Network Working Group
Request for Comments: 4116
Category: Informational

J. Abley
ISC
K. Lindqvist
Netnod Internet Exchange
E. Davies
Independent Researcher
B. Black
Layer8 Networks
V. Gill
AOL

July 2005

IPv4 Multihoming Practices and Limitations

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2005).

Abstract

Multihoming is an essential component of service for many Internet sites. This document describes some implementation strategies for multihoming with IPv4 and enumerates features for comparison with other multihoming proposals (particularly those related to IPv6).

Table of Contents

1. Introduction 3
2. Terminology 3
3. IPv4 Multihoming Practices 4
   3.1. Multihoming with BGP 4
        3.1.1. Addressing Considerations 4
        3.1.2. AS Number Considerations 6
   3.2. Multiple Attachments to a Single Transit Provider 6
   3.3. NAT- or RFC2260-based Multihoming 7
4. Features of IPv4 Multihoming 7
   4.1. Redundancy 7
   4.2. Load Sharing 8
   4.3. Performance 8
   4.4. Policy 8
   4.5. Simplicity 9
   4.6. Transport-Layer Survivability 9
   4.7. Impact on DNS 9
   4.8. Packet Filtering 9
   4.9. Scalability 9
   4.10. Impact on Routers 10
   4.11. Impact on Hosts 10
   4.12. Interactions between Hosts and the Routing System 10
   4.13. Operations and Management 10
   4.14. Cooperation between Transit Providers 10
5. Security Considerations 10
6. Acknowledgements 10
7. Informative References 11

1. Introduction

Multihoming is an important component of service for many Internet sites. Current IPv4 multihoming practices have been added on to the Classless Inter Domain Routing (CIDR) architecture [RFC1519], 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 satisfy a number of high-level requirements. It is widely used in the IPv4 Internet. There are some practical limitations, however, including concerns as to how it would scale with future Internet growth. This document aims to document common IPv4 multihoming practices and enumerate their features for comparison with other multihoming approaches.

There are a number of different ways to route and manage traffic in and out of a multihomed site: the majority rely on the routing policy capabilities of the inter-domain routing protocol, the Border Gateway Protocol, version 4 (BGP) [RFC1771]. This document also discusses a multi-homing strategy which does not rely on the capabilities of BGP.

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 [RFC1918].

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 and its own direct customer networks. 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 has more than one point of layer-3 interconnection to a single transit provider.

Provider-Independent (PI) addresses are globally-unique addresses which are not assigned by a transit provider, but are provided by some other organisation, usually a Regional Internet Registry (RIR).

Provider-Aggregatable (PA) addresses are globally-unique addresses assigned by a transit provider to a customer. The addresses are considered "aggregatable" because the set of routes corresponding to the PA addresses are usually covered by an aggregate route set corresponding to the address space operated by the transit provider, from which the assignment was made.

Note that the words "assign" and "allocate" have specific meanings in Regional Internet Registry (RIR) address management policies, but are used more loosely in this document.

3. IPv4 Multihoming Practices

3.1. Multihoming with BGP

The general approach for multihoming with BGP is to announce a set of routes to two or more transit providers. This provides the rest of the Internet with multiple paths back to the multihomed sites, and each transit provider provides an additional possible path for the site's outbound traffic.

3.1.1. Addressing Considerations

3.1.1.1. PI Addresses

The site uses PI addresses, and a set of routes covering those PI addresses is announced or propagated by two or more transit providers.

Using PI addresses has long been the preferred approach for IPv4 multihoming. 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 obtain a /24 PI assignment. Since then, RIR address management policies have become less liberal in this respect. Not all RIRs support the assignment of address blocks to small, multihomed end-users, and those that do support it require justification for blocks as large as a /24, which cannot be met by small sites. As a consequence, PI addresses are not available to many sites who wish to multihome.

Each site that uses PI addresses introduces an additional prefix into the global routing system. If this scheme for multihoming became widespread, it would present scaling concerns.

3.1.1.2. PA Addresses

The site uses PA addresses assigned by a single transit provider. The set of routes covering those PA addresses (the "site route set") is announced or propagated by one or more additional transit providers. The transit provider which assigned the PA addresses (the "primary transit provider") originates a set of routes which cover the site route set. The primary transit provider often originates or propagates the site route set as well as the covering aggregates.

The use of PA addresses is applicable to sites whose addressing requirements are not sufficient to meet the requirements for PI assignments by RIRs. However, in the case where the site route set is to be announced or propagated by two or more different transit providers, common operational practice still dictates minimum /24 prefixes, which may be larger than the allocation available to small sites.

