IP Anycast: Point to (Any) Point Communications

Draft 0.3

Chris Metz, chmetz@cisco.com

Introduction

The Internet supports several different communication paradigms. Unicast is defined as a point-to-point flow of packets between a source (client) and destination (server) host. The server is identified by a unique IP unicast address (e.g. 155.12.1.1) and this address is contained in the header of each packet sent from the client. The network of routers will then make a best-effort attempt to deliver the packet to the destination host identified by this unicast address. Web-browsing (employing HTTP and TCP) and file transfer (FTP) are two examples of unicast applications.

Multicast is defined as the point-to-multipoint flow of packets between a single source host and one or more destination hosts. Rather than send a copy of the same packet to the unicast address of each destination host, the source host sends a single copy of the packet to a group address (e.g. 227.12.33.2). Any number of unique destination hosts that wish to receive the multicast packet will be configured with a multicast group address. The network of routers will then make a best-effort attempt to deliver the multicast packets to all destination hosts identified by the multicast group address. Broadcast-style video-conferencing is an example of an application that employs IP multicast.

Anycast is defined as the point-to-point flow of packets between a single client and the "nearest" destination server identified by an anycast address. The idea behind anycast is that a client would like to send packets to any one of several possible servers offering a particular service or application but does not really care which one. To accomplish this, a single anycast address is assigned to one or more servers contained within an anycast group. A client sends packets to an anycast server by placing the anycast address in the packet header. The network of routers will then attempt to deliver the packet to a server with the matching anycast address.

The notion of anycast is illustrated in figure 1. Two servers are configured with an anycast address of "A" and located in different areas of the network. Packets originating from clients and addressed to "A" will automatically be delivered by the routing system to the closest destination server with the anycast address of "A".

