A Scalable Overlay Framework for Internet Anycasting Service

Chi-Jen Wu
Institute of Information Science
Academia Sinica, Taiwan
cjwu@iis.sinica.edu.tw

Ren-Hung Hwang
Dep. of CSIE
National Chung Cheng University, Taiwan
rhhwang@cs.ccu.edu.tw

Jan-Ming Ho
Institute of Information Science
Academia Sinica, Taiwan
hoho@iis.sinica.edu.tw

ABSTRACT

This study presents IAS, a scalable and efficient global overlay routing framework for Internet Anycasting Service. We introduce a new routing group concept and adopt the overlay network mechanism to achieve scalable and efficient inter-domain anycast routing. We show that the routing table size of an anycast router can be bounded by $O(\sqrt{N})$, where N denotes the number of anycast groups. We conduct simulations on a AS topology to verify this bound and show that routes found by IAS are very close to the shortest path when the size of anycast group is reasonably large.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Network Protocol-Applications

General Terms

Design, Performance

Keywords

Routing Protocol, Anycasting, Overlay

1. INTRODUCTION

The anycast service has been proposed in RFC 1546 [1]. Anycast is defined in the next generation network (IPv6) addressing architecture [2, 3] as a special routing model which allows a sender to access at least one in a group which shares the same anycast address. Ideally, the packet is sent to the nearest one in the group, where "nearest" can be defined according to the routing metric used by the routing protocol. For example, in Figure 1, sender 1 and sender 2 are sending packets to the same anycast group. If hop count is the routing metric, the anycast routing protocol will deliver sender 1's packet to member 1 and sender 2's packet to member 2, as they are the nearest servers, respectively.

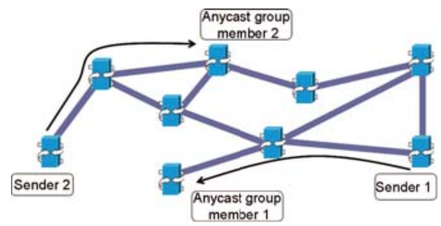

Figure 1: An example of network-layer anycast

Anycast can be used to develop numerous potential killer applications, e.g., DNS, replicated ftp or www servers. In most of these applications, anycast can be used to locate an appropriate server from a group of available servers, such as multiple mirrored web servers that share a single anycast address. Improving system reliability, global load distribution among distributed servers, and host auto-configuration are other promising anycast applications. In addition, anycast can significantly improve the network performance by routing packets to the nearest server [4]. For example, anycasting can support multiple rendezvous routers in PIM-SM multicast routing protocol.

Currently, how to provide a scalable and efficient global anycast routing is a challenge [7-8]. The current standard does not define any protocol for performing Anycast routing due to the lack of a scalable and feasible solution. A router to an Anycast address is treated by the routing system as a host router. As a consequence, a backbone router needs to have a routing entry for each anycast address. In other words, route aggregation cannot be done for anycast addresses since different anycast addresses share the same network prefix may have different next hop at each backbone router. For example, if the router supports 10,000 distinct anycast groups, 10,000 additional routing entries will appear in the routing table. As the anycast service becomes more and more popular, the routing tables of backbone routers for Anycast service will become too large to be tractable.

Our study is to design a novel scalable and feasible anycast routing protocol, called IAS, for global Internet. We propose a new routing group concept and adopt the overlay network mechanism to achieve scalable and efficient inter-domain anycast routing. Based on a global hash function with uniform partition property, anycast routers and anycast addresses are divided into routing groups in a distributed manner. Anycast routers of the same routing group will self-organize into an overlay network. They exchange routing information based on an extension of the BGP protocol such that each router knows how to route anycast addresses of the same routing group. By dividing anycast routers and addresses into $\sqrt{N}$ routing groups, the size of routing table for anycast can be bounded by $O(\sqrt{N})$, where N denoted the number of anycast groups. Besides, IAS allows IAS-enabled routers coexist with non-IAS-enabled routers which facilities gradual deployment of Internet anycast service.

