J. Abley
Internet Systems Consortium, Inc.
B. Black
Layer8 Networks
V. Gill
AOL
K. Lindqvist
Netnod Internet Exchange

October 17, 2004

IPv4 Multihoming Motivation, Practices and Limitations
draft-ietf-multi6-v4-multihoming-02

Status of this Memo

Abstract

Multihoming is an essential component of service for sites which are part of the Internet. This draft describes some of the motivations,

practices and limitations of multihoming as it is achieved in the IPv4 world today. The analysis is done in order to serve as underlaying documentation to the discussions in the "Site multihoming for IPv6" working group of the IETF, who are working to a longer term solution to some of the issues that arise from doing multihoming in the ways as are described here.

1. Introduction

Multihoming is an important component of service for many sites which are part of the Internet. Current IPv4 multihoming practices have been added on to the CIDR architecture [1], which assumes that routing table entries can be aggregated based upon a hierarchy of customers and service providers.

Multihoming is a mechanism by which sites can currently satisfy a number of high-level requirements, and is widely used in the IPv4 network today. There are some practical limitations, however, including concerns of how well (or, if) the current practice will scale as the network continues to grow.

The preferred way to multihome in IPv4 is to announce an independent block of address space over two or more ISPs using BGP. Until the mid-1990s this was relatively easy to accomplish, as the maximum generally accepted prefix length in the global routing table was a /24, and little justification was needed to receive a /24. However, RIR policies today do not generally support the assignment of netblocks to small multi-homed end-users (e.g. those who might only be able to justify a /24 assignment), and as a consequence provider-independent (PI) space is not available to many sites who wish to multi-home.

An alternative way to multihome in IPv4 is to get address space allocated to the site from a upstream service provider. The site can then subscribe to a second Internet connection from a second service provider, and ask that the second service provider either announce or accept the announcement over BGP of the address block allocated to the site from the first service provider.

This second practice has two advantages and two disadvantages for the multihomed site, as well for the Internet at large. The first advantage is that this practice is applicable to sites whose addressing requirements are not sufficient to meet the requirements for PI assignments by RIRs. The second advantage is that even if the more specific announcement is filtered, they are still reachable over the primary ISP by virtue of the aggregate announced by this ISP. Even when the circuit to the primary ISP is down, this often works because the primary ISP will generally accept the announcement over the secondary ISP, so traffic flows from the filtering network to the primary ISP and then to the secondary ISP in order to arrive at the multihomed network. While this is common, it is also not universally true.

The first disadvantage with this approach is that the multihomed network must depend on the primary ISP for the aggregate. If the

primary ISP goes down, this will impact reachability to networks that filter. Most ISPs will cooperate with this "punching holes in an aggregate" solution to multihoming, but some are reluctant. And when the multihomed network leaves the primary ISP, they are generally expected to return the address space because otherwise this ISP would have to route traffic for a non-customer. The second disadvantage with the approach of announcing more specific routes out of a PA block, is that if the site were to change service provider away from the provider that has allocated the PA block, they will have to renumber. It is worth noting that the site can change all the other service providers without having to renumber.

2. Terminology

A "site" is an entity autonomously operating a network using IP, and in particular, determining the addressing plan and routing policy for that network. This definition is intended to be equivalent to "enterprise" as defined in [2].

A "transit provider" operates a site that directly provides connectivity to the Internet to one or more external sites. The connectivity provided extends beyond the transit provider's own site. A transit provider's site is directly connected to the sites for which it provides transit.

A "multihomed" site is one with more than one transit provider. "Site-multihoming" is the practice of arranging a site to be multihomed.

The term "re-homing" denotes a transition of a site between two states of connectedness due to a change in the connectivity between the site and its transit providers' sites.

A "multi-attached" site is one with more than one point of layer-3 interconnection to a single transit provider.

3. Motivations for Multihoming

There exists a great wealth of reasons why any single person or entity would like to multihome their network, in one way or the other. The reasons for doing this, are to achieve one or more of the goals as outlined in [5]. A more detailed analysis of the reasoning behind multihoming configurations are considered out-of-scope for this document.

4. Current methods used for IPv4 multihoming

There are a number of ways that a site which wishes to become multihomed can achieve its objectives today using IPv4. These methods can broadly be split into five categories as described below.

4.1 Multihoming with PI addresses and AS

The site uses provider-independent (PI) addresses assigned by an RIR. The routes corresponding to the PI addresses are announced with an origin AS associated with the multi-homed site to two or more transit providers.

4.2 Multihoming with your own AS, but PA addresses

The site uses provider-aggregatable (PA) addresses assigned by one transit provider. The routes corresponding to those PA addresses are announced with an origin AS associated with the multi-homed site to two or more transit providers. One of those transit providers originates a covering supernet for the site's routes.

4.3 Multihoming with your own addresses, and private AS

Another possible way of multihoming is with addresses owned by the site wishing to multihome, but advertising them without having a public AS, or with no AS at all. This is done with the site either sourcing the prefixes in a private AS [3], and having their upstreams remove those on announcement to the rest of the world, or the upstreams simply sourcing the prefixes in their AS and then routing to the organization.

4.4 Multiple attachments to the same ISP

Fourth option is to have multiple connection to the same ISP. This is fairly popular, but will not have an impact on the global routing table as both paths are covered by the ISPs aggregate route. A site that have solved their multihoming needs in this way is commonly referred to as "multi-attached".