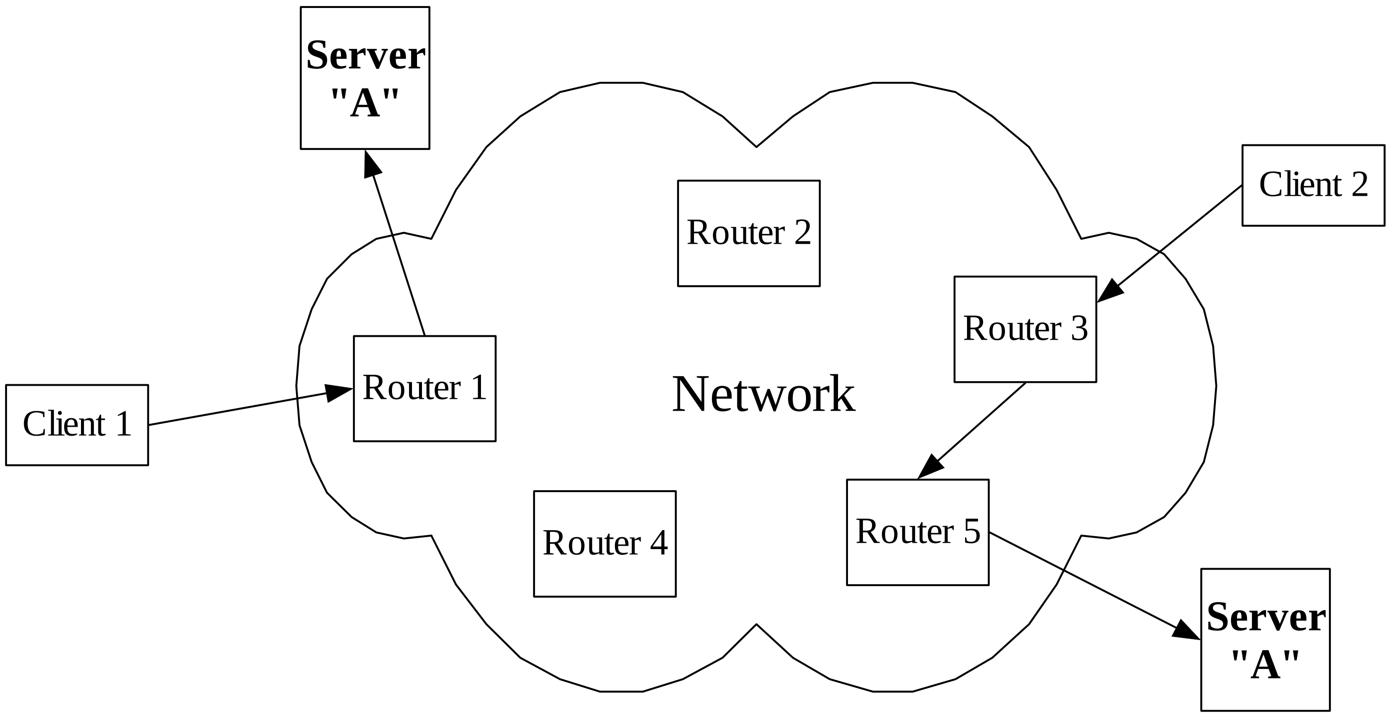

Figure 1 IP Anycast

The criteria for selecting the destination server that will receive anycast-addressed packets is based on different metrics and ultimately determines the type of anycasting scheme that is deployed in the network. If the choice is based on network topology such as the fewest number of router hops or lowest cost (different router links may have different costs) as computed by the routing algorithms inside the routers, then this is called network-layer (or IP) anycast. If the choice is based on server or application metrics such as available capacity, measured response times, number of active connections and so on, then this is called application-layer anycasting[1]. The key difference between the two is that IP anycast relies solely on the network to make the destination anycast server selection. Application-layer anycasting depends on an external entity that monitors the location and status of the many destination servers. The client must then query this external entity to determine the location of best destination server to contact.

The general notion of anycasting is realized today in the Internet as a means of balancing the connection load across multiple web servers sharing the same content. One technique is to configure a cluster of mirrored web servers (called a server farm) on a local area network (LAN) with a single "virtual" IP address. A special device called a server load balancer front-ends the server farm and intercepts and directs HTTP connection requests towards an available server. Another technique is to configure a Domain Name Server (DNS) with a single server name that maps to multiple unicast IP addresses that each represents a unique destination server containing the content. A client will query the DNS to resolve a server name with an IP address and the DNS will return one of the IP addresses in a round-robin fashion. Both techniques allow many connection requests to be distributed among many different server hosts thus contributing to better performance and improved scalability.

With the possible exception of the web server farm, the use of IP anycast is limited in its deployment and still relatively new. But momentum for IP anycast is beginning to pick up given its inherent benefits:

- Reduction in router and link resources. Standard IP routing will deliver packets over the shortest path to closest available host.

- Simplified configuration. A client only needs to be configured with a single anycast address that identifies one of a group of possible servers that offers a particular service or application. It then becomes a simple task for the client to locate a server.

- Network resiliency. If a server in the anycast group goes away, the network will deliver packets to the next closest anycast server.

- Load balancing. Anycast servers distributed over the network topology will have the effect of balancing the traffic load from many clients.

The IETF has adopted anycast as prominent feature of the IPv6 protocol specification[2]. An IPv6 anycast address format has been defined and several solutions for router and service discovery have been proposed. Other technologies that may employ elements of IP anycast include mobile ad-hoc networking for routing wireless messages to the nearest server or forwarding element and content routing for directing content request packets to the nearest content server. It also remains a topic of active research.

IP Anycast Basics

The basic architecture of IP anycast was first described in document issued by the Internet Research Task Force (IRTF) in 1993[3]. The primary motivation behind the development of an anycast service was a desire to simplify the host configuration and routing processes required to contact a particular server.

The original definition of anycast as spelled out in RFC1546 is: " .. host transmits a datagram to an anycast address and the internetwork is responsible for providing best effort delivery of the datagram to at least one, and preferably only one, of the servers that accept datagrams for the anycast address ..".