The rest of this paper is organized as follows. In Section 2 we describe current related work of anycast, and revisit the problems of IP anycast. Section 3 and 4 illustrates the overview and the design details of our anycast routing protocol. Numerical results are shown in Section 5. Finally, Section 6 concludes this paper and presents our future work.

2. RELATED WORK

Anycast routing research can be broadly classified into network-layer and application-layer. Network-layer (or IP) anycast focuses on developing anycast routing algorithms which route packets to one of the anycast servers with fewest hops or lowest cost. Since routes cannot be aggregated as that has been done in Classless Inter-Domain Routing (CIDR), scalability (routing table size) and efficiency (IP look up process) are the major challenges. Developing a new anycast routing protocol also handicaps the speed of deployment, in particular, if the new routing protocol needs to be installed in all routers. On the other hand, application-layer anycast [9] tries to solve the problem at the application layer, without the involvement of routers. It focuses on the problem of server selection which tries to choose one of many identical servers based on some performance metrics, such as server capacity, response time, or server loading, etc. In the following, we summarize current research.

Network-layer anycast routing can be inter-domain and intra-domain. In this article, we focus on inter-domain anycast routing problem, because some intra-domain routing protocols, like RIP, already have the ability to provide anycast service [5-6], whether the scheme scales or not is a different issue.

The first noticeable paper that discussed the global network layer anycast routing problem is Global IP Anycast (GIA) [5]. In GIA, anycast address, violates the definition in [3] has its own address space. Anycast addresses are further classified into three classes, according to their popularity. Periodically, GIA router exchanges information with its neighbors to maintain good routes to popular anycast addresses. Therefore, packets destined to popular anycast addresses will be routed to the nearest anycast server based on this information. On the other hand, packets destined to unpopular anycast addresses will be routed to the home network of the anycast address. GIA offers a scalable solution for network layer anycast, but it is an on-demand query-based protocol. As a consequence, as anycast service becomes commonly used, enormous query messages may become a scalability problem.

Recently, Ballani *et al.* proposed the PIAS [10]. The basic idea of PIAS is to solve the scalability problem of network-layer anycast by deploying an overlay of proxies. Specifically, a large number of anycast proxies are deployed around the Internet and the anycast address information are logged at the proxies. Routers are not responsible for delivering anycast packets. Rather, anycast packets are delivered to the nearest anycast proxy which in turn delivers the packet directly to an AS using unicast. Although the burden of routers has been alleviated, deploying a large number of proxies may take a while.

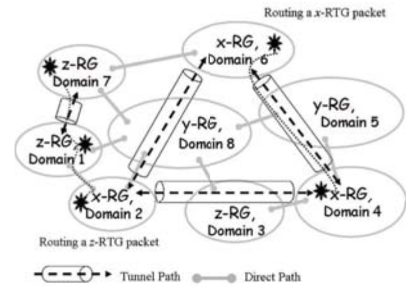

Figure 2: Overview of Internet anycasting service

3. OVERVIEW OF IAS

To address the scalability problem of anycast routing, we introduce a novel routing group (RG) concept. The basic idea of RG is to ***divide and conquer*** the large number of anycast groups (addresses). Specifically, a global hash function is used to map anycast group addresses into routing groups. A backbone router also uses the same hash function to decide which routing group it belongs to, using one of its anycast addresses as the input (or any unicast addresses if it does not have any anycast addresses). An anycast address is mapped to a routing group using SHA-1[11] as a based hash function in our routing framework. The SHA-1 is employed in the design of peer-to-peer systems [12] extensively. In our design, the global hash function is a very important component; it should distribute anycast routing entries to produce a balanced distribution in our routing group. Although SHA-1 distributes data uniformly with high probability, further study is still undergoing to verify its performance on distributing anycast groups. The number of routing groups is a system parameter. Ideally, it is suggested to set to $\sqrt{N}$, where N is the number of anycast groups. Then, the most important idea of our RG is that all routers in the same routing group, say group A, know how to route anycast packets with addresses that are mapped to group A.

