases.

In consequence, when Host X wishes to communicate with Host Y, it will typically perform a DNS lookup which gives two addresses. Host X will perform default address selection, which picks out the best source/destination address pair. This will be used throughout that particular communication with Host Y.

The acknowledged problem with this model is that many ISPs nowadays perform ingress filtering; that is, they check that the source address of packets coming from their customer is one of those assigned to the customer. For example, ISP A would check that packets coming from Site X would be part of the prefix assigned to Site X by ISP A. With multiple addresses from different providers, many packets would also include the source address from the prefix of ISP B; these packets would be discarded.

There are multiple ways, as described in [35], to deal with this "site exit router selection" issue. The most prominent of them include:

- relaxing source address checks
- source address dependent routing
- source address selection by host

In the first, the ISPs would be contacted out-of-band, and asked to allow the traffic from the specific other source addresses, belonging to different ISPs.

In the second, the routers inside the site would perform routing based on both source and destination addresses, sometimes referred to as "policy routing", instead of just destination addresses – to transport the packets with source address out of particular ISPs prefix to that ISP.

In the third, the site exit routers would inform the hosts to pick a different source address, e.g. by using an ICMP message, or to contact the correct site exit router via some other means, e.g. by using routing header or tunneling.

4.4.4 IPv6 Multihoming at Site Exit Routers

The model of IPv6 multihoming at site exit routers is presented in [36]; it is based on a slightly on a similar technique specified earlier for IPv4 [24].

The idea is that the site connects to more than one ISP using multiple border routers, obtains IP address prefixes from each ISP, and deploys multiple addresses on every node, exactly as described with host-centric multihoming, above.

The difference comes from the fact that site exit routers form an adjacency with all ISPs. This could be done using physical links, but that is expensive; instead, IPv6-over-IPv6 or IPv6-over-IPv4 tunneling techniques are proposed. This is shown in figure 4.2; the dotted lines are separate tunneled connections to the other ISP(s).

The benefit of this mechanism is that when one of the ISPs or the link to it becomes unreachable, and hosts are still using addresses out of that ISPs prefix. Then the connections and connectivity are not lost, but can be resumed, although suboptimally, using the link through any other ISP, and using an interconnection between the two ISPs. This makes a rather reasonable assumption that the interconnection and the ISP's upstream connectivity are still functional.

This model shares the same issues of source address selection as described in host-centric multihoming.

4.4.5 Geographic Address Allocation

There are, and have been for a long time, a lot of proposals to allocate IP addresses in some geographic fashion, making them independent of the network topology. Such proposals have always been rejected in the past; with recent proposals, the twist is that the "geo-PI" addresses would not replace but augment current provider-based address allocations.

A lot of proposals have been made recently, like [37]. Typically, these consist of some party advertising an aggregated prefix of a whole region, and their upstreams advertising even less specific routes to limit the number of routes in the global routing table.

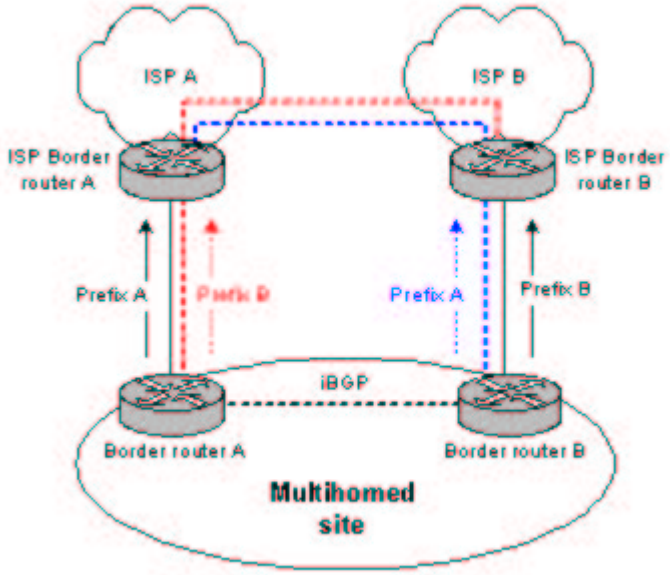

Figure 4.2: Multihoming at site exit routers [27]

The fundamental, and acknowledged, problem with these is that the model, more or less, requires a neutral party – like an Internet Exchange with redundant-enough upstream connectivity – to do it, or a billing structure for the advertisement service from someone who does. This is typically a very problematic situation.

4.4.6 Provider Independent Addressing Derived from AS Numbers

A proposal [38] suggests a possibility of creating Provider Independent address space for end-sites from the 16-bit AS number space. This would be done by including AS number as an identifier in the IP address prefix. The mapping is only defined for currently-used AS numbers, that is, the first half of the 16 bits.

There are multiple possibilities of how such a mapping could be done. To take two examples with AS1741, 06CD in hexadecimal, the address space could be e.g. 2000:6CD::/32 or 2001:0:6CD::/48.

This approach would give implicit provider-independent addresses immediately to everyone who has an AS number, without additional address space applications from RIRs or LIRs.

This model would expand the group who can advertise routes globally from ISPs to include the current end-sites which have acquired an AS number.

4.4.7 Other Mechanisms

Here, miscellaneous other ideas are presented in a brief format.

Advertising More Specific Routes

[39] proposes that as the number of multihomers is not yet that large, advertising longer prefixes belonging to end-sites could be a temporary way forward. This is similar to practice effective today with IPv4.

IPv6 Multihoming with Route Aggregation

[40] proposes a model where a site connects to multiple ISPs, and obtains IP address prefix from only one of them; only that primary ISP advertises the IP addresses of the site to the Internet. The secondary ones have an interconnection to the primary ISP, and act as backup connectivity against link and router failures.

End-to-end Multihoming

[41] proposes a model where the end-nodes themselves make the decisions on which destination and source addresses to use; this is to be done e.g. by making hosts have the global routing table, or portions thereof, with site-controlled metrics. Transport and application layers would be modified to be aware of all the addresses and their preferences.

Traffic Steering Using Routing Header

For outbound traffic, the packets could be steered by including a routing header in the packet. For example, an end-host could pick the site exit router corresponding to the selected source address using one, or nodes could choose ISPs for their special purposes.

This does not work for inbound traffic, as nodes in the Internet initiating new connections cannot know how to use it, and replies to packets which include a routing header must not use this information for specifying the "return path" unless the packet was authenticated with IPsec, due to security concerns.

MHAP

Multihoming Aliasing Protocol (MHAP) [42] proposes a protocol to be used at the edges of the network which would rewrite addresses to a special format when entering from the site, and would rewrite the addresses back to the normal form when reaching the other edge. The rewriting would happen between special "singlehomed" and "multihomed" prefixes.

Router Renumbering

Router Renumbering proposal [43] describes an architecture where routers have the router renumbering protocol to communicate changes in the network topology and take them into account automatically. This has been proposed to enable additions, changes and expirations of prefixes throughout the site.

According to the proposal, site exit routers would be able to notify e.g. the internal router network in the case of link failures and similar events.

If sites would be able to renumber easily, for example using this mechanism, there would be less need for independence and even requirements for redundancy could be more easily met.

4.5 Multi-connecting

Multi-connecting means connecting multiple times to a single ISP. As noted in the definitions, multi-connecting is not considered a multihoming mechanism in this thesis. However, it still deserves a brief description.

Multi-connecting is typically achieved by having multiple site border routers and connecting each of them to separate routers at the ISP, usually in different locations.

Often, a routing protocol is run between the site and the ISP, to be able to quickly detect errors in the connectivity and to switch to alternative paths. Routing protocol used is often BGP with private AS numbers, as described in section 3.1.1.

Sometimes, typically when the ISP is also in charge of the management of site border routers, some other protocols such as OSPF or IS-IS are also possible – this often requires strict filtering of advertisements at the ISP end, to prevent compromising the ISP's core routing system. In some cases, simply static routes may also be possible; typically this requires a link-layer medium which provides notifications if the end-to-end link-layer path becomes nonoperational.

Multi-connection is a flexible mechanism and relatively easy to accomplish: it requires no coordination between ISPs, no address applications for provider independent address space, no AS number applications for BGP AS numbers or the like. Yet it provides protection from the most common set of problems, as analyzed in section 6.1.8.

If one link fails, the changes do not need to be propagated further than the ISP; therefore, this is a very scalable approach when considering the global routing table.

Chapter 5

Current IPv4 Multihoming Practices

Having described the scope, the motivations, the background, the data collection and processing, and the site multihoming, it is time to analyze current IPv4 multihoming practices. In the next chapter, the situation with IPv6 is analyzed at length.

First, different observed route advertisement types which could be applicable to multihoming are listed. Second, other means to gain insight into route advertisements are described. Third, the collected, categorized and processed data is presented. Fourth, multi-connecting is described, and last, organizations are classified to different multihoming types based on the work in this chapter.

5.1 Categorizing Route Advertisements

In this section, different observed route advertisement types which could be applicable to multihoming are listed.

These are broken down to four categories: clearly multihomed, possibly multihomed, unclear cases and multihomed by transit. As appropriate, the specific mechanisms are described under these categories.

In particular, it has not been the intent to separate the route advertisements into multiple very strict types and categories, as it seems clear that there is a large unknown variable to most of the advertisements – and such classification would likely not be reliable or too generic to be useful.

One should note that it is impossible to prove that multihoming, by some means, is not being done; the only thing that can be guaranteed is whether it's done by means that are recognizable at the observation point.

The descriptions include simple figures which try to illustrate the cases, as well as a concrete example of the route advertisement as appropriate; these make it easier to

understand the cases and how they differ from each other.

The notation used in the figures combines two facts: the advertised address space and the AS numbers in the route advertisement path. In particular, the AS number objects closest to the address space circles are the advertisers of the address space. In one particular case, the address space is advertised by two equally close AS numbers. Also, in two cases, address space advertised by one is a subset of the one advertised by the other; this is depicted as a smaller inner circle.

5.1.1 Clearly Multihomed

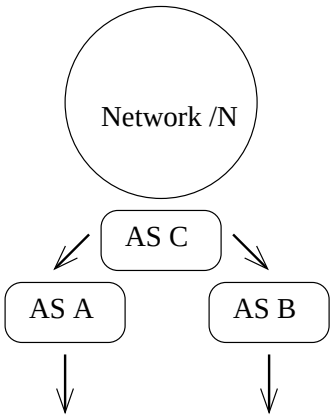

Figure 5.1: Clearly multihomed

If an identical prefix is being advertised via two different paths, and the origin AS is the same, it is clearly multihomed, as shown in figure 5.1; the origin site has obtained an AS number and possibly IP addresses, and is doing multihoming to more than one ISP using BGP.

An example of this is AS5420 behind two ISPs:

```
137.33.0.0/16 (3 entries, 1 announced)
     Nexthop: 212.226.101.142
     AS path: 9060 5420 I

137.33.0.0/16 (3 entries, 1 announced)
     Nexthop: 212.226.101.150
     AS path: 3246 5420 I
```

5.1.2 Possibly Multihomed

This category lists two possibly multihomed cases; it is not assumed that all such cases are in fact multihomed, but some of them certainly are.

An Identical Prefix from a Different Origin

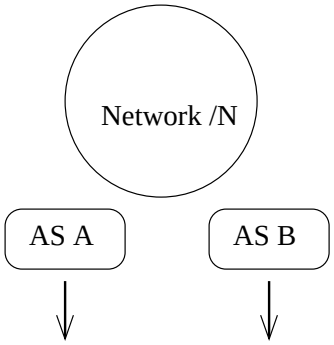

Figure 5.2: An identical prefix from a different origin

If an identical prefix is being advertised via different paths using different origin AS, as shown in figure 5.2, this could be either:

1. the site is connecting to two ISPs but with a private AS number or static routes, and these do not show in the AS path in the global routing table, or
2. multiple parties are advertising the same prefix intentionally.

The latter is interesting; there are some cases where multiple origins intentionally advertise the same prefix. One example of such is providing "anycast" -service using the routing protocol: multiple servers providing the same service, and the traffic always going to the closest one; one such service in an experimental anycast root DNS [44].

Often the prefixes are published in a routing registry; see section 3.1.5 for more. In this case, every instance that advertises the prefix should add an entry to the registry.