Another definition appeared several years later when the anycast address portion of the IPv6 addressing architecture was defined: "..An identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the "nearest" one, according to the routing protocols' measure of distance) .."[4]

The basic components of the IP anycast architecture are the anycast address space, anycast routing, anycast LAN communications and anycast group membership announcement.

Anycast Addresses

All servers contained inside an anycast group are given the same address. But unlike multicast group addresses, an anycast address identifies one instance of a server – the closest server as determined by the routing system. Therefore an anycast address is treated by the network as a specific host address.

A pool of anycast addresses can be carved out of the existing IP unicast address space. That will make the job of routing packets much easier because the routers will treat anycast-addressed packets just like any other IP packet. However that there may be a drawback to that approach because it might be necessary for the client or server to understand that it is dealing with anycast packet as might be the case with TCP or an IPv6 host. Another drawback is the possibility that the same anycast address could be assigned to different servers in different groups. Thus the client's packets could be unwittingly delivered to a server that cannot offer the service or application that was expected by the client.

It is also possible to define a separate address class for anycast as was suggested in RFC1546. Given the current constraints on the available IPv4 address space, that is not likely to happen. Therefore in the IPv4 environment, anycast addresses will be allocated from the available unicast address pool.

The developers of IPv6 decided that anycast and unicast addresses should be indistinguishable. Experimentation with IPv6 host-based anycasting was and is still underway so the developers imposed several restrictions that govern the use of IPv6 anycast addresses:

- Can only be configured on a router interface, not a host.
- Cannot be a source address in an IPv6 packet.

An IPv6 subnet-router anycast address has been defined that enables hosts or routers to send a packet to any router on the subnet specified in the anycast address. An IPv6 router is required to support this address and in addition a pool of per-subnet anycast addresses have been reserved for future use[5].

Anycast Routing

The forwarding of anycast packets remains unchanged from standard unicast forwarding so packets are forwarded on a hop-by-hop basis to the closest server based on the shortest path. However IP routing is connection-less which means that a change in the topology (a link or router goes down) may result in a new shortest path leading to a different server in the anycast group. The consequences are that successive anycast-addressed packets originating from a client may arrive at different servers that share the same anycast address. Needless to say this may be confusing to the different servers that receive them.

Anycast routes may be injected into the routing system as host routes. This means they will be appear in the routing table in routers as an individual host entry rather than as a single network entry representing many hosts. For example if the routing system is required to support 1000 distinct anycast groups then 1000 additional entries will appear in the routing table. This places a burden on the routers that require more memory and CPU resources to store and maintain the routing table.

The fact remains that IPv4 and IPv6 anycast addresses cannot be easily aggregated. This is because as part of the unicast address space, a single anycast address can only be contained inside a single larger network prefix (address block). If the anycast servers and their accompanying addresses all reside in a network covered by that prefix then it is not necessary to advertise the anycast address with a host route. If the anycast address resides in networks where the anycast address is not part of the allocated block of network prefixes, then it becomes necessary to advertise the anycast address as a host route.

Another possible application of anycast routing relates to IPv6 policy based routing. A routing header appended to IPv6 packets could contain the anycast address of a particular service provider. This is a simple technique for ensuring that packets flowing from a source to a destination will transit at least one of the ISP's routers with the anycast address. It is questionable if this can ever be deployed in the IPv6 Internet since it would require the first the ISP router anycast addresses are carried in the routing table and second, an external source would need to discover the anycast address of the ISP routers to place inside the routing header.

Anycast LAN Communications

A client and one or more anycast servers may be present on the same LAN (e.g. ethernet). To communicate, the client must resolve the destination IP anycast address with a link-layer (called a media access control or MAC) address. One technique to accomplish this is to use the Address Resolultion Protocol (ARP). The clients broadcasts an ARP request to all hosts on the LAN and caches the MAC address contained in the first ARP reply from a server. If the current anycast server goes down then the client must wait for the cache to timeout before it can broadcast another ARP request. To ensure resilient operations in this environment, small ARP cache timeout values should be used which comes at the expense of generating additional broadcast traffic on the LAN.

