INTERNET-DRAFT

Danny McPherson
Arbor Networks
Dave Oran
Cisco Systems

Expires: July 2009
January 6, 2009
Intended Status: Informational

Architectural Considerations of IP Anycast
<draft-mcpherson-anycast-arch-implications-00.txt>

Status of this Memo

Copyright Notice

Abstract

This memo discusses architectural implications of IP anycast.

Table of Contents

1. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119].

2. Overview

IP anycast is used for at least one critical Internet service, that of the Domain Name System [RFC 1035] root servers. As of early 2008, at least 10 of the 13 root name servers were using IP anycast [RSSAC 29]. Use of IP anycast is growing for other applications as well. It has been deployed for over a decade for DNS resolution services and is currently used by several DNS Top Level Domain (TLD) operators. IP anycast is also used for other services in operational environments, to include Network Time Protocol (NTP) [RFC 1305].

Anycast addresses are syntactically indistinguishable from unicast addresses. Allocation of anycast addresses typically follow a model similar to that of unicast allocation policies. Anycast addressing is inherent to that of unicast in multiple locations, and leverages unicast destination routing to deliver a packet to either zero or one interface among the interfaces asserting the address. The expectation of delivery is to the "closest" instance as determined by unicast routing topology metric(s). There is also an expectation of load- balancing that exists among equal cost routes.

Unlike IP unicast, it is not considered an error to assert the same anycast address on multiple interfaces within the same or multiple systems.

Some consider anycast a "deceptively simple idea". That is, many pitfalls and subtleties exist with application and transport, as well as for routing configuration and operation, when IP anycast is employed. In this document, we aim to capture many of the architectural implications of IP anycast.

3. Anycast History

As of this writing, the term "anycast" appears in 126 RFCs, and ~360 Internet-Drafts (since 2006).

The first formal specification of anycast was provided in "Host Anycasting Service" [RFC 1546]. The authors of this document did a good job of capturing most of the issues that exist with IP anycast today.

One of the first documented uses of anycast was in 1994 for a "Video Registry" experiment [IMR 9401]. In the experiment, a UDP query was transmitted to an anycast address to locate a server, and TCP was used by the client to query the server, and then multicast was used to distribute the server database. There is also discussion that ISPs began using anycast for DNS resolution services around the same time, although no public references to support this are available.

The IAB clarified in [RFC 2101] that IPv4 anycast addresses were pure "locators", and could never serve as an "identifier" (of a host, interface, or anything else).

In 1998 the IAB conducted a routing workshop [RFC 2902]. Interestingly, of the conclusions and ouput actions items from the report, an Anycast section is contained in S 2.10.3. Specifically called out in the conclusions section is the need to describe the advantages and disadvantages of anycast, and the belief that local-scoped well-known anycast addresses will be useful to some applications. In the subsequent section, an action item was outlined that suggested a BOF should be held to plan work on progress, and if a working group forms, a paper on the advantages and the disadvantages of anycast should be included as part of the charter.

4. Use of Anycast in RFCs

SNTPv4 [RFC 2030] defined how to use anycast for server discovery. This was extended in [RFC 4330] as an NTP-specific "manycast" service, in which anycast was used for the discovery part.

IPv6 defined some reserved subnet anycast addresses [RFC 2526] and assigned one to "Mobile IPv6 Home-Agents" [RFC 3775].

The original IPv6 transition mechanism [RFC 2893] made use of IPv4 anycast addresses as tunnel endpoints for 6-over-4 tunnels, but this was removed in the revision [RFC 4213]. Huitema's Relay Router scheme [RFC 3068] for 6to4 translation also used anycast in a similar fashion.

DNS use of anycast was first specified in [RFC 3258]. Is is noteable that it used the term "shared unicast address" rather than "anycast

address" for the service.

Anycast was used for routing to rendezvous points (RPs) for MSDP and PIM as documented in [RFC 4610].

[RFC 4786] deals with how the routing system interacts with anycast services, and the operation of anycast services.

[RFC 4892] cites the use of anycast with DNS as a motivation to identify individual nameserver instances' [RFC 5001] defines the NSID option for doing so.

"Reflections on Internet Transparency" [RFC 4924] briefly mentions how violating transparency can also damage global services that use anycast.

5. Anycast in IPv6

The original IPv6 addressing architecture [RFC 1884], carried forward in [RFC 2373] and [RFC 3513], severely restricted the use of anycast addresses. In particular, they provided that anycast addresses MUST NOT be used as a source address, and MUST NOT be assigned to an IPv6 host (i.e., only routers). these restrictions were finally lifted in 2006, with the publication of [RFC 4291].

In fact, the recent "IPv6 Transition/Co-existence Security Considerations" [RFC 4942] overview now recommends:

> "To avoid exposing knowledge about the internal structure of the network, it is recommended that anycast servers now take advantage of the ability to return responses with the anycast address as the source address if possible."