Otherwise, advertising the same prefix as someone else can be troublesome; the route is viewed questionable at best, illegitimate at worst. This is because the advertisement cannot be distinguished from any other unauthorized advertisement; indeed, the latter case of advertising the same, or more specific, route is sometimes called "route hijacking".

An example of this is:

```
192.49.189.0/24 (5 entries, 1 announced)
     Nexthop: 212.226.101.146
     AS path: 1759 5515 ?

192.49.189.0/24 (5 entries, 1 announced)
     Nexthop: 212.226.101.49
```

```
     MED: 100
     AS path: 719 ?
```

More Specific Routes from a Different Origin

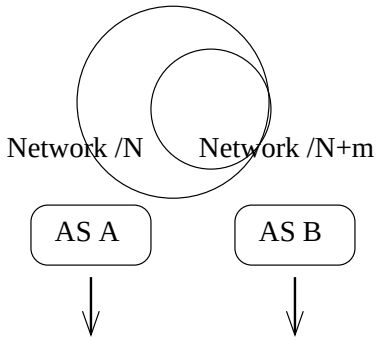

Figure 5.3: More specific routes from a different origin

If one route is advertised by one origin AS, and someone else advertises a more specific route, as shown in figure 5.3, this could indicate either:

1. the site having changed ISPs but taking the IP addresses with it, e.g. by economical transactions like payments, or
2. possibly multihomed; the other ISP where the more specific route is advertised is the "primary" ISP and the one advertising the less specific aggregate is there for backup.

A more generic form of advertising more specific routes where the origin AS is not considered is commonly referred to as "punching a hole in an aggregate".

An example of this is:

```
157.124.0.0/16 (2 entries, 1 announced)
     Nexthop: 212.226.101.146
     AS path: 1759 5515 ?

157.124.16.0/21 (4 entries, 1 announced)
     Nexthop: 212.226.101.49
     MED: 100
     AS path: 719 1738 I
```

It's assumed that typically, this is not a sign of multihoming, but changing providers.

One possibly interesting usage scenario would be a site obtaining a new primary ISP but keeping the old one for backup, at least for a while. This would be multihoming.

There is one particular special case when the advertisement of the more specific route is only coming inside the same ISP, but with a different origin AS. This is rather rare, and does not indicate multihoming or changing ISPs.

5.1.3 Unclear Cases

Here, four relatively unclear cases are described; it is typically rather difficult to detect the extent of multihoming in these cases.

More Specific Routes from the Same Path and Origin

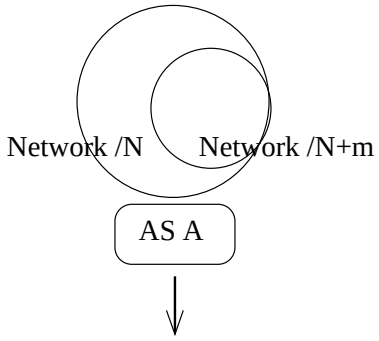

Figure 5.4: More specific routes from the same path and origin

If a less specific route is advertised from the same origin AS and is on the same path as the more specific route, as shown in figure 5.4, this is probably one of:

1. mistake by the advertiser or his ISP; this is probable – the route could have been aggregated, or
2. some form of traffic engineering (TE); e.g. the route may include some TE communities.

In this particular case, the atomic aggregation attribute, as an indicator of automatic rather than manual aggregation, in the less specific route is often a sign of misconfiguration rather than traffic engineering.

It seems unlikely that multihoming is being done in this case.

A Route with AS-path Prepending

If a route is received with AS-path prepending in it, as described in section 3.1.1 and shown in figure 5.5, it is very probably either:

1. traffic-engineered, or

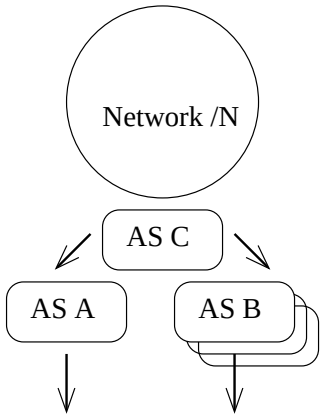

Figure 5.5: Route with AS-path prepending

2. multihomed.

This is because there is no reason to prepend a route unless it is being advertised via more than one path.

Particularly interesting case of this would be those AS-path prepended routes for which no other similar prefixes exist, signifying that the other route(s) have been considered better on their paths.

An example of this is:

```
81.22.160.0/20 (3 entries, 1 announced)
     Nexthop: 212.226.101.122
     MED: 1000
     AS path: 8434 24713 I

81.22.160.0/20 (3 entries, 1 announced)
     Nexthop: 212.226.101.150
     AS path: 3246 3246 3246 3246 24713 I
     Communities: 3246:13 24713:358
```

However, one should note that prepending in itself is not a sign of multihoming. It is always combined with some other technique.

One should also note the case where Nordic prefixes are being advertised in FICIX using traffic engineering with AS-path prepending, and probably without prepending elsewhere, over "primary" peering sessions. This is assumed to be relatively commonplace among those whose main business area is outside of Finland, and the sites are outside of Finland.

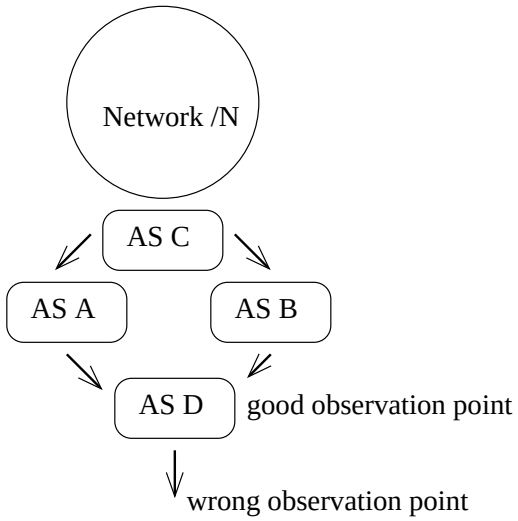

Figure 5.6: Multihoming behind Neighbor ISPs

Multihoming behind Neighbor ISPs

If the observation point is selected in such a fashion that there are two layers of ISPs, anything done strictly inside the first ISP is "hidden" there and impossible to notice. This is shown in figure 5.6.

For example, as in the figure, if "AS D" offers transit to both "AS A" and "AS B", and site multihoming is done by "AS C" between "AS A" and "AS B", but "AS A" and "AS B" have no other transit than "AS D", any "internal" operations will go undetected if only advertisements from "AS D" are seen at the observation point.

This is particularly the case with approaches using more specific routes and/or private AS numbers; either could be filtered out by "AS D".

Some Degree of Multihoming Using NAT

If the site is using Network Address Translation (NAT) [16], they can gain operator independence easily; simultaneous multihoming is a more difficult task. This has been described in section 4.3.2.

5.1.4 Multihomed by Transit

If an identical prefix with an AS path is received from one neighbor and the same prefix is received from another neighbor where the AS-path fully includes the shorter AS-path, the neighbor ISP is doing multihoming and this is not an evidence of site multihoming, but rather a peering and backup agreement between ISPs, and as such, out of scope; this is shown in figure 5.7.

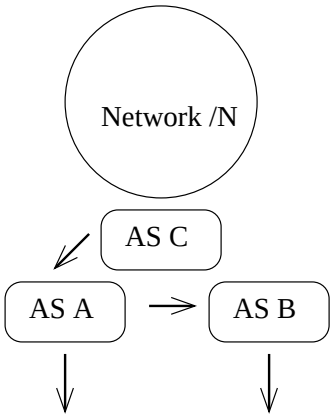

Figure 5.7: Multihomed by transit

An example of this is AS790 also being advertised through AS6667 and AS8434, like:

```
62.142.0.0/16 (3 entries, 1 announced)
     Nexthop: 212.226.101.122
     MED: 1000
     AS path: 8434 6667 790 I

62.142.0.0/16 (3 entries, 1 announced)
     Nexthop: 212.226.101.105
     AS path: 790 I
```

To re-iterate, this is not considered site multihoming.

5.2 Collecting Information by Other Means

In addition to analyzing the route advertisements and categorizing them, other methods were also used.

First, the typical operational practices have been followed on various fora. Second, queries were sent to the major ISPs in Finland to ask for clarifications on certain methods and their views on multihoming. Last, a particularly questionable set of route advertisements was compared to an older one, to gain a time development perspective.

5.2.1 Following the Used Practices

Experience and following and participating in routing discussions on various fora have contributed to the classification of the route advertisements significantly.

Even though there are plausible technical explanations for doing certain kinds of route advertisements, often attributed to traffic engineering or multihoming, observing the operational practices and reality gives an entirely different perspective.

This gives two summarizing thoughts which have been taken into consideration in the classification in the previous section:

- "when in doubt, suspect a configuration mistake", and
- "bigger organizations do everything they can to avoid renumbering even though their ISP changes".

5.2.2 Queries to Major ISPs on Multihoming Practices

A lot of possibly multihomed (5.1.2) or unclear cases (5.1.3) surfaced, as expected. It seemed prudent to try to gain at least some additional knowledge about the relative amount of these uncertainties.

Therefore, a query was sent to all full FICIX members, asking them to answer a few questions on questionable multihoming practices, and asking them to try to describe the questionable route advertisements specific to that operator.

The query and an example operator-specific appendix can be seen in appendix A.

Responses were received from about a third of the membership, constituting a significant portion of Internet traffic in Finland. The responses are summarized below.

The responses for the query will be described in the same order as questions asked; for more information, refer to appendix A.

Identical Prefix from Different Origin

Some ISPs have advertised identical prefixes from their own AS; typically this has been done when changing the operator when the customer has its own addresses but no public AS number.

More Specific Routes from Different Origin

It's generally not considered acceptable for someone to advertise more specifics of your aggregates; also, yourself advertising more specifics from someone else's aggregate is also considered a bad practice. However, sometimes it "must" have been done, mainly due to "commercial reasons".

A part of these is multihoming, while it seems the majority of these cases is making provider-aggegatable addresses provider independent.

One cited reason for "punching holes in aggregates" is that previously the correctness of advertisements was not so worrisome, and changing operators without changing addresses was considered acceptable, ie. historical reasons.

When analyzing the more specifics with a different path there are some sources of confusion: for example, there are some traditional network blocks which are not even meant to be aggregated, but some advertise the aggregates anyway – this gives a false positive. However, there are only a few of them, so this is not extremely significant.

The more specific advertisements are not considered to be temporary; some consider that, as policies have been established, new cases should at least be temporary.

Some allow more specific routes from FICIX to their own aggregates. Some also have secondary network connections to such end-sites, indicating multihoming.

More Specific Routes from Same Path and Origin

All agree that advertising more specific routes from the same origin is usually a sign of a configuration mistake; typically due to tests or network topology changes or forgetting to update aggregation configuration. Operators themselves do not use this method for traffic engineering purposes, but some of the customers might.

Multi-connecting

None of the operators were aware of other multi-connecting mechanisms than those listed, that is, BGP, OSPF/IS-IS or manual fail-over.

Many customers have multiple BGP sessions, typically with private AS-numbers. With some operators, the use of OSPF is even more typical than BGP.

The reasons for picking multi-connecting instead of multihoming seem to be that some end-sites want to maximize usability and do not trust the provider, or there is a policy decision.

Multi-connecting is also typically a much cheaper, and easier to configure than the full multihoming solution; faster convergence was also cited as a reason for multi-connecting.

In fact, one study [45] – which has since then been criticized, shows that convergence may actually be much slower with multihoming.

Multihoming behind Neighbor ISPs

There are a few cases of small ISPs which are both customers of the same bigger ISP which were queried here; in such a case, some information about multihoming may be lost in the analysis.

Some Degree of Multihoming Using NAT

Typically the routing operations people are not aware of NAT being used as a primary multihoming mechanism, but as it's invisible to the network and many do use NAT, many guessed that some in fact had used it to change the operators when needed, at least.

Miscellaneous Observations

Some advertisements which used prepending with only one path were configuration mistakes. Some multihomed cases with private AS numbers were explicitly identified. There were some configuration mistakes, both in more specific routes from the same and different path. The former were mostly configuration mistakes, but there were a case or two of private AS-number use which may have been traffic engineering or multihoming.