There have been well-documented examples of sites filtering long-prefix routes which are covered by a transit-providers aggregate. If this practice were to become very widespread, it might limit the effectiveness of multihoming using PA addresses. However, limited filtering of this kind can be tolerated because the aggregate announcements of the primary transit provider should be sufficient to attract traffic from autonomous systems which do not accept the covered site route set. The more traffic that follows the primary transit provider's aggregate in the absence of the covered, more-specific route, the greater the reliance on that primary transit provider. In some cases, this reliance might result in an effective single point of failure.

Traffic following the primary transit provider's aggregate routes may still be able to reach the multihomed site, even in the case where the connection between the primary transit provider and the site has failed. The site route set will still be propagating through the site's other transit providers. If that route set reaches (and is accepted by) the primary transit provider, connectivity for traffic following the aggregate route will be preserved.

Sites that use PA addresses are usually obliged to renumber if they decide not to retain connectivity to the primary transit provider. While this is a common requirement for all sites using PA addresses (and not just those that are multihomed), it is one that may have more frequent impact on sites whose motivation to multihome is to facilitate changes of ISP. A multihomed site using PA addresses can still add or drop other service providers without having to renumber.

3.1.2. AS Number Considerations

3.1.2.1. Consistent Origin AS

A multihomed site may choose to announce routes to two or more transit providers from a globally-unique Autonomous System (AS) number assigned to the site. This causes the origin of the route to appear consistent when viewed from all parts of the Internet.

3.1.2.2. Inconsistent Origin AS

A multihomed site may choose to use a private-use AS number [RFC1930] to originate routes to transit providers. It is normal practice for private-use AS numbers to be stripped from AS_PATH attributes before they are allowed to propagate from transit providers towards peers. Therefore, routes observed from other parts of the Internet may appear to have inconsistent origins.

When using private-use AS numbers, collisions between the use of individual numbers by different transit providers are possible. These collisions are arguably best avoided by not using private-use AS numbers for applications which involve routing across administrative domain boundaries.

A multihomed site may request that their transit providers each originate the site's routes from the transit providers' ASes. Dynamic routing (for the purposes of withdrawing the site's route in the event that connectivity to the site is lost) is still possible, in this case, using the transit providers' internal routing systems to trigger the externally-visible announcements.

Operational troubleshooting is facilitated by the use of a consistent origin AS. This allows import policies to be based on a route's true origin rather than on intermediate routing details, which may change (e.g., as transit providers are added and dropped by the multihomed site).

3.2. Multiple Attachments to a Single Transit Provider

Multihoming can be achieved through multiple connections to a single transit provider. This imposes no additional load on the global routing table beyond that involved in the site being single-attached. A site that has solved its multihoming needs in this way is commonly referred to as "multi-attached".

It is not a requirement that the multi-attached site exchange routing information with its transit provider using BGP. However, in the event of failure, some mechanism for re-routing inbound and outbound traffic over remaining circuits is required. BGP is often used for this purpose.

Multi-attached sites gain no advantages from using PI addresses or (where BGP is used) globally-unique AS numbers, and have no need to be able to justify address assignments of a particular minimum size. However, multi-attachment does not protect a site from the failure of the single transit provider.

3.3. NAT- or RFC2260-based Multihoming

This method uses PA addresses assigned by each transit provider to which the site is connected. The addresses are either allocated to individual hosts within the network according to [RFC2260], or the site uses Network Address Translation (NAT) to translate the various provider addresses into a single set of private-use addresses [RFC1918] within the site. The site is effectively singlehomed to more than one transit provider. None of the transit providers need to make any accommodations beyond those typically made for a non-multihomed customer.

This approach accommodates a wide range of sites, from residential Internet users to very large enterprises, requires no PI addresses or AS numbers, and imposes no additional load on the Internet's global routing system. However, it does not address several common motivations for multihoming, most notably transport-layer survivability.

4. Features of IPv4 Multihoming

The following sections describe some of the features of the approaches described in Section 3, in the context of the general goals for multihoming architectures presented in [RFC3582]. Detailed descriptions and rationale for these goals can be found in that document.

4.1. Redundancy

All the methods described provide redundancy, which can protect a site from some single-point failures. The degree of protection depends on the choice of transit providers and the methods used to interconnect the site to those transit providers.

4.2. Load Sharing

All of the methods described provide some measure of load-sharing capability. Outbound traffic can be shared across ISPs using appropriate exit selection policies; inbound traffic can be distributed using appropriate export policies designed to influence the exit selection of remote sites sending traffic back towards the multihomed site.

In the case of RFC2260/NAT multihoming, distribution of inbound traffic is controlled by address selection on the host or NAT.

4.3. Performance

BGP-speaking sites can employ import policies that cause exit selection to avoid paths known to be problematic. For inbound traffic, sites can often employ route export policy, which affords different treatment of traffic towards particular address ranges within their network.