To reduce the amount of ARP broadcast traffic on the LAN, the client could use a link-layer multicast scheme to reach all anycast servers on the LAN. Thus only the anycast servers will receive the packets but then there is no formal protocol mechanism that the client can use to establish communications with just a single anycast server.

IPv6 overcomes the limitations of the ARP broadcast scheme through the use of a multicast-based address resolution mechanism and a periodic solicitation to determine host reachability on the LAN[6].

Anycast Group Membership Announcement

Short of manual configuration there needs to be a way that a server can announce its membership in an anycast group. More specifically the host needs to tell the nearest router that it is a destination server for one or more specific anycast addresses. The router in turn will then install the anycast address as a host route in its routing table and use standard unicast routing protocols (e.g. OSPF) to advertise these anycast routes to other routers in the network.

One option is to extend the Internet Group Membership Protocol (IGMP) to enable servers to announce anycast group membership as well as multicast group membership. Another option is to run a routing protocol such as RIPv2 or OSPF on the server and have it advertise anycast host routes to the nearby router. Most likely IPv6 will address this problem by extending the Multicast Listener Discovery (MLD) protocol to carry anycast addresses[7]. The IETF is looking into a solution to this problem and recently formed the Multicast and Anycast Group Membership Announcement (MAGMA) Working Group to develop a solution.[8]

TCP and Anycast

The Transmission Control Protocol (TCP) is a connection-oriented transport layer protocol used by many applications. It provides a reliable transport and ordered delivery of packets between a client and server. Of course it is possible that IP anycast routing could deliver TCP packets from a particular client-server connection to different servers. This can result in degraded TCP throughput or worse, lost TCP sessions. Therefore there needs to be a way to ensure that TCP anycast packets from the client always arrive at the same server in the anycast group.

Several techniques for solving this problem have been suggested through the years. RFC1546 proposed that when a server receives a TCP connection request message from a client (called a TCP SYN message) with an anycast address, it should respond with its unique unicast address in the response (called the TCP SYN-ACK). The client could then complete the three-way TCP connection setup by directing the TCP-ACK message to the unicast address of the server. All subsequent TCP packets would then use the client and server's unicast addresses.

This technique is illustrated in figure 2. The TCP client addresses the TCP SYN message to the server using anycast address "A". The server responses by placing its unicast address in the source address field of the TCP SYN-ACK message. The client then addresses the TCP ACK message to unicast address of the server. The downside of this approach is that it requires a change to the TCP end-points and they must recognize anycast addresses.