5.2.3 Development of Certain More Specific Routes

The development of certain more specific routes was examined to gain insight in a sometimes-heard argument that when changing ISPs, the old addresses were advertised through the new ISP only temporarily.

First, certain general developments on the number of prefixes advertised were noted; these are table 5.1. The values are mainly discussed in section 5.3.2 and tables 5.6 and 5.7.

Table 5.1: Development of the number of prefixes advertised

	16 July, 2002	29 January, 2003
Clearly multihomed	26	52
.. more than dual-homed	1	5
Multihomed w/ different origin AS	2	0
Multihomed by transit	0	117
More specifics	200	212
More and less specifics	1	3
Less specifics	56	90
More specifics with a different path	101	159
More specifics with the same path	100	53

From there, the only thing which was compared was the longevity of more specific routes with a different path. That is, advertisements about 6 months apart were compared by a manual check: whether the same more specific routes had been advertised in the past.

Table 5.2: Development of more specifics from a different path

	Amount
New prefixes	71
Still advertised prefixes	83
Removed prefixes	15

The changes in the advertisements of more specific prefixes with a different path are noted in table 5.2.

There is a slight inaccuracy in the counting, which is why the total number does not equal to results in table 5.7; this is not significant or important. Due to this uncertainty factor, percentages and other more precise characteristics are not calculated.

However, the analysis shows clearly three things:

- advertising more specifics with a different path has gained significantly in popularity,
- the number of long-lived more specifics with a different path is high, and
- the number of withdrawn prefixes is small compared to the number of new or still advertised prefixes.

To conclude, it seems safe to say that only few use this mechanism as a temporary mechanism: instead, the method appears to be gaining in popularity and being a rather permanent arrangement.

5.3 Categorized and Processed Data

Now, after considering the other information sources, the categorized and processed route advertisement data is presented. The data was based on a snapshot taken on January 29, 2003.

First, some generic, not necessarily multihoming-related, data is observed. After that, the multihoming-specific data is presented.

5.3.1 Generic Data about Advertisements

The number of prefixes advertised, broken down by prefix length, is summarized in table 5.3. Included are two percentages: absolute, from the number of all prefixes, and relative, from the covered address space. The prefix lengths which had no advertised prefixes were left out for brevity and clarity.

Table 5.3: Prefix breakdown by prefix length

Prefix length	Amount	Absolute (%)	Relative (%)
/14	1	0.1	2.08
/15	6	0.4	6.25
/16	114	6.9	59.41
/17	23	1.4	5.99
/18	39	2.3	5.08
/19	209	12.6	13.61
/20	116	7.0	3.78
/21	52	3.1	0.85
/22	79	4.8	0.64
/23	147	8.9	0.60
/24	834	50.2	1.70
/25	3	0.2	0.00
/27	6	0.4	0.00
/28	5	0.3	0.00
/29	13	0.8	0.00
/30	11	0.7	0.00
/32	3	0.2	0.00
all	1661	100	100

The covered address space is calculated by giving each prefix length a weight associated with its size; this is slightly inaccurate when the prefixes overlap, but should be enough for illustrative purposes.

So, one can conclude that the vast minority of those who advertise the more specifics take the majority of the resources.

The numbers and percentages of prefixes with community, atomic aggregation or MED attributes are listed in table 5.4.

Table 5.4: Characteristics of advertised prefixes

Characteristic	Amount	Percentage (%)
Prefixes w/ community values	84	5.1
Prefixes w/ atomic aggregation	157	9.5
Prefixes w/ MED	831	50.0

Last, the numbers of neighbor autonomous systems, transit AS's and originating AS's are listed in table 5.5. Certain other characteristics are also listed in the same table: the number of AS's which are both neighbors and transits, the number of AS's which do not advertise any prefixes, and the number of AS's which use AS-

path prepending.

Table 5.5: Characteristics of autonomous systems

Characteristic	Amount
Neighbor AS's	15
Transit AS's	50
Originating AS's	273
Both neighbor and transit AS's	3
Non-originating AS's	1
AS-path prepending AS's	47

5.3.2 Multihoming-specific Data

Now, having considered some generic data on advertisements, let's have a look at multihoming-specific information.

Table 5.6 lists the numbers and percentages of prefixes of:

- multihomed by transit ("ISP multihoming"), as described in section 5.1.4,
- uniquely multihomed, as described in section 5.1.1,
- of which those that are uniquely multihomed to more than two ISPs, and
- those multihomed with the same prefix but different origin AS, as described in section 5.1.2.

In comparison, unique multihomers have 24 different origin AS numbers, and 13 of them originate only one prefix.

Table 5.6: Advertisements of the same prefix

Characteristic	Amount	Percentage (%)
Clearly multihomed	52	3.1
.. more than dual-homed	5	0.3
Multihomed w/ different origin AS	0	0
Multihomed by transit	117	7.0

In table 5.7, the numbers and percentages of more or less specific routes are listed; "more and less specifics" refers to those routes which have both of them. In the latter part, percentages relative to the number of both all prefixes and more specifics only are separated. These cases are described in sections 5.1.2 and 5.1.3.

Table 5.7: Advertisements of the more and less specific routes

Characteristic	Amount	Percentage (%)
More specifics	212	12.8
More and less specifics	3	0.2
Less specifics	90	5.4
More specifics with a different path	159	9.6 / 75.0
More specifics with the same path	53	3.2 / 25.0

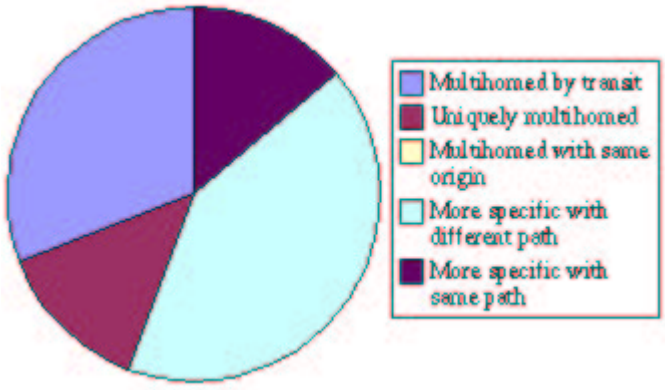

Figure 5.8: Distribution of different techniques

The distribution of different techniques relating to multihoming is shown in figure 5.8.

The total effect of site multihoming to the global routing table is difficult to measure due to the uncertainties in the route classification.

There is a relatively high number of end-sites which multihome by advertising a more specific route to their IP address via another path. These may or may not have a public AS number. It is difficult to measure how many of these exist, but it would seem safe to assume that they are at least 30-40% of the more specifics with a different path.

By that assumption, the amount of site multihoming in Finnish networks seems to be 6-8% (with assumptions of 30% and 50% of more specifics being multihoming, respectively) of all prefixes.

Looking at the development of advertisements during the 6 months, it seems that multihoming is in the rise as noted in table 5.2; the figures in table 5.1 are easily comparable as the introduction of ISP multihoming makes it more difficult to make reliable projections.

This may not apply to other countries. In particular, the number of large enterprises is rather low in Finland. A different data collection model will be needed for more

precise analysis.

5.4 Multi-connecting

Multi-connecting was described in section 4.5, as well as queried from the operators above, in section 5.2.2.

In summary, multi-connecting seems to be a relatively typical practice. Most often, it is done with BGP using private AS numbers, but some operators also use other protocols, such as OSPF; this would also seem to depend on the service model of the ISP and the customer – whether the ISP manages the site's border routers or not.

The primary justification for multi-connecting is that it's rather easy to achieve, and requires no AS numbers nor specific IP addresses, if done properly. The convergence time, considering for example service times for the primary network connection, also seems superior to the BGP multihoming scenario.

The main reason why multi-connecting is not considered to be enough seems to be some form of policy decision and desire for complete independence, or not trusting the operator to work properly.

5.5 Classifying the Organizations

After having described the IPv4 multihoming and multi-connection techniques and operational practices, it's time to try to create some rough classifications for organizations. Then, some observed motivations for different multihoming types are described.

5.5.1 Classification

First, there are a number of ISPs which choose to multihome through other ISPs as noted in section 5.1.4; these are outside of the scope of this thesis.

Second, there are a number of end-sites which have a public AS number and their own IP addresses; these are clearly multihomed. Almost all of them multihome to two ISPs, not more. Let's consider this as type A multihoming.

Third, there is a relatively high number of end-sites which multihome by advertising a more specific route to their IP address via another path. These may or may not have a public AS number. At least 30-40% of these are really multihoming, as described in section 5.3.2 – the rest have just switched ISPs. This is type B multihoming.

Fourth, there is the category of end-sites which have their own addresses but no public AS number; considering how easy it is to apply for an AS number relative to addresses, this is not so high. This case results in identical route advertisements if

they'd multihome to switch ISPs; currently, there are none of this kind. This is type C multihoming.

Fifth, multi-connecting is a relatively common practice in certain circles. There are multiple ways to accomplish it, but few of those are visible in the route advertisement data. This is not multihoming as such, but in this classification, still considered type D multihoming.

Last, NAT and other such techniques may be used for multihoming purposes as described earlier. Such methods seem to be commonplace when switching ISPs, but not so much when connecting simultaneously to multiple ISPs. This is type E multihoming.

5.5.2 Motivations

Type A multihoming seems typically the way it is done by large end-sites, who have both the resources and the will to do multihoming "all the way".

Type B multihoming seems usually the way it is done by smaller end-sites which do not care whether the solution is the right one or not, as long as it works for them.

Type C multihoming is very rare due to the relatively easy way to move into type A category.

Type D multihoming is rather commonplace and solves most needs; indeed, it could be an option for some other multihoming types too.

Type E multihoming does not seem to be used that much; taking one to use is a bit complicated when other options such as type B are also available – otherwise, it might be more popular.

To summarize, the number of ways of multihoming by end-sites seems rather limited. This is probably due to the fact that currently there are also relatively easy ways to obtain multihoming for an end-site for example, by convincing the ISP to advertise a more specific route. Consequently, there does not seem to be a real demand for multihoming mechanisms that would be architecturally correct and scalable. A change in thinking is likely to be required when deploying IPv6 multihoming, it seems.

Chapter 6

Applicability of IPv6 Multihoming Solutions

Having described the scope, the terminology, the background, the data collection and processing, the site multihoming, and the current IPv4 multihoming practices, it's time to analyze the situation with IPv6 at length. In the next chapter, conclusions are presented.

First, the IPv6 multihoming mechanisms presented in 4.4 are analyzed. Then, a classification for different organization types and their multihoming requirements is done. Last, methods for choosing the multihoming mechanism are described; in particular, the applicability of multihoming mechanisms is analyzed and the solutions meeting the requirements of different organization types are examined.

As noted in the previous chapter, the situation with IPv4 is rather chaotic; many end-sites and ISPs use mechanisms which are the easiest for them, with little regard to the global routing system. Avoiding renumbering when changing ISPs seems to be the number one priority, having redundancy and other benefits of operator-independence coming close behind.

Now with IPv6, it's either possible to repeat the same mistakes or try to fix them, even if it meant having some growing pains. I'll focus on the latter, as the adoption of IPv6 gives a possibility to do start from scratch.

6.1 Analysis of IPv6 Multihoming Mechanisms

The mechanisms outlined in 4.4 are now revisited for analysis in the same order.

Further applicability, deployability and possible use cases in different scenarios are examined later, in section 6.3.1.

Challenges and future work are mainly described later, in section 7.2.

6.1.1 Transport Solutions

Transport solutions, such as TCP modifications and SCTP have a major problem: such modifications would have to be deployed everywhere so that they could be depended on, and all the connection-oriented, and maybe even some connectionless, protocols would have to solve the exactly same problem multiple times – if a solution would even be possible.

Ignoring the initial resistance to these proposals, several points can be underlined.

The modifications for TCP must be backward-compatible due to the huge installed base, to the extent of classic TCP sessions. It might be marginally acceptable to allow a solution which would work without major drawbacks with all implementations but would provide additional features only if both the endpoints supported the modified protocol.

SCTP is likely to be deployed at least in a few relatively restricted environments, but not as a global solution; however, the application interface also has a TCP-like syntax, so changing the applications to use SCTP rather than TCP is a small effort. In fact, it might even be possible to create an abstraction layer in the host systems where all applications would think they're using TCP but would use SCTP instead.