6. DNS Anycast

"Distributed Authoritative Name Servers via Shared Unicast Addresses" [RFC 3258] described how to reach authoritative name servers using anycast. It made some interesting points:

- asserted (as an advantage) that no routing changes were needed

- recommended stopping DNS processes, rather than withdrawing

routes, to deal with fail-over.

- argued that failure modes involving state were not serious, because:

  - the vast majority of DNS queries are UDP
  - large routing metric disparity among authoritative server instances would localize queries to a single instance for most clients
  - when the resolver tries TCP and it breaks, the resolver will move to a different server instance (where presumably it doesn't break)

7. BCP 126 Revisited

BCP 126 [RFC 4786] was a product of the IETF's GROW working group. The primary design constraint considered was that routing "be stable" for significantly longer than a "transaction time", where "transaction time" is loosely defined as "a single interaction between a single client and a single server". It takes no position on what applications are suitable candidates for anycast usage.

Furthermore, it views anycast service disruptions as an operational problem, "Operators should be aware that, especially for long running flows, there are potential failure modes using anycast that are more complex than a simple 'destination unreachable' failure using unicast."

The document primary deals with global Internet-wide services provided by anycast. Where internal topology issues are discussed they're mostly regarding routing implications, not application design implications. BCP 126 also views networks employing per-packet load balancing on equal cost paths as "pathological".

8. Layering and Resiliency

Preserving the integrity of a modular layered design for IP protocols on the Internet is critical to its continued success and flexibility. One such consideration is that of whether an application should have to adapt to changes in the routing system.

Higher layer protocols should make minimal assumptions about lower layer protocols. E.g., applications should make minimal assumptions about routing stability, just as they should make minimal assumptions about congestion and packet loss. When designing applications, it would perhaps be safe to assume that the routing system may deliver each packet to a different service instance, in any pattern, with termporal re-ordering being a not-so-rare phenomenon.

Stateful transport protocols (TCP, DCCP, SCTP), without modification, do not understand the properties of anycast and hence will fail probabilistically, but possibly catastrophically, in the presence of "normal" routing dynamics.

9. Anycast Addresses as Destinations

Anycast addresses are "safe" to use as a destination addresses for an application if:

- A request message or "one shot" message is self-contained in a single transport packet

- A stateless transport (e.g., UDP) is used for the above

- Replies are always sent to a unicast address; these can be multi-packet since the unicast destination is "stable"

  * Note: this constrains the use of anycast as source addresses as reply messages to that address may reach a device that was not the source that initially triggered it.

- The server side of the application keeps no hard state across requests

- Retries are idempotent; in addition to not assuming server state, they do not encode any assumptions about loss of requests versus loss of replies.

10. Anycast Addresses as Sources

Anycast addresses are "safe" to use as source addresses for an application if:

- o No reflexive (response) message is generated by the receiver with the anycast source used as a destination

  * unless the application has some private state synchronization that allows for the reflexive message arriving at a different instance

- o The source anycast address is a bona fide interface address if reverse path forwarding (RPF) checking is on, or a service address explicitly provisioned to bypass RPF

11. Regarding Widespread Anycast Use

Widespread use of anycast for global Internet-wide services or inter-domain services has some scaling challenges. Similar in ways to multicast, each service generates at least one unique route in the global BGP routing system. As a result, additional anycast instances result in additional paths for a given prefix, which scales super-linearly as a function of denseness of inter-domain interconnection within the routing system (i.e., more paths result in more resources, more network interconnections result in more paths)..

12. Service Discovery

Applications able to tolerate an extra round trip time (RTT) to learn a unicast destination address for multi-packet exchanges can safely use anycast destination addresses for service instance discovery.

- o "Instance discovery" message sent to anycast destination address

- o Reply sent from unicast source address of the interface that received the discovery message

- o Subsequent exchanges use the unicast address

13. Middleboxes and Anycast

Middleboxes (e.g., NATs, firewalls) may cause problems when used in

conjunction with anycast. In particular, a switch from anycast to unicast requires may require a new binding, and this may not exist in the middlebox.

14. Transport Implications

UDP is the "lingua franca" for anycast today. Stateful transports could be enhanced to be more anycast friendly. It seems as though this was anticipated in Host Anycasting Services [RFC 1546], specifically:

"The solution to this problem is to only permit anycast addresses as the remote address of a TCP SYN segment (without the ACK bit set). A TCP can then initiate a connection to an anycast address. When the SYN-ACK is sent back by the host that received the anycast segment, the initiating TCP should replace the anycast address of its peer, with the address of the host returning the SYN-ACK. (The initiating TCP can recognize the connection for which the SYN-ACK is destined by treating the anycast address as a wildcard address, which matches any incoming SYN-ACK segment with the correct destination port and address and source port, provided the SYN-ACK's full address, including source address, does not match another connection and the sequence numbers in the SYN-ACK are correct.) This approach ensures that a TCP, after receiving the SYN-ACK is always communicating with only one host."

Multi-address transports (e.g., SCTP) might be more amenable to such extensions than TCP.