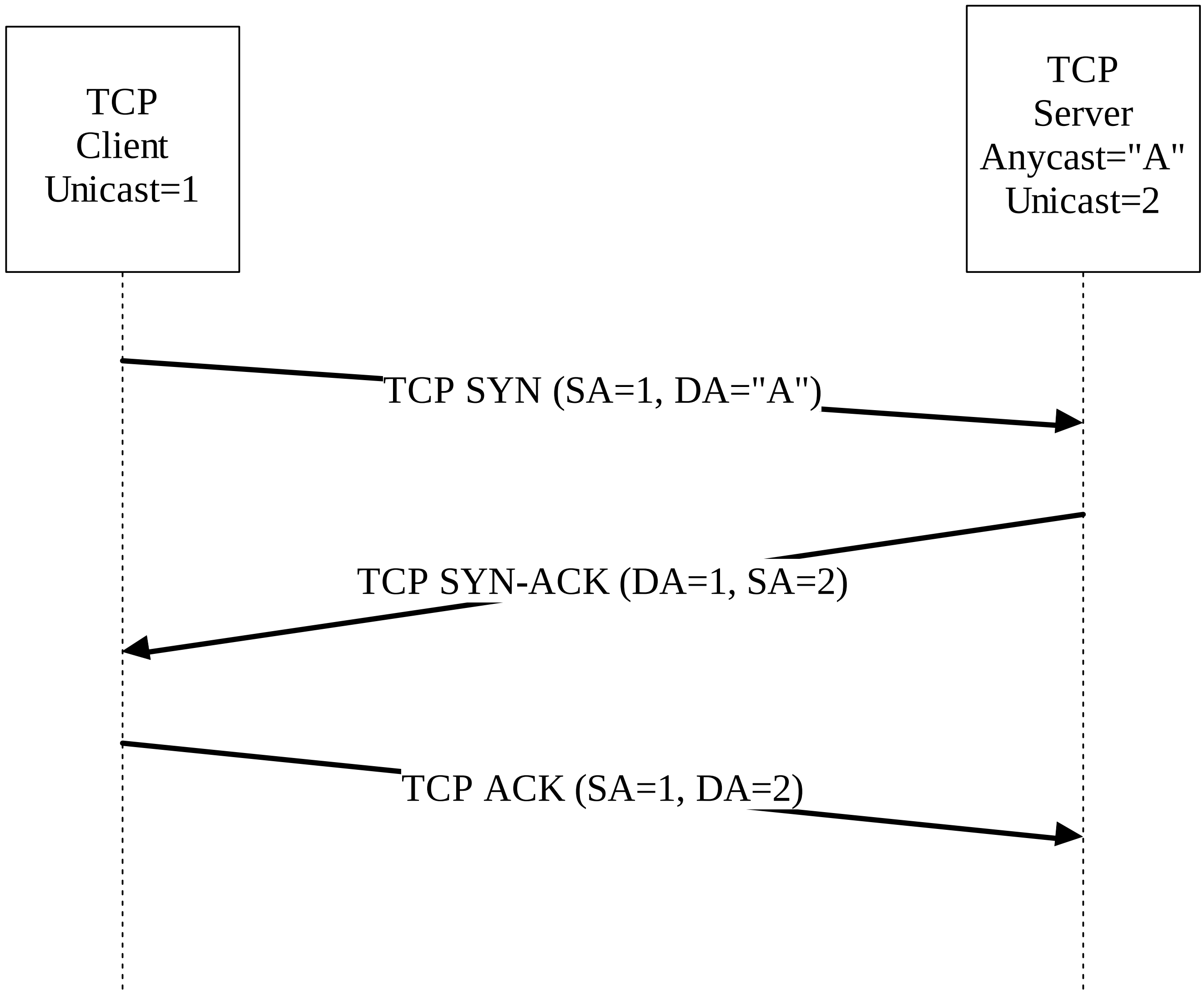

Figure 2 One technique for TCP Anycast Connection Setup

Another option that was proposed was to include an additional IP options field in all packets that would effectively bind the server's anycast address to a unique unicast address[9]. The main drawback of this approach is the possibility of degraded performance because the intermediate routers must examine and process the additional IP options field in each packet even if it has nothing to do with the router itself.

The reasonable solution to this problem is to develop an "anycast-to-unicast" address resolution or discovery mechanism that is performed separate from the TCP protocol. The idea is that the client will send packets addressed to the anycast server and in return will receive information about its unicast address. The HTTP protocol has a technique for redirecting client connections to a different host. It could work by having the anycast server complete a TCP connection setup with a client (yes the anycast address would be used), recognize it was dealing with an anycast address (this must be configured on the hosts somehow), and then issue an HTTP redirect that instructs the client to establish a TCP connection with the unicast address of the server. Another technique under investigation by the IPv6 working group is for the client to use extensions to the ICMPv6 protocol to query the anycast server for its unicast addresses. The client could then use one of these unicast sessions to establish a TCP connection with the server[10].