Summary

Looking at the big picture, transport solutions could be used to complement the host-centric multihoming approach. However, it looks like the proposed approaches are, in the global scale, too little and too late: if a major change is needed, why not go directly to the locator/identifier -separation approach which provides similar functionality for all the protocols? It could also be argued that the window of opportunity for global deployment has also already passed.

Protocols such as SCTP are undoubtedly useful in certain specific applications such as the application of node multihoming in telephone signaling transport, but are unlikely to become popular in the global perspective.

Transport solutions build on having multiple addresses per node and solve the connection survivability problem.

6.1.2 Identifier and Locator Separation

LIN6

The identifier and locator separation approaches seem to be interesting approaches in the long term, as an architectural decision.

In LIN6 [33], the mapping function is done between basically an interface-ID derived from the assumed-globally-unique Ethernet MAC-address; the mappings are entered in a database. It is not evident that this kind of mapping can be made reason-

ably secure or scalable, or that the particular approach to deriving globally unique identifiers would the best one.

In addition, LIN6 has patents or patent claims associated with it, making the approach unacceptable as a requirement for protocols providing multihoming.

Therefore, LIN6 is not considered useful as a way forward.

HIP

In HIP, the emphasis is on security from the start. However, one could say that security is also a weak point, if one would consider it being adopted globally.

HIP uses a rendezvous service to map Host Identities to addresses. A problem is that the mapping must be secured using DNSSEC [46] or a similar mechanism; otherwise there is no guarantee that the other endpoint's IP address that was returned from the rendezvous service really belongs to the party with the Host Identity.

Another approach to this is employing an "opportunistic" mode, like introduced with Secure Shell [47] key management: instead of building a key infrastructure, obtain the key when first contacting the other node, store it, and start using it. The keys can also be manually verified e.g. by creating short fingerprints of them, to be verified by off-band methods. If, at some point, the other end does not match the previously stored key, one can suspect that something bad is going on. Of course, this model has drawbacks as well: it is mostly suitable for communications which occur between previously known nodes. The model bases its security on the assumption that typically services have not been compromised or hijacked.

Another issue with security, from a completely different point of view, is the use of encryption on all the traffic. It is typical that different entities like enterprises want to be able to monitor and set policy on the traffic, for example in firewalls. This is not possible with encryption; the policy-making process is either pushed to the end-nodes or one must use some kind of "security gateways" to act as a proxy for all the traffic. As can be seen, the use of encryption may not necessarily fit well in all the current operational environments.

Of course, the HIP protocol could probably be modified so that the end-nodes could also negotiate an unencrypted but authenticated association, depending on the wishes of the end-hosts, or use a proxied HIP to begin with.

The experience and specification of HIP is still at an early stage; whether it can be made to work smoothly and whether the Internet is ready for a change like that remains to be seen in a few years.

Mobile IPv6

Mobile IPv6 has been proposed to be used to solve the problem of connections breaking when the IP address changes in the multihoming context; Mobile IPv6

currently provides such functionality for moving, mobile nodes.

However, this does not really solve the problem as is, just shifts it around: an approach like this would require the Home Agents to be addressed and located in such a place in the network topology that they would not be affected by the outages. This is typically not the case. A Homeless Mobile IPv6 approach has been proposed but withdrawn by the authors; in this case, something like HIP is closer to the right solution.

The issue is that in order for the Correspondent Node to verify the Binding Update sent by a multihomed node the primary connectivity of which has failed, it must verify that the Care-of and Home Addresses are routed at the same node. However, in the case of multihoming-related network outage, by definition the previously used Care-of Address is no longer operational – and the verification will not succeed.

If a global Public Key Infrastructure was in place, it might be possible to authenticate the modified Binding Updates so that connections could survive; however, this does not seem to be realistic in the short term.

The proposal for the operation of the modified Binding Update has not been written out, so it is difficult to analyze whether sufficient security properties could be obtained without an explicit home agent.

Summary

To summarize the locator/identifier mechanisms, LIN6 does not seem reasonable, standard Mobile IPv6 does not work, but a modified version could be explored, and HIP appears to be the most interesting architectural proposal in the long term but not fleshed out sufficiently to say.

Locator/identifier solutions build on having multiple addresses per node and solve the connection survivability problem.

6.1.3 Host-Centric IPv6 Multihoming

The host-centric IPv6 multihoming is a framework of all nodes having multiple addresses and the hosts being in control of many of the multihoming decisions.

The most prominent issue with nodes using addresses from different providers is the correct source address selection. Disabling ingress filtering between the ISP and the site is not a realistic recommendation for the safety of the Internet [48].

Using multiple addresses also requires additional solutions if connection survivability is desired.

The requirements for the mechanism to work are:

1. source nodes will be able to pick a working source address, if not first, at least eventually, and

2. packets with the source address belonging to ISP X will be routed to the site-border router with an active interface to ISP X, and forwarded there.

The first requirement may not have to be solved if one can assume that connectivity to all the providers is always maintained, e.g. through a tunnel as described next. If this cannot be assumed, the information of the failure of one link must be propagated to the end nodes somehow. This can be done by e.g. routers using ICMP to suggest picking the different source address – this could possibly even be propagated to the first hop routers, by distributing information to reject traffic by the use of special routes. A few other techniques might also be possible.

The second requirement can be solved by requiring the use of source-based routing in all routers of the site – or at least the site exit routers but that would induce a possible extra hop(s) for the traffic, the use of routing header by the source nodes to the site exit routers or tunneling to the site exit routers.

A special case is when only one ISP is in primary use; this can be achieved with a default route with a very good metric which is propagated through the site's routing system which overrides all the other sites' routes everywhere in the site. The end-nodes will have to support source-based routing or have their default address selection policy database [49] modified to pick the source address belonging to the primary site; either of these could be automated in e.g. route advertisement messages. Then, no additional infrastructure to support the model would be needed.

Summary

Host-centric IPv6 multihoming in itself provides multiple addresses per node, but does not solve the connection survivability problem. There are some details to be worked out yet. Outbound, and inbound in particular, load-balancing may be difficult if the policy has to be set at the edges as decisions are made by end-hosts.

6.1.4 IPv6 Multihoming at Site Exit Routers

This model builds on and expands the host-centric multihoming approach; in addition, a tunneled backup connection is established to all the ISPs through other ISPs.

This is very useful: in the normal case, if a link or router fails, connectivity to an ISP may be disturbed and connections using the addresses belonging to that ISP will break. Now, the connectivity is quickly restored and goes tunneled via a slightly more suboptimal path through the other ISP.

This also makes the source address selection problem slightly easier to solve, as one can assume at least almost always having connectivity to all the ISPs; in practice, this yields the benefits of multi-connecting, described in sections 4.5, 5.4 and 6.1.8.

However, if the ISPs, or one of their upstreams perform ingress filtering, there is still

the problem noted above, routing the right traffic to the right site exit routers. In the case of two equally used site exit routers, the problem could also be worked around by forwarding the "wrong" ISP's traffic one router receives over to the "backup" tunnel. However, this seems a bit questionable.

A concern specific to this model is that the Maximum Transmission Unit (MTU) of the physical links to both ISPs must be sufficiently large: over 1500 (MTU on Ethernet) bytes would be best. This is because tunneling will increase the packet size by at least 20 or 40 bytes, and the maximum sized packet that could originate in the site should still fit in the link even in the encapsulated format with extra bytes. If this is not possible, one will have to rely on Path MTU Discovery, which is suboptimal, or configure a lower MTU everywhere in the site.

Summary

Multihoming at site exit routers in itself provides multiple addresses per node, does not solve the connection survivability problem completely but mitigates it dramatically so that it may not any longer be a problem.

There are some details to be worked out yet. The sites should seriously consider link MTU's on connectivity they obtain. Outbound, and inbound in particular, load-balancing may be difficult if the policy has to be set at the edges as decisions are made by end-hosts.

6.1.5 Geographic Address Allocation

Geographic addressing may first sound like a good idea, but in practice it is a very challenging concept.

Avoiding renumbering when changing operators and deploying a simple IP address plan without multiple addresses per node is what may seem desirable to many sites.

However, tying the numbering to geography may not be as useful as one might hope. It is not uncommon for sites to move from one place to another, at least in the mid-long term like 4-7 years. Renumbering would seem inevitable then anyway. This is not really all that different from the lifetime of ISPs.

Another significant problem with topology-independent addressing is global routing. Who is willing to advertise the aggregate for a region, when not all sites in that region are your customers and pay you for transit service? The alternative is advertising more specific "geo-PI" addresses from different ISPs, but that only makes the idea of geographic aggregation and allocation useless. What would be needed is either end-sites, or their direct ISP, paying explicitly to a sufficient number of separate transit ISPs to advertise the regional aggregates: a billing structure with a granularity of an end-site. An alternative is depending on neutral parties, such as Internet exchanges with enough redundancy and upstream connectivity, to perform the route advertisement. However, it appears that such an exchange would then be an ISP

with a different name, with similar trade-offs.

For the model to work to the sufficient degree, a requirement also has to be that regional traffic can be routed regionally, without going through a long way through other providers – as this would increase the number of more specifics in the routing table. This is quite problematic especially in the areas of lower Internet connectivity penetration: it is rather usual that the traffic is transported a long distance until it is spread out to the Internet. This causes problems when there are multiple providers in the area who all are doing this, instead of exchanging the routes directly, in non-existent exchanges or non-existent and non-profitable bilateral interconnections.

In any case, it is difficult to argue on technical grounds why the model could or could not work; however, the economic and organizational realities would seem to indicate a strong reluctance to committing to one.

Summary

Geographic address allocation provides a long-lived address per node, and thus connection survivability is not a problem. Typically only outbound load-balancing is possible. It does not seem to be an operationally or economically realistic approach.

6.1.6 Provider Independent Addressing Derived from AS Numbers

The model creates a PI address space ("ASN-PI") syntactically for every end-site that has an AS number – that is, the majority of people who could be multihoming today.

The approach is not the best one, as it would lead to a possible "land-rush" for AS numbers, and then in turn, AS number space would be exhausted, forcing the move to longer ones and a protocol change. This is explicitly forbidden by the specification, but the pressure to change it might rise once the number of people allowed to multihome using this mechanism was reached.

However, the proposed solution is considered significantly better than some alternatives, like using more specific routes. In this model, the routes are clearly controllable and distinguishable from the rest, and cannot lead to a mess of more specific routes nobody knows what they're for, as with IPv4 today.

At the end of 2002, origin-only AS numbers currently being used were about 10000 [50]. This doesn't seem like too high a number to satisfy current multihomers' needs.

One should note, however, that at least some multihomers, especially those of type B as described in section 5.5, may not have AS numbers at the moment.

Summary

"ASN-PI" allocation provides a stable address per node for existing AS number holders, and thus connection survivability is not a problem. Both outbound and inbound load-balancing are possible. The model might have some uncertainties regarding the applicability as a long-term policy.

However, if such advertisements would come from specific, well-defined address blocks, this might be a usable approach.

6.1.7 Other Mechanisms

Advertising More Specific Routes

The model where any ISP or even anyone is able to advertise more specific routes globally without control seems an unacceptable burden for the Internet, at least in anything other than the short term.

IPv6 Multihoming with Route Aggregation

The proposed model is basically the same as with "IPv6 Multihoming at Site Exit Routers", with the difference that no tunnels are built, but the ISPs agree between themselves to let the more specific routes, with the wrong source address, through.

Addresses are obtained from only one ISP, though – so this model is closer to multi-connecting than multihoming, as it does not offer any provider independence.

ISPs are typically competitors and might not be too enthusiastic with an approach like this; in addition it requires some added complexity especially with routing policies, so tunneling could be considered to be a bit more elegant approach.

An approach like this could be very usable when multi-connecting to a single ISP, but different locations which might have an address aggregation border between them.

End-to-end Multihoming

The model proposes having all nodes receive the routing data with site-specific metrics and be thus able to select the most preferred destination and source addresses when the transport and application layers have been modified to be able to receive all such addresses and sort them.

This is an interesting solution which could be classified as long-term, as quite a few modifications would be required, starting from making the routing table available to every node, modifying the applications and such. The details have not been fleshed out appropriately to make any concrete determination whether this approach would prove successful.

Traffic Steering Using Routing Header

As routing header can only be, in practice, used for outbound connections, its usability for traffic steering seems relatively limited. At most, it could complement other mechanisms.