It should be noted that this is not a comprehensive capability. In general, there are many traffic engineering goals which can only be loosely approximated using this approach.

In the case of RFC2260/NAT multihoming in the absence of BGP routing information, management of outbound traffic is not possible. The path taken by inbound traffic for a particular session can be controlled by source address selection on the host or NAT.

4.4. Policy

In some circumstances, it is possible to route traffic of a particular type (e.g., protocol) via particular transit providers. This can be done if the devices in the site which source or sink that traffic can be isolated to a set of addresses to which a special export policy can be applied.

An example of this capability is the grouping of budget, best-effort Internet customers into a particular range of addresses that is covered by a route which is announced preferentially over a single, low-quality transit path.

In the case of RFC2260/NAT multihoming, policies such as those described here can be accommodated by appropriate address selection on the host or NAT. More flexible implementations may be possible for sessions originated from the multihomed site by selecting an appropriate source address on a host or NAT, according to criteria such as transport-layer protocols and addresses (ports).

4.5. Simplicity

The current methods used as multihoming solutions are not without their complexities, but have proven to be sufficiently simple to be used. They have the advantage of familiarity due to having been deployed extensively.

4.6. Transport-Layer Survivability

All BGP-based multihoming practices provide some degree of session survivability for transport-layer protocols. However, in cases where path convergence takes a long time following a re-homing event, sessions may time out.

Transport-layer sessions will not, in general, survive over a re-homing event when using RFC2260/NAT multihoming. Transport protocols which support multiple volatile endpoint addresses may be able to provide session stability; however, these transport protocols are not in wide use.

In all the methods described in this document, new transport-layer sessions are able to be created following a re-homing event.

4.7. Impact on DNS

These multihoming strategies impose no new requirements on the DNS.

4.8. Packet Filtering

These multihoming practices do not preclude filtering of packets with inappropriate source or destination addresses at the administrative boundary of the multihomed site.

4.9. Scalability

Current IPv4 multihoming practices are thought to contribute to significant observed growth in the amount of state held in the global inter-provider routing system. This is a concern because of both the hardware requirements it imposes and the impact on the stability of the routing system. This issue is discussed in greater detail in [RFC3221].

Of the methods presented in this document, RFC2260/NAT multihoming and multi-attaching to a single transit provider provide no additional state to be held in the global routing system. All other strategies contribute to routing system state bloat.

Globally-unique AS numbers are a finite resource. Thus, widespread multihoming that uses strategies requiring assignment of AS numbers might lead to increased resource contention.

4.10. Impact on Routers

For some of the multihoming approaches described in this document, the routers at the boundary of the multihomed site are required to participate in BGP sessions with transit provider routers. Other routers within the site generally have no special requirements beyond those in singlehomed sites.

4.11. Impact on Hosts

There are no requirements of hosts beyond those in singlehomed sites.

4.12. Interactions between Hosts and the Routing System

There are no requirements for interaction between routers and hosts beyond those in singlehomed sites.

4.13. Operations and Management

There is extensive operational experience in managing IPv4-multihomed sites.

4.14. Cooperation between Transit Providers

Transit providers who are asked to announce or propagate a PA prefix covered by some other (primary) transit provider usually obtain authorisation first. However, there is no technical requirement or common contractual policy which requires this coordination to take place.

5. Security Considerations

This document discusses current IPv4 multihoming practices, but provides no analysis of the security implications of multihoming.

6. Acknowledgements

Special acknowledgement goes to John Loughney for proof-reading and corrections. Thanks also goes to Pekka Savola and Iljitsch van Beijnum for providing feedback and contributing text.

This work was supported by the US National Science Foundation (research grant SCI-0427144) and DNS-OARC.

7. Informative References

[RFC1519] Fuller, V., Li, T., Yu, J., and K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993.

[RFC1771] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC1930] Hawkinson, J. and T. Bates, "Guidelines for creation, selection, and registration of an Autonomous System (AS)", BCP 6, RFC 1930, March 1996.

[RFC2260] Bates, T. and Y. Rekhter, "Scalable Support for Multi-homed Multi-provider Connectivity", RFC 2260, January 1998.

[RFC3221] Huston, G., "Commentary on Inter-Domain Routing in the Internet", RFC 3221, December 2001.

[RFC3582] Abley, J., Black, B., and V. Gill, "Goals for IPv6 Site-Multihoming Architectures", RFC 3582, August 2003.

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

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

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

Elwyn B. Davies
Independent Researcher
Soham, Cambridgeshire CB7 5AW
UK

Phone: +44 7889 488 335
EMail: elwynd@dial.pipex.com

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

Full Copyright Statement

Copyright (C) The Internet Society (2005).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

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