Practical Problems Solved by Anycast

While it is true that IP anycast is not widely deployed on the Internet today, there are instances where it can and will be used in situations to enhance network performance, efficiency and reliability. One such implementation that has been around for a couple of years is called Anycast RP[11]. It makes use of IP anycast to provide enhanced performance and recovery in intra-domain multicast networks running Protocol Independent Multicast (PIM). It works by having more than one PIM Rendezvous Points (RP) router configured with the same unicast IP address – an anycast address. All multicast senders and receivers are configured with this RP address. The senders and receivers also use normal IP shortest path routing to send multicast control and data packets to the RP address of the RP router.

As illustrated in figure 3, if the RP router1 on the left goes away, all sources and receivers will automatically forward their multicast control and data packets to the next closest RP router (RP Router2) with the same RP address, "A". In steady state, the Anycast RP scheme provides a way of balancing control and data traffic across multiple RP routers and in a failure situation, it provides an automatic backup to the next closest RP router.

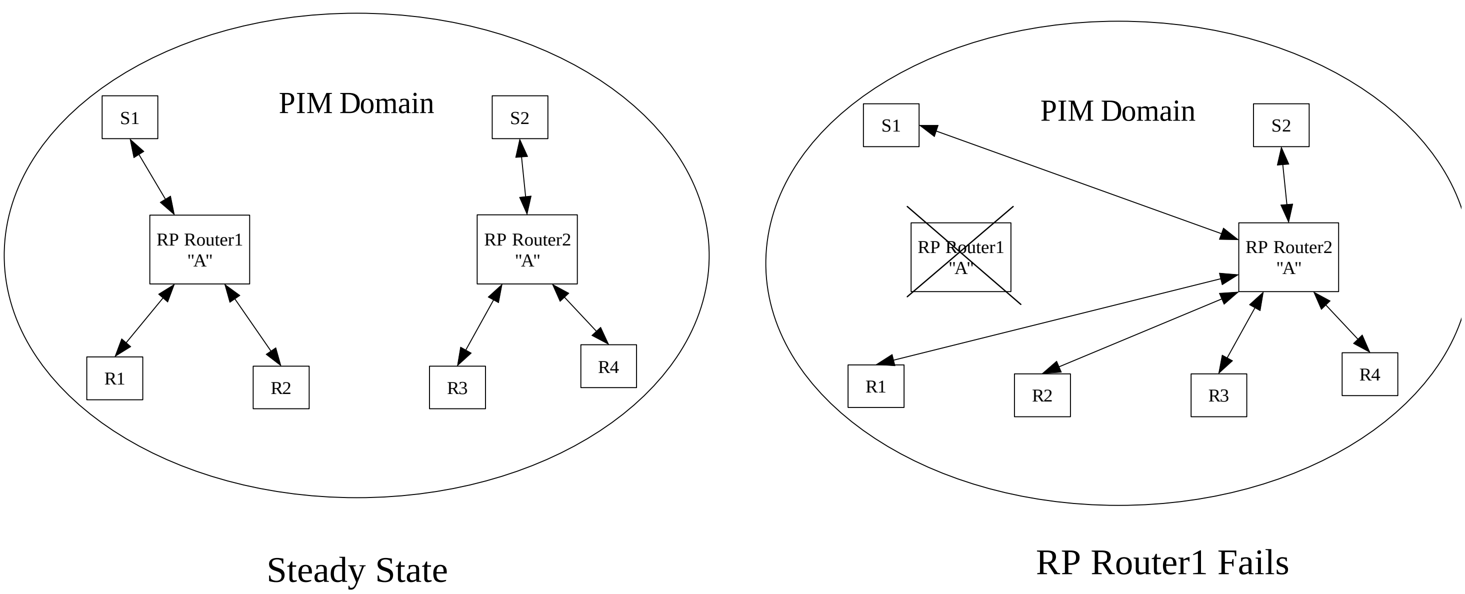

Figure 3 Anycast RP

Another proposed use of IP anycast is to simplify the transition from IPv4 to IPv6. Initially IPv6 networks will be configured as islands that are connected to each other through a network of IPv4 tunnels – that is IPv6 packets will be carried inside IPv4 packets. 6to4 is a technique that uses a special IPv6 address to automatically build IPv4 tunnels between IPv6 islands. Over time these IPv6 islands called 6to4 sites may wish to communicate with the burgeoning native IPv6 network. This requires that packets sent from a 6to4 site border router must pass through a 6to4 relay router that is connected to the native IPv6 network.