MHAP

The procotol proposes an additional layer of indirection, shifting the scalability problems of site multihoming to a separate protocol. Address rewriting could be quite a fragile tool, as it is difficult, in an Internet-wide environment, guarantee that it will always be rewritten back and that no intermediate party, e.g. a router responding with an ICMP message, will need to access the original information.

Router Renumbering

Automatic renumbering is a very difficult problem – one that many have forgone due to its operational complexity. With current understanding, the mechanisms are not sufficiently extensive or robust to be able to handle rapid renumbering – that is, revoking prefixes when an ISP becomes unreachable, updating DNS to reflect for the changes, etc.

However, even though the initial approach has proved to be too complex, some experience from the protocol, for example communication between routers on changes of prefix availability, could be salvaged and considered to be used for the purposes of host-centric multihoming.

Summary

None of the mechanisms described here seem to be directly usable for multihoming purposes; however, some, like router renumbering, traffic steering, multihoming with route aggregation or advertising more specific routes seem to have some salvageable parts.

6.1.8 Multi-connecting

Multi-connecting is a trivial procedure and is available immediately in IPv6, compared to the multihoming mechanisms. In contrast, in IPv4 there was only a slight difference in the ease of use to the favor of multi-connecting.

Having said that, one should consider the implications of multi-connecting versus multihoming. Of the motivations listed in section 2.4, multi-connecting:

- doesn't provide any provider-independence,

- provides a significant amount of redundancy; typically, only such rare large scale failures which affect all the customers of the ISP cannot be protected against,
- provides load sharing, but only within the ISP,
- doesn't typically provide performance -related solutions, and
- doesn't typically provide policy -related solutions.

The concerns in the latter two usually exceed the ISP; in some cases, ISPs may be able to offer some restricted-scope solutions for these too, of course.

As can be seen, multi-connecting actually provides quite a bit of what's typically required of a multihoming solution except independence.

Especially if relying solely on multi-connecting, it usually becomes important to pick the ISP with care; in particular, one should look at the number of customers and the reliability as a whole. Based on this, one may be able to make an estimate how quickly the ISP would notice and fix critical problems affecting most customers – the rare large scale failures mentioned above.

Redundancy gained with multi-connecting always incurs faster convergence times than with multi-homing mechanisms; this is due to routing updates requiring only a limited amount of propagation, and the possibility to use faster-converging protocols if necessary. Therefore, if particularly fast convergence is a requirement, multi-connecting or virtual multi-connecting – that is, the multihoming at site exit routers approach – is recommended. In fact, one study [45] – which has since then been criticized, shows that convergence may actually be much slower with multihoming.

To conclude, it seems useful to keep in mind that two solutions, one which provides redundancy and, to some degree, load-sharing, and the other which provides independence may be able to solve the whole problem quite nicely.

6.2 Classification of Organizations and Motivations

Now, based on the earlier experience, classifications with IPv4 and such, the different types of organizations are separated, and their possible motivations explored separately.

In this way, it's possible to try to break the big "site" and "multihoming" concepts into smaller pieces.

First, a rough classification is presented, and then each of the four types are described at more length.

Analysis on which methods could suit each type best are described in the next section.

6.2.1 Classification

In the light of the classification in section 5.5, queries, experience and research as described above, a matrix of multihoming motivations and types of organizations is created; it is shown in the table below.

	Independence	Redundancy	Load sharing
Minimal	-	-	-
Small	?	?	-
Large	?/x	x	?
International	x	x	x

In the table "-" represents a non-important issue, "?" a possibly important motivator and "x" a certainly important motive.

Note that the motivations "Performance" and "Policy" are not included above: in practice, they do not seem to be prime motivators, and it is difficult to gauge how desirable features they are.

The term "IP-users" is used in the classifications. This is used to loosely refer to those people who have an operational need for Internet access to get their work done.

The organizations are classified roughly in four categories:

1. Minimal: a very small or restricted end-site, for example a typical home network, or a small office of less than 10 IP-users
2. Small: a small-to-mid-size enterprise, for example with less than 50-150 IP-users
3. Large: a large enterprise, for example a regional or national enterprise of typically less than 1000 IP-users
4. International: a large or very large enterprise, with a significant amount of international activity

The distinction between the last two comes from the assumption that an international organization is assumed to have major activity in multiple countries or regions: in practice, one with a separate significant Internet connectivity with different ISPs in many areas. Naturally, even smaller organizations are likely to have some relatively minor international activity, but these are not likely counted as International.

Now, the reasons and motivations specific to each organization class are described. These give a bit more elaboration on why the categories were written down as they were.

6.2.2 Minimal

Very minimal end-sites, such as typical home networks or very small enterprises, are quite small and typically do not include mission-critical activities.

Naturally, anyone would be willing to achieve multihoming benefits, but usually the associated costs, e.g. caused by obtaining physical connectivity to two ISPs, do not justify it.

In the very rare case that some multihoming is required, it can be done using the "Small" model.

6.2.3 Small

Small end-sites are typically small-to-mid-size enterprises, which operate mainly in one geographical location; branch offices are also possible, but these are typically handled using Virtual Private Networks (VPNs) or similar.

Small end-sites typically use the Internet in such a manner that they might find redundancy and independence important, depending on the costs and complexity. However, typically an outage of some minutes or rarely even hours is not yet critical, and due to the size, operator-independence is not vital.

6.2.4 Large

Large end-sites are typically largish enterprises which operate in one or typically more geographical locations, however, mostly inside a region or a country. Relatively minor international branch offices are possible, but there are no major internal, private interconnects – physical or link-layer connectivity that does not go through the Internet – to the offices in other countries. Typically, if there are multiple locations inside a region, there are internal, private interconnects between the locations.

Redundancy is very important due to increased size: longer outages are unacceptable for productivity. Also, as the number of the nodes in the site is quite high, the effort of renumbering is also high, and some level of operator independence valuable but not always strictly necessary if significant enough ISPs operate in the area. Only in some relatively rare cases the network capacities or other parameters are exceeded so that some form of load-sharing is required.

6.2.5 International

International end-sites are typically large or very large enterprises which have significant employment internationally, connect to the Internet with high-speed links in each of these significant locations and may often have internal, private interconnects in place.

Typically, ISPs are not global enough to supply connectivity to all the locations, and being locked to scenic routing paths caused by only a few regional ISPs is not considered a serious option.

As the international enterprise typically exceeds the geographic and topological scope of any single ISP, and the network is so large, operator independence is considered extremely important. Redundancy is also critical. The required capacities and usage patterns may vary a lot, but often also some form of load sharing is necessary: all the traffic to the end-site cannot go through a single location, and it might not be reasonable either, due to geography.

6.3 Methods for Choosing a Multihoming Mechanism

In previous sections, the IPv6 multihoming mechanisms have been analyzed and organizations split into four main classifications; now it's possible to summarize the mechanisms and see if particular mechanisms could be made to meet the organizations' needs.

The first thing to realize is that the scope and breadth of multihoming motivations and requirements vary a lot. It is highly unlikely that a "one size fits all" solution could be found: indeed, it seems fruitless to even try to look for one except possibly in purely research terms. Further, generic solutions are not probable to be found without larger architectural changes.

Therefore, by applying some rough classification one can hope to break the problem space into pieces and try to evaluate whether applicable solutions can be found for those pieces.

6.3.1 Viable Multihoming Mechanisms

First, it is important to take a look back at section 6.1 to see which mechanisms could be viable and how they could be positioned.

These are grouped to three categories: immediate, short-, and long-term. The definitions of the latter two are intentionally left vague, but short term solutions should not take more than 1-3 years to implement and deploy, at most.

Immediate Approaches

Multi-connecting seems to be the only obvious way to work around multihoming problems right now.

Host-centric multihoming and multihoming at site exit routers would also be applicable immediately, if no ISP would implement ingress filtering.

Short-term Approaches

Host-centric IPv6 multihoming and multihoming at site-exit routers, fully fleshed out, are both short-term solutions; both still need some work.

Provider independent addressing based on AS numbers is a short-term solution; the same applies to advertising more specific routes, if done from specific allocations to make them distinguishable.

Parts of multihoming with route aggregation, traffic steering using routing header and router renumbering could be used in the short term, but the mechanisms themselves are not usable.

Long Term Approaches

All transport solutions are only generally applicable in the long term, if at all.

Identifier and locator separation solutions are long-term solutions; HIP most credible of them all. LIN6 has fundamental issues which does not make it globally deployable. Mobile IPv6, if the address ownership and key management issue could be worked around, could be a short-term solution, a "hack". Architecturally, it seems a bit ill-fit as a long-term solution.

Geographic address allocation is a long-term solution if it's considered viable.

End-to-end multihoming is a long-term solution; it requires some major changes in thinking, and may be difficult to accept.

MHAP is a long-term solution; it moves into an area where most do not want to go architecturally, and it's likely it could not be made to work in practice.

Conclusion on Approaches

In the following analysis, none of the long-term approaches are seriously considered; they all need much more work to be adoptable.

Of the long-term solutions, it seems likely that at least identifier and locator solutions will prevail in some form or another. However, they do not solve the whole multihoming problem themselves.

6.3.2 Applicable Mechanisms for Organization Types

Now, the mechanisms outlined in the multihoming analysis and considered viable, above, are applied to the rough end-site organization categories above.

Minimal

Minimal end-sites do not seem to require a multihoming mechanism. Of course, some would still find one preferable.

If they would, one building on "host-centric multihoming" approach would seem sufficient; this might not even be all too expensive using e.g. multiple xDSL connections.

Small

Small end-sites might find a multihoming mechanism desirable for redundancy and independence reasons.

The possible approaches are three-fold:

1. multi-connecting to one ISP, if redundancy is important,
2. host-centric multihoming, if independence is important, or
3. multihoming at site exit routers, if both are important.

Multi-connecting or multihoming at site exit routers provide a very extensive amount of redundancy and convergence; independence is obtained by connecting to multiple ISPs and deploying IP addresses from those, which would make the renumbering much less of a problem if it had to be done.

Solving the multihoming problem of small end-sites with an approach like "ASN-PI" or more specific routes is unscalable.

Large

Large end-sites typically require redundancy and possibly independence, sometimes desiring load sharing as well.

Often, the same mechanisms applied to small end-sites will also provide the sufficient level of redundancy and independence for also large end-sites.

One might be able to handle the scalability issues associated with approaches like "ASN-PI" with large end-sites, but such mechanisms should be avoided.

International

International end-sites typically require redundancy, independence, and load sharing.

As their geographical scope typically exceed ISPs', assigning addresses from multiple ISPs, as with previous approaches, would lead to suboptimal routing.

One approach would be to depend on a change of the IP addressing and routing model: each significant international part of the organization would obtain separate

IP address assignments from the local ISPs, and use an appropriate multihoming mechanism with that – probably like with large end-sites above – and interconnect the offices using mechanisms like VPN's; not even trying to obtain a homogeneous address space for all of the company.

This would be a "divide-and-conquer" approach to the problem: break it to smaller pieces and solve them independently.

If this approach can't be adopted, it seems the only realistic model would be to use something like "ASN-PI", more specific routes or separate address allocations, or similar.

Chapter 7

Conclusions

First, the multihoming subject was introduced and the terminology, the scope and the motivations were clarified. Then, background information and the data collection environment were described. Third, site multihoming in particular was elaborated. Next, current IPv4 multihoming practices were analyzed. Last, the situation with IPv6 multihoming was analyzed and considered. Now, the conclusions and future work are presented.

Site multihoming is a very difficult issue which poses a scalability problem for the current routing system if one tries to solve it there, and not by using some other mechanisms.

Based on the analysis, there seem to be three-four main mechanisms which are used to achieve at least some of the multihoming benefits in IPv4: obtaining their own address space and an AS number and advertising those, advertising more specific routes with a different path, using multi-connecting and leveraging NAT. The first two which are roughly measurable constitute about 6-8% of Finnish prefixes.

In IPv6, the first two mechanisms which are considered architecturally unscalable have been operationally prevented for now, while the fourth does not exist; therefore, many have felt that IPv6 does not offer a site multihoming solution, at least one they could adopt.