Some similarities exist between what is needed for anycast and what is needed for address discovery when doing multi-homing in the transport layer. **NEED TO EXPAND ON THIS***

15. Security Considerations

Anycast is often employed to mitigate or at least localize the effects of distributed denial of service (DDOS) attacks. For example, with the Netgear NTP fiasco [RFC 4085] anycast was used in a distributed sinkhole model to mitigate the effects of embedded globally-routed Internet addresses in network elements.

"Internet Denial-of-Service Considerations" [RFC 4732] notes that "A number of the root nameservers have since been replicated using anycast to further improve their resistance to DoS".

[RFC 4786] cites DoS mitigation, constraining DoS to localized regions, and identifying attack sources using spoofed addresses as some motivations to deploy services using anycast. Multiple anycast service instances such as those used by the root name servers also add resiliency when network partitioning occurs (e.g., as the result of transoceanic fiber cuts or natural disasters).

16. Deployment Considerations

This document covers issues associated with the architectural implications of anycast. Operators should heed these considerations when evaluating the use of anycast in their specific environments.

17. IANA Considerations

No IANA actions are required.

18. Acknowledgments

Many thanks for Dave Thaler and Kurtis Lindqvist for their early review and feedback on this document.

Your name could be here....

19. References

19.1. Normative References

19.2. Informative References

[IMR 9401] "INTERNET MONTHLY REPORT", January 1994, http://mirror.facebook.com/rfc/museum/imr/imr9401.txt

[RSSAC 29] "RSSAC 29 Meeting Minutes", December 2, 2007, http://www.rssac.org/meetings/04-08/rssac29.pdf

[RFC 1035] Mockapetris, P., "DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION", RFC 1035, November 1987.

[RFC 1305] Mills, D., "Network Time Protocol (Version 3) Specification, Implementation and Analysis", RFC 1305, March 1992.

[RFC 1546] Partridge, C., Mendez, T., Milliken, W., "Host Anycasting Service", RFC 1546, November 1993.

[RFC 1884] Hinden, R., Deering, S., "IP Version 6 Addressing Architecture", RFC 1884, December 1995.

[RFC 2030] Mills, D., "Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI", RFC 2030, October 1996.

[RFC 2101] Carpenter, B., Crowcroft, J., Rekhter, Y., "IPv4 Address Behaviour Today", RFC 2101, February 1997.

[RFC 2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.

[RFC 2373] Hinden, R., Deering, S., "IP Version 6 Addressing Architecture", RFC 2373, July 1998.

[RFC 2526] Johnson, D., Deering, S., "Reserved IPv6 Subnet Anycast Addresses", RFC 2526, March 1999.

[RFC 2893] Gilligan, R., Nordmark, E., "Transition Mechanisms for IPv6 Hosts and Routers", RFC 2893, August 2000.

[RFC 2902] Deering, S., Hares, S., Perkins, C., Perlman, R., "Overview of the 1998 IAB Routing Workshop", RFC 2902, August 2000.

[RFC 3068] Huitema, C., "An Anycast Prefix for 6to4 Relay Routers", RFC 3068, June 2001.

[RFC 3258] Hardie, R., "Distributing Authoritative Name Servers via Shared Unicast Addresses", RFC 3258, April 2002.

[RFC 3513] Hinden, R., Deering, S., "Internet Protocol Version 6 (IPv6) Addressing Architecture", RFC 3513, April 2003.

[RFC 3775] Johnson, D., Perkins, C., Arkko, J., "Mobility Support in IPv6", RFC 3775, June 2004.

[RFC 4085] Plonka, D., "Embedding Globally-Routable Internet Addresses Considered Harmful", RFC 4085, June 2005.

[RFC 4213] Normark, E., Gilligan, R., "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.

[RFC 4291] Hinden, R., Deering, S., "IP Version 6 Addressing Architecture", RFC 4291, February 2006.

[RFC 4330] Mills, D., "Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI", RFC 4330, January 2006.

[RFC 4610] Farinacci, D., Cai, Y., "Anycast-RP Using Protocol Independent Multicast (PIM)", RFC 4610, August 2006.

[RFC 4732] Handley, M., Rescorla, E., IAB, "Internet Denial-of-Service Considerations", RFC 4732, November 2006.

[RFC 4786] Abley, J., Lindqvist, K., "Operation of Anycast Services", RFC 4786, December 2006.

[RFC 4892] Woolf, S., Conrad, D., "Requirements for a Mechanism Identifying a Name Server Instance", RFC 4892, June 2007.

[RFC 4924] Aboba, B., Davies, E., " Reflections on Internet

Transparency", RFC 4924, July 2007.

[RFC 4942] Davies, E., Krishnan, S., Savola, P., "IPv6 Transition/Coexistence Security Considerations", RFC 4942, September 2007.

[RFC 5001] Austein, R., "DNS Name Server Identifier (NSID) Option", RFC 5001, August 2007.

20. Authors' Addresses

Danny McPherson
Arbor Networks, Inc.
Email: danny@arbor.net

Dave Oran
Cisco Systems
Email: oran@cisco.com

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).