Currently the administrators of these 6to4 borders routers must manually configure a default route across the IPv4 network to a 6to4 relay router or configure complex IPv6 interdomain routing protocols. A simple solution to this configuration headache is for the 6to4 relay routers advertise their presence in the IPv4 network (and to other 6to4 border routers) using an anycast network address[12]. This enhances performance (shortest path to any 6to4 relay router will be used), simplifies configuration and enhances scalability (more 6to4 relay routers can be added). An abstract illustration of this mechanism is shown in figure 4.

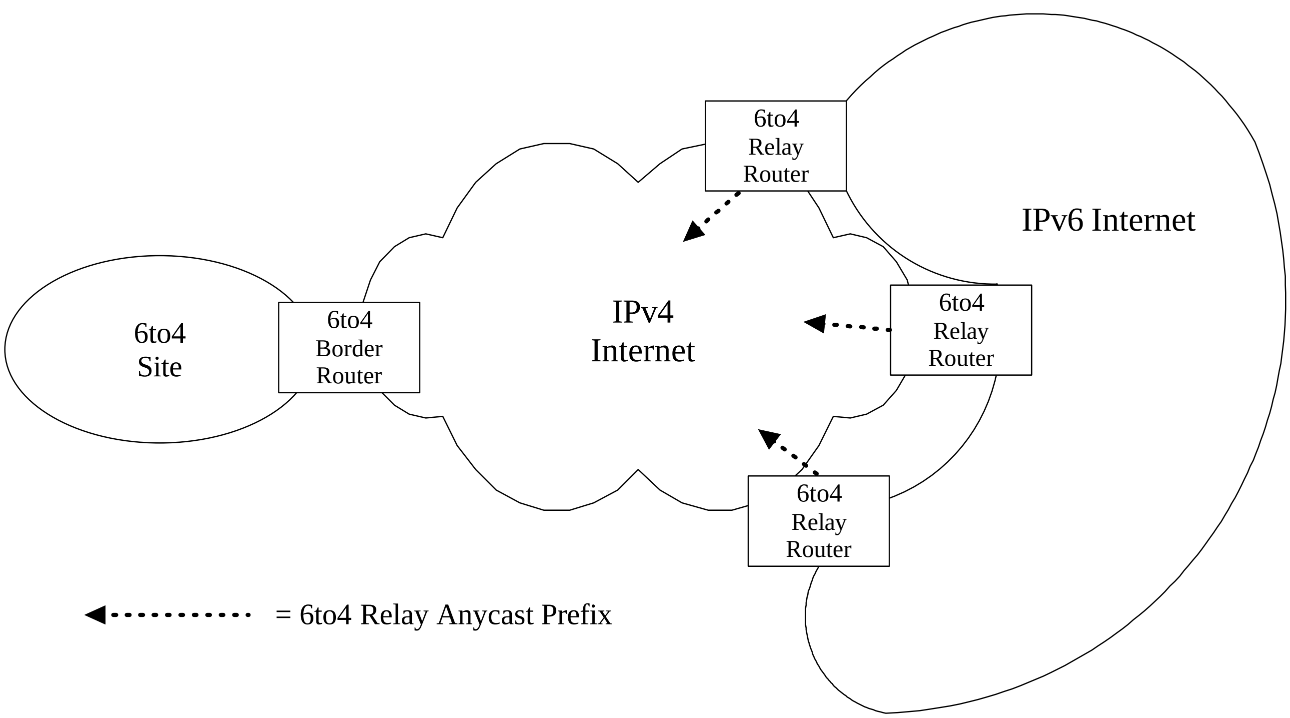

Figure 4 6to4 Relay Router Anycast

Most of the IP anycast development is happening in the IPv6 environment. A design team from the IPv6 working group recently considered a number of ways that an IPv6 client could dynamically discover and contact the DNS for the purpose of performing name-to-address resolution. It concluded that anycast could be used to either discover the unicast address(es) of the DNS or DNS queries could be addressed to the DNS anycast address. If the latter is used then the DNS reply will contain a unicast address in the source field (remember IPv6 does not permit source anycast addresses) and the DNS client should not discard the packet (as is usually the case) because the source address of the reply is different the destination address of the query. In addition a set of site-scoped (limited to a particular network) DNS anycast addresses will be defined for use in individual IPv6 networks.

Anycast Developments

One of the barriers to Internet-scale deployment (as opposed to LAN or network scale) of IP anycast is the inability to aggregate the anycast address space. Thus for anycast routing to work beyond a confined topological region, one needs to advertise these anycast host routes beyond that region or across autonomous system (AS) boundaries which providers are loath to do. In fact providers are even filtering out longer addresses (e.g. host routes) to contain the growth in the Internet routing table.

A novel technique called Global IP-anycast (GIA) introduced a mechanism whereby it would be possible to deploy a global anycast routing and addressing scheme[13]. It was built on the interesting (and true) observation that anycast addresses represent a service and that some services are more popular then others. Therefore it makes sense to give precedence to those anycast routes that represent popular services.