Measurements and operator query reveal that the situation in IPv4 is rather chaotic: many use mechanisms which are the easiest for them, with little regard to the global routing system. Avoiding renumbering when changing ISPs seems to be the number one priority, having redundancy and other benefits of operator-independence coming close behind. Now with IPv6, there is an opportunity to do this properly from the start, without historical weight.

Many solutions and ideas for a different multihoming model have been presented, but there has been a lack of consensus on how to proceed. In this thesis, based on literature analysis and the experience learned from IPv4, I present a possible approach how to focus the work by analyzing the strengths and weaknesses of many

multihoming proposals and by creating a roadmap on how to proceed with IPv6 site multihoming in the short term while continuing the research and study of long-term approaches.

It is apparent that only a limited amount of work is needed to enable sufficiently good multihoming mechanisms which should provide enough features to satisfy the requirements: multi-connecting, multihoming at site exit routers and possibly separate address assignments. However, as the mechanisms are unarguably more difficult for the end-site, while taking the whole Internet better into account, whether they might be adopted remains to be seen.

7.1 Related Work

There are few if any similar studies on the multihoming impact on the routing tables.

Concurrently, other work [51] which focuses on the more generic analysis of route advertisements, was pending publication; feedback has been exchanged on both sides.

The more generic issue of analyzing routing tables without any specific focus has been explored before, for example in [52] and its unofficial successor [50]. The latter also lists some other interesting related work. [50] provides very interesting characteristics of the routing table, and studying it is highly recommended. The results, where applicable, are similar to the ones I've noted in the data analysis.

[53] is another source which provides insight into the development of the routing tables; some analysis and reports have also been done based on the data. In addition to the typical features, it also provides an easy access to historical route advertisement information.

7.2 Future Work

As always, there is a lot of work to be done. This is separated in two separate subjects, how the work here could be extended and what seems to be needed in the IPv6 site multihoming.

Following Work

Node multihoming and ISP multihoming were decreed out of scope. Focusing on the first might give some insight whether there are some missing pieces as the mobility and node multihoming become more important all the time. On the other hand, it is believed that focusing on ISP multihoming in general might not give all that much information – which was why it was defined out of scope – but by concentrating on some specific subject, it still might be possible to get some new insight on methods and possibilities.

More scientific study of multihoming motivations, as described in section 2.4, might be appropriate. The classification was done based on a first draft of a work in progress, enhanced by operational experience and strengthened by the analysis of IPv4 multihoming practices. In particular, the possible effect of performance and policy reasons would use more study. Also, the taxonomy could be rethought: the border between load sharing and performance, for example, seems to be rather blurry.

The data processing was slightly inaccurate and unscalable for larger use. The most difficult factor was being able to eliminate ISP multihoming from the results so one could only examine site multihoming. This was not trivial, and a redesign and rewrite of the processing algorithms would have been necessary to cope with the problem completely. That was not done, as this level of detail and error margin seemed sufficient. Moreover, the algorithms and mechanisms used – including but not limited to the use of the Perl programming language, made the processing system very heavy; it is completely unsuitable for analyzing significantly larger data amounts, such as the whole Internet routing table, but then again, this was a known engineering trade-off from the beginning: the method didn't have to be perfect to be usable in this context.

The data collection process could be improved; instead of taking snapshots of the data, BGP update and withdrawal messages could be monitored. Sooner or later a path will go down. If this is observed long enough, some of the otherwise-undetected failure modes will be noticed; now such would only be noticeable if a failure is occurring when the data snapshot is taken.

In particular, in cases described in sections 5.1.2 and 5.1.3, this could give additional insight in this particular observation point.

With real-time monitoring, when examining more and less specific routes for multihoming, if the destination is reachable through the less specific route when the more specific – through a different network service provider – is nonoperational, this is a sign of some real multihoming contract between the two. This could be observed e.g. by sending ICMP echo messages to the destination when a failure occurs and checking whether there are replies, and if so, from where.

Another approach to gain some more data might have been to analyze the routing policies stored in the routing registries, as described in section 3.1.5, to gain full knowledge of all routes, not just those visible at a particular time. This would have been an especially interesting approach if one could assume the data entered there would be complete and consistent.

The introduction of the generic scalability problem of site multihoming, in section 4.1, was just an introduction – on purpose. The topic would warrant a much more detailed analysis, being a paper or thesis on its own right. In particular, the possible impact of network capacity was considered to be a non-important factor; this might not be true under all circumstances.

It was also considered to send a short query to the administrative contacts of those sites which seem to be using some "Possibly Multihomed" mechanisms (5.1.2), but

this did not seem to be worthwhile when a query had already been sent to their ISPs. It seemed probable that most end-sites would not even have had any idea what the query was all about. However, doing something like that could shed some more light into current practices especially in the unclear cases.

IPv4 multihoming mechanisms such as [24] were mentioned but they were not believed to be in much use. This belief could also be wrong. The situation should be examined at more length, possibly relating to the query, above.

Also, as the route advertisements had been saved for the duration of more than half a year, some time development analysis was possible. However, only one particular subject was quickly examined in a rather unscientific way, just to get some rough results. A more detailed analysis both in the short term and longer multihoming trends could be useful to properly understand the trends and the future of multihoming.

Last, it is not clear whether classification of organizations and their requirements is necessarily the best one: such classifications and needs should be analyzed separately at more length.

IPv6 Site Multihoming

The most important of all, the IPv6 site multihoming work needs to finish the documentation of existing practices, practical requirements for multihoming, et cetera to be able to properly address the challenges of multihoming.

Soon, a roadmap will be needed on how to proceed; that is, whether to pursue shorter term goals or aim for long-term approaches immediately, and if so, which of them. The work in this thesis serves as one possible starting point for this work.

Some shorter term approaches need more work; typically not much, but some. In particular, techniques using multiple addresses from different ISPs need to tackle the problem of source address selection. For mechanisms aimed for bigger organizations, such as the "ASN-PI" or "longer prefixes" model, there must be consensus which mechanism would be most appropriate, and more importantly, how the limiting of the organizations "privileged" for such methods could be done.

Long-term approaches need much more work; as the amount of work needed is very high, it should suffice for now, without going into details, that all of them need more attention yet.

Additionally, it is very important to start documenting the process on how to make renumbering as easy as possible; in particular, how to avoid the use of IP addresses in places where changing them would be difficult. It is unrealistic to believe renumbering would become trivial, but if one is able to specify operational procedures and mechanisms which make it easier, the sites are more likely to accept multihoming mechanisms which depend on multiple addresses.

Bibliography

[1] Joe Abley, Benjamin Black, and Vijay Gill. IPv4 Multihoming Motivation, Practices and Limitations, Jun 2001. draft-ietf-multi6-v4-multihoming-00.txt. Expired.

[2] J. Noel Chiappa. A message on multi6 mailing-list on 2.1.2003. http://ops.ietf.org/lists/multi6/multi6.2002/msg01048.html. Referred 8.1.2003.

[3] David B. Johnson, Charles E. Perkins, and Jari Arkko. Mobility Support in IPv6, Feb 2003. draft-ietf-mobileip-ipv6-21.txt. Work in progress.

[4] Robert G. Moskowitz and et al. The Host Identity Payload Homepage. http://homebase.htt-consult.com/HIP.html. Referred 2.1.2003.

[5] Masahiro Ishiyama and et al. LINA: A New Approach to Mobility Support in Wide Area Networks. *IEICE Transactions on Communications*, E84-B(8):2076–2086, 2001.

[6] Yakov Rekhter and Tony Li. A Border Gateway Protocol 4. RFC1771, Mar 1995. Draft Standard.

[7] Ravi Chandra, Paul Traina, and Tony Li. BGP Communities Attribute. RFC1997, Aug 1996. Proposed Standard.

[8] Yakov Rekhter and Tony Li. An Architecture for IP Address Allocation with CIDR. RFC1518, Sep 1993. Proposed Standard.

[9] Robert Hinden, Mike O'Dell, and Steve Deering. An IPv6 Aggregatable Global Unicast Address Format. RFC2374, Jul 1998. Proposed Standard.

[10] Alain Durand and Christian Huitema. The Host-Density Ratio for Address Assignment Efficiency: An update on the H ratio. RFC3194, Nov 2001. Informational.

[11] Cengiz Alaettinoglu and et al. Routing Policy Specification Language (RPSL). RFC2622, Jun 1999. Proposed Standard.

[12] David Meyer and et al. Using RPSL in Practice. RFC2650, Aug 1999. Informational.

[13] Merit Networks. Routing Assets Database project. http://www.radb.net. Referred 15.1.2003.

[14] RIPE NCC. RIPE Whois Database. http://www.ripe.net/ripencc/pub-services/db/. Referred 15.1.2003.

[15] RIPE NCC. Internet Routing Registry Toolset Project. http://www.ripe.net/ripencc/pub-services/db/irrtoolset/. Referred 15.1.2003.

[16] Pyda Srisuresh and Kjeld Egevang. Traditional IP Network Address Translator. RFC3022, Jan 2001. Informational.

[17] Yakov Rekhter and et al. Address Allocation for Private Internets. RFC1918, Feb 1996. Best Current Practise 5.

[18] CSC. Funet Network Connections 2002. http://www.csc.fi/suomi/funet/verkko.html.en. Referred 30.1.2003.

[19] Jorma Mellin. FICIX presentation in RIPE-41 EIX, Jan 2002. http://www.ficix.fi/tiedotteet/ficix-eix-41.ppt. Referred 30.1.2003.

[20] Aki Anttila. Private email discussion on 30.1.2003.

[21] Statistics Finland. Finland in Figures - Enterprises. http://www.stat.fi/tk/tp/tasku/taskue_yritykset.html. Year 2000 data. Referred 3.3.2003.

[22] U.S. Census Bureau. Statistics of U.S. Businesses. http://www.census.gov/csd/susb/susb2.htm. Year 2000 data. Referred 3.3.2003.

[23] Praveen Akkiraju, Kevin Delgadillo, and Yakov Rekhter. Enabling Enterprise Multihoming with Cisco IOS Network Address Translation (NAT). http://www.cisco.com/warp/public/cc/pd/iosw/ioft/ionetn/tech/emios_wp.htm. Referred 2.1.2003.

[24] Tony Bates and Yakov Rekhter. Scalable Support for Multi-homed Multi-provider Connectivity. RFC2260, Jan 1998. Informational.

[25] Radware. LinkProof - Internet Link Application Switching product whitepaper. http://www.radware.com/content/products/link.asp. Referred 2.1.2003.

[26] FatPipe. The WARP multihoming product whitepaper. http://www.fatpipeinc.com/warp/. Referred 2.1.2003.

[27] Martin Dunmore and Christopher Edwards (eds.). Report on IETF Multihoming Solutions, Oct 2002. http://www.6net.org/publications/deliverables/D4.5.1.pdf. EU IST-2001-32603 6NET project deliverable 4.5.1.

[28] Frank Kastenholz (ed.). Requirements For a Next Generation Routing and Addressing Architecture, Apr 2002. draft-irtf-routing-reqs-groupa-00.txt. Expired.

[29] Peter R. Tattam. Preserving Active TCP sessions on Multihomed IPv6 Networks, Aug 2001. http://jazz-1.trumpet.com.au/ipv6-draft/preserve-tcp.txt. Referred 24.2.2003.

[30] Randall R. Stewart and et al. Stream Control Transmission Protocol. RFC2960, Oct 2000. Proposed Standard.

[31] Lode Coene (ed.). Multihoming issues in the Stream Control Transmission Protocol, Feb 2002. draft-coene-sctp-multihome-03.txt. Work in progress.

[32] Eliot Lear and Ralph Droms. What's In A Name: Thoughts from the NSRG, Mar 2003. draft-irtf-nsrg-report-09.txt. Work in progress.

[33] Fumio Tereoka and et al. LIN6: A Solution to Mobility and Multi-Homing in IPv6, Aug 2001. draft-teraoka-ipng-lin6-01.txt. Expired.

[34] Francis Dupont. Multihomed routing domain issues for IPv6 aggregatable scheme, Sep 1999. draft-ietf-ipngwg-multi-isp-00.txt. Expired.

[35] Christian Huitema and Richard Draves. Host-Centric IPv6 Multihoming, Jun 2002. draft-huitema-multi6-hosts-01.txt. Expired.

[36] Jun ichiro Hagino and Hal Snyder. IPv6 Multihoming Support at Site Exit Routers. RFC3178, Oct 2001. Informational.