4.5 NAT or RFC2260 based multihoming

This last method might very well be the most commonly used method in terms of volume. Simply because this is what most residential users are normally referred to. However, this method is also in use by very large enterprises, as this is an easy way for large enterprise networks to avoid advertising multiple prefixes when they connect at multiple locations.

This method uses addresses from each of the upstream that an organization is connected to. Either the addresses are allocated to nodes inside the network according to the proposal in [4], or the site uses NAT to translate into private addresses inside the site.

5. Features of IPv4 Multihoming

The following section analyzes some of the features driving the choices for various multihoming approaches in todays IPv4 Internet. As the "Site multihoming for IPv6" working group progresses, they will have to take similar considerations into account, learning from IPv4. These considerations are listed in [5], and some of the operational considerations that needs to be thought of for new multihoming mechanisms can be found in [6]. Not all approaches described in this document will support all the features listed below. Particularly solutions based on RFC2260/NAT [4] based designs will support a significantly smaller subset of these features.

5.1 Simplicity

The current methods used as multihoming solutions are not all without complexity, but in practice it is quite straightforward to deploy and maintain by virtue of the fact that they are well-known, tried and tested.

5.2 Transport-Layer Survivability

The current multihoming solution provides session survivability for transport-layer protocols in most cases; i.e. exchange of data between devices on the multi-homed site network and devices elsewhere on the Internet may proceed with no greater interruption than that associated with the transient packet loss during a re-homing event. However, there are cases where the current multihoming solutions does not provide transport-layer survivability. One example is that due to the large number of ASes in the current Internet routing table, BGP convergence take longer and longer and the transport-layer might time-out while waiting for BGP to converge. There are also BGP implementations in wide use that will take a long time for failure detection and that might cause TCP timeouts.

In the case a re-homing event occurs, new transport-layer sessions are able to be created.

5.3 Inter-Provider Traffic Engineering

A multi-homed site may influence routing decisions beyond its immediate transit providers by advertising a strategic mixture of carefully-aimed long prefixes and covering shorter-prefix routes. This precise effects of such egress policy are often difficult to predict, but an approximation of the desired objective is often easy to accomplish.

5.4 Load Control

The current multihoming solution places control of traffic flow in the hands of the site responsible for the multi-homed interconnections with transit providers. A single-homed client a multi-homed site may vary the demand for traffic that they impose on the site, and may influence differential traffic load between transit providers; however, the basic mechanisms for congestion control and route propagation are in the hands of the site, not the client.

5.5 Impact on Routers

The routers at the boundary of a multi-homed site are usually required to participate in BGP sessions with the interconnected routers of transit providers. Other routers within the site have no special requirements beyond those of single-homed site routers.

5.6 Impact on Hosts

There are no requirements of hosts beyond those of single-homed sites hosts.

5.7 Interactions between Hosts and the Routing System

There are no requirements for interaction between routers and hosts beyond those of single-homed site routers and hosts.

6. Limitations of IPv4 Multihoming

6.1 Scalability

Current IPv4 multihoming practices contribute to the significant growth currently observed in the state held in the global inter-provider routing system; this is a concern both because of the hardware requirements it imposes and also because of the impact on the stability of the routing system.

The only method described in this document that doesn't add to the growth of state in the global inter-provider routing system is the RFC2260/NAT based method. The use of this method might explains the relatively higher growth in multihomed sites, compared to the growth of the state in the global inter-provider routing system.

These mechanisms also add to the consumption of public AS number resources, when small site wishing to multihome obtain an AS number specifically for only that purpose. Using a different mechanism would help to conserve the 16-bit AS number space, and avoid the move to 32-bit AS numbers.

This issue is discussed in great detail in [7].

7. Security Considerations

This memo analyzes the IPv4 multihoming practices. This analysis only includes the description of the mechanisms and partially how they affect the availability of the site deploying the IPv4 multihoming mechanism. Other security properties of the IPv4 multihoming mechanisms are not analyzed.

8. IANA Considerations

This document requests no action by IANA.

9. Acknowledgements

Thanks goes to Pekka Savola and Iljitsch van Beijnum for providing feedback and suggestions on the text as well as text.

10 Informative references

[1] Fuller, V., Li, T., Yu, J. and K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993.

[2] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. and E. Lear, "Address Allocation for Private Internets", RFC 1918, February 1996.

[3] Hawkinson, J. and T. Bates, "Guidelines for creation, selection, and registration of an Autonomous System (AS)", RFC 1930, March 1996.

[4] Bates, T. and Y. Rekhter, "Scalable Support for Multi-homed Multi-provider Connectivity", RFC 2260, January 1998.

[5] Black, B., Gill, V. and J. Abley, "Goals for IP Multihoming Architectures", RFC 3582, August 2003.

[6] Lear, E., "Things MULTI6 Developers should think about", Internet-Drafts draft-ietf-multi6-things-to-think-about-00, June 2004.

[7] Huston, G., "Analyzing the Internet's BGP Routing Table", January 2001.

Authors' Addresses

Joe Abley
Internet Systems Consortium, Inc.
950 Charter Street
Redwood City, CA 94063
USA

Phone: +1 650 423 1317
EMail: jabley@isc.org

Benjamin Black
Layer8 Networks

EMail: ben@layer8.net

Vijay Gill
AOL
12100 Sunrise Valley Dr
Reston, VA 20191
US

Phone: +1 410 336 4796
EMail: vgill@vijaygill.com

Kurt Erik Lindqvist
Netnod Internet Exchange
Bellmansgatan 30
Stockholm S-118 47
Sweden

Phone: +46 8 615 85 70
EMail: kurtis@kurtis.pp.se

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.