GIA defined a new anycast address space that contains a field for the anycast provider’s home network. It placed anycast addresses into one of three classes: internal (inside the home network), external-unpopular (outside the home network and not frequented often) and external-popular (outside the home network visited frequently). Packets addressed to an internal anycast address were forwarded business as usual inside the home network. Packet addressed to external-unpopular address required routers to extract the home network address from the anycast address on a per-packet basis. And packets addressed to external-popular sites used a BGP-based query and response protocol to discover and then install anycast routes.

Investigation is underway into network services and applications that can benefit from IP anycast. Routing in mobile networks is a possibility given its ad-hoc nature and need to find the closest, useful routing node or server. Content routing is another area that can benefit from IP anycast. In this environment, a content client could contact the closest content server, who in turn to download the content to the client or initiate a search or query for a copy of the content. These two examples illustrate the applicability of IP anycast when network resources and content are distributed and the client will be satisfied with the first available instance. With the growth of the Internet coupled with the eventual move towards IPv6, IP anycast will certainly play more of a role in our Internet experience

[1] S. Bhattacharjee, M. Ammar, E. Zegura, V. Shah and Z. Fei. “Application-Layer Anycasting”. IEEE Infocom '97, Kobe, Japan.
[2] ftp://ftp.isi.edu/in-notes/rfc2460.txt
[3] http://ietf.org/rfc/rfc1546.txt?number=1546
[4] http://ietf.org/rfc/rfc2373.txt?number=2373
[5] ftp://ftp.isi.edu/in-notes/rfc2526.txt
[6] ftp://ftp.isi.edu/in-notes/rfc2461.txt
[7] ftp://ftp.isi.edu/in-notes/rfc2710.txt
[8] http://ietf.org/html.charters/magma-charter.html
[9] Basturk, E., Engel, R., Haas, R., Peris, Vinod, Saha, Debanhan, “Using Network Layer Anycast for Load Distribution in the Internet”, IBM Research Report, RC20938, July, 1997
[10] http://ietf.org/html.charters/ipv6-charter.html
[11] http://www.cisco.com/univercd/cc/td/doc/cisintwk/intsolns/mcst_sol/anycast.htm
[12] http://ietf.org/rfc/rfc3068.txt?number=3068
[13] Katabi, D., Wroclawski, J., “A Framework for the Scalable Global IP-Anycast (GIA)”, ACM/Sigcomm Proceedings, 2000