[37] Tony Hain. Application and Use of the IPv6 Provider Independent Global Unicast Address Format, Feb 2003. draft-hain-ipv6-pi-addr-use-04.txt. Work in progress.

[38] Pekka Savola. Multihoming Using IPv6 Addressing Derived from AS Numbers, Jan 2003. draft-savola-multi6-asn-pi-00.txt. Work in progress.

[39] Kurt E. Lindqvist. Multihoming in IPv6 by Multiple Announcements of Longer Prefixes, Dec 2002. draft-kurtis-multihoming-longprefix-00.txt. Work in progress.

[40] Jieyun Yu. IPv6 Multihoming with Route Aggregation, Aug 2000. draft-ietf-ipngwg-ipv6multihome-with-aggr-01.txt. Expired.

[41] Masataka Ohta. The Architecture of End to End Multihoming, Nov 2002. draft-ohta-e2e-multihoming-03.txt. Work in progress.

[42] Michel Py. Multi Homing Aliasing Protocol, Apr 2002. draft-py-mhap-01a.txt. Expired.

[43] Mathew Crawford. Router Renumbering for IPv6. RFC2894, Aug 2000. Proposed Standard.

[44] Masataka Ohta. Root Name Servers with Inter Domain Anycast Addresses, Nov 2002. draft-ietf-dnsop-ohta-shared-root-server-02.txt. Work in progress.

[45] Craig Labovitz and et al. Understanding the Large-Scale Dynamics of Internet Routing Protocols. Workshop - The Global Internet: Measurement, Modeling and Analysis, Sep 2000. http://www.renesys.com/projects/leiden/, referred 2.1.2003.

[46] Donald E. Eastlake 3rd. Domain Name System Security Extensions. RFC2535, Mar 1999. Proposed Standard.

[47] Tatu Ylönen and et al. SSH Protocol Architecture, Sep 2002. draft-ietf-secsh-architecture-13.txt. Work in progress.

[48] Tom Killalea. Recommended Internet Service Provider Security Services and Procedures. RFC3013, Nov 2000. Best Current Practice 46.

[49] Richard Draves. Default Address Selection for IPv6. RFC3484, Feb 2003. Proposed Standard.

[50] Geoff Huston. BGP Table Data project. http://bgp.potaroo.net. Referred 12.3.2003.

[51] X.Meng, Z.Xu, L.Zhang, and S.Lu. An Analysis of BGP Routing Table Evolution. Technical Report, Computer Science Department, UCLA, Feb 2003.

[52] Tony Bates and Philip Smith. CIDR Report project. http://www.cidr-report.org. Referred 12.3.2003.

[53] RIPE NCC. RIPE Route Information Service. http://www.ripe.net/ris/. Referred 15.1.2003.

Appendix A

Query to ISPs

Hello,

I'm researching multihoming mechanisms by analyzing route advertisements from different ISPs in FICIX. The research is done to better understand the path to IPv6 multihoming, in particular, issues with end-site multihoming. For that purpose, background and reasons for multihoming in IPv4 is analyzed.

It would be greatly appreciated if you could spare a few minutes to fill in this query, and hopefully be able to explain a few route advertisements that relate to your association (below).

For some (not too extensive) background material on different route advertisements, a draft version of categorization can be found at:
[URL pointing to a version of the thesis removed]

Responses will be anonymized to sufficient detail so they cannot be connected to any organizations. The only intent is to gain insight into motives of route advertisements, not gather data on "market shares", "illegal route advertisers", or whatever.

QUERY
=====

(Please refer to the attached current route advertisements, as appropriate)

1. Advertising Identical Prefixes

   Discussion:

   For some purposes, for example some "anycast service", like 192.88.99.0/24, identical prefixes are sometimes advertised from different origin AS's. This could also be a form of multihoming, e.g. BGP with private AS-numbers or static routes set by the ISP. Currently, there are no such advertisements except for 192.88.99.0/24 in FICIX.

   (See "Possibly Multihomed - Identical Prefixes from a Different Origin")

   Question(s):

   Have you ever intentionally advertised an identical prefix as someone else for other than anycast purposes?

   If so, was it multihoming or something else -- what?

2. Advertising More Specifics from a Different Origin

   Discussion:

   Often, someone advertises a more specific route from a different origin AS than the less specific aggregate.

   It is the belief that most of these advertisements are due to customers

```
   moving from one ISP to another and taking their PA addresses with them
   instead of renumbering.

   However, this could also be a form of multihoming: the more specific is
   the main network connection, and the less specific aggregate will be used
   for backup.  One particular case here is when a customer changes ISPs
   but keeps the old one for backup (at least for a while).

   Our data collection over a 6 month cycle suggests only few of these
   more specific are temporary in nature.

   (See "Possibly Multihomed - More Specific Routes from Different Origin")

  Question(s):

   Could you summarize the reasons and/or legitimacy of:

    a) someone else advertising a more specific route to one of your networks
    b) you advertising a more specific route to someone else's networks

   Are you aware of this method being used for multihoming?

   How big a portion of these routes are for only "making PA addresses PI"?

   Are you aware how "temporary" these advertisements, in practice, are?

   If someone else is advertising a more specific route, do you also have
   a (secondary) network connection to that site, or do you accept the
   more specific routes to your aggregates from someone else?

3. Advertising More Specifics from the Same Origin

  Discussion:

   One related case to the above is when the same origin advertises both
   the less and more specific route.  This does not seem to have any
   multihoming implications.  Most of these are assumed to be either
   a form of traffic engineering, or humane configuration mistakes.

   (See "Unclear Cases - More Specific from the Same Path and Origin")

  Question(s):

   Are there reasons why you are advertising these routes?

   Do you use them for traffic-engineering purposes?

4. Methods and the Extent of Multi-Connecting

  Discussion:

   By some terminology, multihoming also includes multiple connections (for
   redundancy) to a single ISP.  This is generally undetectable to the
   outside observers.  Here, for "multi-connecting", it is required that
   the whole address space of the customer is routed to multiple
```

```
   connections (that is, branch offices with separate address spaces are
   not considered multi-connecting.)

   Typical ways to do multi-attachment are either:
    1) customer-ISP BGP sessions with (private) AS numbers
    2) customer-ISP OSPF/IS-IS adjacency for fail-over (e.g. "floating
       static routes") when ISP is in charge of CPE.
    3) manually configured fail-over

   (See "Unclear Cases - Multi-attachment inside an ISP")

  Question(s):

   What is the amount of multi-connecting with these mechanisms?

   Are you aware of other ways multi-connecting has been done?

   Do you know reasons why some customers have required "complete"
   multihoming (own AS numbers, two providers, two identical prefixes)
   and why this is enough for some?


5. Multihoming behind Neighbor ISPs

  Discussion:

   If multihoming is done between two smaller ISPs, and the upstream
   connectivity of both goes via you (no other routes), observing
   route advertisements in FICIX does not show more than one route, so data
   gets lost due to BGP only advertising the best path.

   (See "Unclear Cases - Multihoming behind Neighbor ISPs")

  Question(s):

   Are you aware of any smaller ISPs as your customers which are not
   connected to other ISPs?

6. Multihoming Using NAT

  Discussion:

   There are some solutions based on NAT which will achieve a limited grade
   of multihoming and provider-independence.  NAT is also commonplace for
   other reasons.

   (See "Unclear Cases - Some Degree of Multihoming Using NAT")

  Question(s):

   Are you aware of customers that have deployed NAT for:
    a) operator independence, and have switched operators easily using it, or
    b) connecting to more than one ISP (simultaneously) for redundancy,
      fail-over, or other such reasons?
```

```
    If so, can you estimate the number of these (respectively), and
    reasoning behind this model, if possible?

APPENDIX
========

Below are those unclear routes that your organization is in some way related
to.  If you can, please specify (possibly coupled with questions above),
what is the purpose of these advertisements.

In the first section, more/less specific routes with the same path are
printed.  These are assumed to be either mistakes or part of
traffic-engineering.

In the next section, more/less specific routes with different origin are
printed.  These are assumed to be "making PA addresses PI" or multihoming.
Mistakes e.g. in atomic aggregation are also possible.

In the third section, routes which haven't been received from other FICIX
peers but which have AS-path prepending are printed.  These may be
multihomed behind a neighbor ISP, traffic engineering, or other so foreign
networks that AS-path prepending is done for some non-FICIX purposes.

Last, routes which have been aggregated by a private AS number are printed.
These are assumed to be typically a product of multi-attachment inside the
ISP.
```

```
More specifics with different origin
------------------------------------

Prefix             Nexthop         MED Aggr? Communities AS path
157.124.0.0/16     212.226.101.146                       1759 5515
 157.124.16.0/21   212.226.101.105 400                   790 1738
 157.124.16.0/21   212.226.101.49  100                   719 1738
159.152.0.0/16     212.226.101.146     64578             1759 5515
 159.152.255.128/2 212.226.101.49  100                   719
192.130.0.0/16     212.226.101.146     5515              1759 5515
 192.130.143.0/24  212.226.101.105 400       1           790 20774
 192.130.143.0/24  212.226.101.122 100                   8434 6667 790 20774
 192.130.198.0/23  212.226.101.146                       1759 5515 764
 192.130.217.0/24  212.226.101.49  100                   719
 192.130.67.128/27 212.226.101.49  100                   719
192.163.32.0/19    212.226.101.49  100 719               719
 192.163.54.0/24   212.226.101.146                       1759 5515
192.194.0.0/16     212.226.101.146     5515              1759 5515
 192.194.114.0/24  212.226.101.49  100                   719
 192.194.252.0/24  212.226.101.146                       1759 5515 3274
 192.194.53.0/24   212.226.101.146                       1759
192.49.16.0/23     212.226.101.146                       1759 5515 764
 192.49.17.0/24    212.226.101.146                       1759 5515
192.49.30.0/23     212.226.101.146                       1759 5515 764
 192.49.31.0/24    212.226.101.146                       1759 5515
192.58.48.0/20     212.226.101.146                       1759 5515
 192.58.49.0/24    212.226.101.122 100                   8434 6667 790 719
 192.58.49.0/24    212.226.101.49  100                   719
 192.58.51.0/24    212.226.101.146                       1759 5515 3274
192.89.0.0/16      212.226.101.146     5515              1759 5515
 192.89.101.0/24   212.226.101.49  100                   719
 192.89.102.0/24   212.226.101.49  100                   719
 192.89.103.0/24   212.226.101.49  100                   719
 192.89.104.0/24   212.226.101.49  100                   719
 192.89.105.0/24   212.226.101.49  100                   719
 192.89.16.0/24    212.226.101.49  100                   719
 192.89.226.0/24   212.226.101.122 100                   8434 6667 790 719
 192.89.226.0/24   212.226.101.49  100                   719
 192.89.248.0/24   212.226.101.49  100                   719
 192.89.249.0/24   212.226.101.49  100                   719
 192.89.250.0/24   212.226.101.49  100                   719
 192.89.251.0/24   212.226.101.49  100                   719
 192.89.4.0/24     212.226.101.49  100                   719
193.184.0.0/15     212.226.101.49  100 719               719
 193.184.135.0/24  212.226.101.146                       1759 5515
 193.184.229.0/24  212.226.101.146                       1759 5515
 193.185.37.0/24   212.226.101.146                       1759 5515
193.208.0.0/16     212.226.101.146     5515              1759 5515
 193.208.66.0/23   212.226.101.146                       1759 5515 764
 193.208.69.0/24   212.226.101.146                       1759 5515 764
 193.208.71.0/24   212.226.101.146                       1759 5515 764
193.209.0.0/16     212.226.101.146     5515              1759 5515
 193.209.25.0/24   212.226.101.49  100                   719
193.210.0.0/16     212.226.101.146     5515              1759 5515
 193.210.240.0/24  212.226.101.49  100                   719
 193.210.241.0/24  212.226.101.49  100                   719
193.64.0.0/15      212.226.101.105 400 790   1           790
 193.64.10.0/24    212.226.101.146                       1759 5515
 193.64.100.0/24   212.226.101.146                       1759 5515
 193.64.11.0/24    212.226.101.146                       1759 5515
 193.64.12.0/24    212.226.101.146                       1759 5515
 193.64.128.0/22   212.226.101.122 100                   8434 6667 790 719
 193.64.128.0/22   212.226.101.49  100                   719
 193.64.140.0/23   212.226.101.122 100                   8434 6667 790 719
 193.64.140.0/23   212.226.101.49  100                   719
 193.64.158.0/23   212.226.101.150                       3246
 193.64.168.0/23   212.226.101.146                       1759 5515
```

```
 193.64.186.0/24   212.226.101.49  100                 719
 193.64.199.0/24   212.226.101.49  100                 719
 193.64.28.0/23    212.226.101.146     64572           1759 5515
 193.64.64.0/20    212.226.101.146                     1759 5515 16044
 193.64.84.0/24    212.226.101.146                     1759 5515 1342
 193.64.87.0/24    212.226.101.146                     1759 5515
 193.65.128.0/23   212.226.101.146                     1759 5515
 193.65.144.0/22   212.226.101.122 100                 8434 6667 790 719
 193.65.144.0/22   212.226.101.49  100                 719
 193.65.148.0/22   212.226.101.146                     1759 5515 3274
 193.65.173.0/24   212.226.101.49  100 64518           719
 193.65.248.0/24   212.226.101.118 50                  16086 16086 12375
 193.65.248.0/24   212.226.101.150                     3246 12375
 193.65.91.0/24    212.226.101.122 100                 8434 6667 790 719
 193.65.91.0/24    212.226.101.49  100                 719
194.110.32.0/20    212.226.101.105 400 790   1         790
 194.110.38.0/24   212.226.101.146                     1759 5515
 194.110.44.0/22   212.226.101.146                     1759 5515
194.111.0.0/16     212.226.101.146     5515            1759 5515
 194.111.121.0/24  212.226.101.49  100                 719
 194.111.122.0/24  212.226.101.146                     1759 5515 3274
194.136.0.0/16     212.226.101.49  100 719             719
 194.136.32.0/19   212.226.101.146                     1759 5515
 194.136.72.0/23   212.226.101.105 400 790   1         790
 194.136.72.0/23   212.226.101.122 100 790             8434 6667 790
194.137.0.0/16     212.226.101.146     5515            1759 5515
 194.137.11.0/24   212.226.101.105 400       1         790
 194.137.11.0/24   212.226.101.122 100                 8434 6667 790
 194.137.159.0/24  212.226.101.105 400       1         790
 194.137.159.0/24  212.226.101.122 100                 8434 6667 790
 194.137.56.0/24   212.226.101.105 400       1         790 20774
 194.137.56.0/24   212.226.101.122 100                 8434 6667 790 20774
194.142.0.0/16     212.226.101.146     5515            1759 5515
 194.142.12.0/24   212.226.101.49  100                 719
 194.142.13.0/24   212.226.101.49  100                 719
 194.142.14.0/24   212.226.101.49  100                 719
 194.142.15.0/24   212.226.101.49  100                 719
194.157.0.0/16     212.226.101.49  100 719             719
 194.157.126.0/23  212.226.101.105 400       1         790
 194.157.126.0/23  212.226.101.122 100                 8434 6667 790
 194.157.232.0/21  212.226.101.146                     1759 5515
194.188.0.0/16     212.226.101.49  100 719             719
 194.188.145.0/24  212.226.101.146     5515            1759 5515
 194.188.30.0/24   212.226.101.105 400       1         790
 194.188.30.0/24   212.226.101.122 100                 8434 6667 790
194.197.0.0/16     212.226.101.146     5515            1759 5515
 194.197.125.0/24  212.226.101.146                     1759 5515 3274
 194.197.155.0/24  212.226.101.105 400       1         790
 194.197.155.0/24  212.226.101.122 100                 8434 6667 790
194.211.0.0/16     212.226.101.49  100 719             719
 194.211.231.0/24  212.226.101.146                     1759 5515 764
194.215.0.0/16     212.226.101.146     5515            1759 5515
 194.215.244.0/24  212.226.101.150                     3246
 194.215.50.0/24   212.226.101.49  100                 719
194.240.0.0/15     212.226.101.49  100 719             719
 194.240.93.0/24   212.226.101.146                     1759 5515
194.251.0.0/16     212.226.101.146     5515            1759 5515
 194.251.182.0/24  212.226.101.150                     3246
 194.251.183.0/24  212.226.101.150                     3246
 194.251.91.0/24   212.226.101.105 400                 790 3274
 194.251.91.0/24   212.226.101.146                     1759 5515 3274
 194.251.91.0/24   212.226.101.150                     3246 3274
194.86.0.0/16      212.226.101.49  100 719             719
 194.86.211.0/24   212.226.101.105 400       1         790
 194.86.211.0/24   212.226.101.122 100                 8434 6667 790
 194.86.93.0/24    212.226.101.146                     1759 5515
 194.86.94.0/24    212.226.101.146                     1759 5515
```

```
194.89.0.0/16      212.226.101.146     5515              1759 5515
 194.89.244.0/24   212.226.101.49  100                  719
 194.89.91.0/24    212.226.101.49  100                  719
195.156.0.0/16     212.226.101.146     5515              1759 5515
 195.156.109.0/24  212.226.101.146                       1759 5515 5473
 195.156.110.0/23  212.226.101.146                       1759 5515 3274
195.84.0.0/16      212.226.101.150     3246              3246
 195.84.157.0/24   212.226.101.146                       1759
 195.84.16.0/24    212.226.101.146                       1759
212.94.64.0/19     212.226.101.146     5515              1759 5515
 212.94.80.0/23    212.226.101.122 100                   8434
 212.94.82.0/23    212.226.101.150                       3246
213.28.0.0/16      212.226.101.146     5515              1759 5515
 213.28.120.0/24   212.226.101.105 400       1           790 24714

More specifics with the same path and origin
--------------------------------------------

Prefix             Nexthop         MED Aggr? Communities AS path
139.157.0.0/16     212.226.101.146                       1759 5515
 139.157.192.0/21  212.226.101.146                       1759 5515
157.144.0.0/16     212.226.101.146                       1759 5515
 157.144.251.0/24  212.226.101.146                       1759 5515
 157.144.252.0/22  212.226.101.146                       1759 5515
192.130.0.0/16     212.226.101.146     5515              1759 5515
 192.130.156.0/24  212.226.101.146                       1759 5515
 192.130.46.0/24   212.226.101.146                       1759 5515
192.49.150.0/23    212.226.101.146                       1759 5515
 192.49.150.0/24   212.226.101.146                       1759 5515
192.49.20.0/23     212.226.101.146                       1759 5515 764
 192.49.20.0/24    212.226.101.146                       1759 5515 764
 192.49.21.0/24    212.226.101.146                       1759 5515 764
192.49.34.0/23     212.226.101.146                       1759 5515 764
 192.49.34.0/24    212.226.101.146                       1759 5515 764
192.58.44.0/22     212.226.101.146     5515              1759 5515
 192.58.46.0/24    212.226.101.146                       1759 5515
192.83.16.0/20     212.226.101.146                       1759 5515
 192.83.25.0/24    212.226.101.146                       1759 5515
193.208.0.0/16     212.226.101.146     5515              1759 5515
 193.208.75.0/24   212.226.101.146                       1759 5515
 193.208.88.0/24   212.226.101.146                       1759 5515
 193.208.66.0/23   212.226.101.146                       1759 5515 764
 193.208.67.0/24   212.226.101.146                       1759 5515 764
193.211.0.0/16     212.226.101.146     5515              1759 5515
 193.211.44.0/24   212.226.101.146                       1759 5515
194.137.0.0/16     212.226.101.146     5515              1759 5515
 194.137.237.0/24  212.226.101.146     64558             1759 5515
213.180.192.0/19   212.226.101.146                       1759 5523 5523 5523 5523 5523 5523 5523 13238
 213.180.192.0/20  212.226.101.146                       1759 5523 5523 5523 5523 5523 5523 5523 13238
 213.180.208.0/20  212.226.101.146                       1759 5523 5523 5523 5523 5523 5523 5523 13238
217.174.96.0/20    212.226.101.146     20655             1759 20655
 217.174.96.0/21   212.226.101.146                       1759 20655
217.74.128.0/19    212.226.101.146     1759              1759
 217.74.128.0/20   212.226.101.146     1759              1759
80.253.192.0/20    212.226.101.146                       1759 21482
 80.253.193.0/24   212.226.101.146                       1759 21482

Prepending with only one path
-----------------------------

Prefix             Nexthop         MED Aggr? Communities AS path
192.188.189.0/24   212.226.101.146                       1759 15756 5467 5467
193.124.156.0/24   212.226.101.146                       1759 5523 5523 5523 5523 5523 5523 5523
193.125.142.0/23   212.226.101.146                       1759 15756 5467 5467
194.226.128.0/20   212.226.101.146                       1759 8825 8825 8825
194.67.128.0/18    212.226.101.146                       1759 5523 5523 5523 5523 5523 5523 5523
194.67.224.0/19    212.226.101.146                       1759 5523 5523 5523 5523 5523 5523 5523
```

```
194.8.160.0/19     212.226.101.146                    1759 24919 24919 24919 24919 24919 6690
194.85.80.0/22     212.226.101.146                    1759 15756 5467 5467
195.131.0.0/16     212.226.101.146                    1759 24919 24919 24919 24919 24919 6690
195.134.224.0/19   212.226.101.146                    1759 5515 8812 8812 8812
195.201.0.0/16     212.226.101.146                    1759 8377 8377 8377 8377
195.234.208.0/22   212.226.101.146                    1759 25258 25258 25258 25258
195.242.0.0/19     212.226.101.146                    1759 8377 8377 8377 8377
195.49.208.0/21    212.226.101.146                    1759 8825 8825 8825 5434 8430
195.82.0.0/19      212.226.101.146                    1759 8825 8825 8825 5434
212.13.128.0/19    212.226.101.146                    1759 8825 8825 8825
212.13.160.0/19    212.226.101.146                    1759 8825 8825 8825
212.58.192.0/19    212.226.101.146                    1759 24919 24919 24919 24919 24919 6690 12380
212.73.96.0/19     212.226.101.146                    1759 20576 20576 12979 12979 12979 12979 [...]
213.140.224.0/19   212.226.101.146                    1759 25515 25515
213.141.160.0/20   212.226.101.146                    1759 8825 8825 8825 5434 20578
213.141.176.0/21   212.226.101.146                    1759 8825 8825 8825 5434 20578
213.141.184.0/21   212.226.101.146                    1759 8825 8825 8825
213.156.128.0/19   212.226.101.146                    1759 20576 20576 12979 12979 12979 12979 [...]
213.158.0.0/19     212.226.101.146                    1759 24919 24919 24919 24919 24919 6690 [...]
213.170.64.0/19    212.226.101.146                    1759 12418 12418 12418
213.170.96.0/20    212.226.101.146                    1759 12418 12418 12418
213.180.192.0/19   212.226.101.146                    1759 5523 5523 5523 5523 5523 5523 5523 13238
  213.180.192.0/20 212.226.101.146                    1759 5523 5523 5523 5523 5523 5523 5523 13238
  213.180.208.0/20 212.226.101.146                    1759 5523 5523 5523 5523 5523 5523 5523 13238
213.211.64.0/18    212.226.101.146                    1759 8825 8825 8825
213.252.64.0/18    212.226.101.146                    1759 5523 5523 5523 5523 5523 5523 5523
213.33.242.0/24    212.226.101.146                    1759 24919 24919 24919 24919 24919 6690 25185
217.15.176.0/20    212.226.101.146                    1759 8825 8825 8825 5434 25534
217.175.128.0/19   212.226.101.146                    1759 15756 13230 20702 20702 20702 20702
217.195.96.0/20    212.226.101.146     20690          1759 24919 13257 20690 20690 20690 20690
81.13.0.0/17       212.226.101.146                    1759 5523 5523 5523 5523 5523 5523 5523
81.28.0.0/20       212.226.101.146                    1759 20576 20576 12979 12979 12979 12979 [...]
81.5.64.0/18       212.226.101.146                    1759 15756 5467 5467 25100

Aggregated by private ASN
-------------------------

Prefix             Nexthop         MED Aggr? Communities AS path
159.152.0.0/16     212.226.101.146     64578             1759 5515
  193.64.28.0/23   212.226.101.146     64572             1759 5515
  194.137.237.0/24 212.226.101.146     64558             1759 5515
```