Anycast packets are routed as follows. When a router receives an anycast packet, it first uses the hash function to determine which group it belongs to. If it belongs to the same group of the router, the router knows how to forward this anycast packet. If not, the router forwards it to a nearby router which belongs to the same group as the anycast packet. The packet can then be successfully routed by routers of that routing group.

Figure 2 shows the example of how to route anycast packets. In this example, there are three routing groups, namely, x-RG, y-RG, and z-RG. Assume that each domain has only one anycast address and one IAS router. Both router and anycast address of the same domain are mapped to the same routing group. So, anycast addresses of domain 1, 3, and 7 belong to z-RG, that of domain 2, 4, and 6 belong to x-RG, and that of domain 5 and 8 belong to y-RG. The first example shows that a packet destined to domain 6 received by domain 4 is directly sent to domain 6 through the tunnel between them. The second example shows that a packet originated from domain 2 (X-RG) destined to domain 7 (Z-RG) is sent to a nearby router of Z-RG first, in this case, domain 1. The packet is then routed to domain 7 based on domain 1's anycast routing table.

If the number of routing groups is and anycast addresses are evenly distributed into routing groups, it is trivial to show that the anycast routing table of each router is bounded by $O(\sqrt{N})$, where N denoted the number of anycast groups. But in the reality circumstance, we recommend using the maximum Autonomous System (AS) number (65536) to be the N value due to the N value isn't easy to obtain. (In this paper, we don't discuss the N value, we remain it to solve in future work)

4. DESIGN DETAILS

This section describes the details of IAS architecture. We assume that anycast addresses are within the unicast address space, but can be differentiated. For example, within each subnet, the highest 128 interface identifier values are reserved for assignment as subnet anycast addresses [13]. Each IAS-enabled router will maintain an anycast routing table which consists of next-hop information to other routing groups and routing table for anycast addresses of its own group. Figure 3 shows the IAS routing table of the router in domain 1 in Figure 2[1].

There are three mechanisms in IAS. First, neighbor discovery is used to construct overlays of routing groups. Address registration is then applied to register anycast addresses. Finally, routing table exchange allows routers of the same routing group exchange routing information of anycast addresses of the same routing group. In the following sections, we describe these three mechanisms in detail.

4.1 Neighbor discovery

Each router needs to know several nearby routers of the same group and one router of each of the rest of routing groups. We propose a search mechanism to discover the routers and construct an overlay network for each routing group. The search mechanism runs at BGP border routers. Two types of message are flooded using IP multi-cast: a search message and a reply message. The message format and discovery operation are very similar the GIA search protocol [5]. The search message is flooded by the source router (SR) as a TTL-scoped expanding-ring searching mechanism that explores the neighboring IAS routers. TTL is properly set such that the search message reaches at least one IAS router of each routing group. Upon receiving a search message which is not duplicate (by checking the message ID), an IAS router will send back a reply message which contains its unicast address and its anycast routing table (with routing cost). At the same time, the receiving router will also flood the search message to its neighbors by decreasing TTL by one, unless TTL reaches 0 after decrement. Upon receiving a reply message, the SR checks if it is from a new IAS router of its own routing group. If yes, a tunnel is built such that this IAS router becomes one of its neighbors in the overlay network. (To avoid too many neighbors, a SR can selectively choose its neighbors according to some routing metric) If not, the SR checks whether the responding

[1]We assume only an IAS router in a domain in this example and the domain refers to a routing domain or an autonomous system (AS).

Domain 1's IAS Routing Table

X-RG	
X-RG	Domain 2

Y-RG	
Y-RG	Domain 8

Z-RG	
Anycast address	Next hop
FE80::2AA:FF:FE9A	Domain 3
FE80::4CA2:FE0C	Domain 7

Figure 3: The IAS routing table of the domain 1 border router. We use the domain to indicate the IP address. Note that we also ignore the routing cost field.

IAS router is a better next hop that the routing group that IAS router belongs to. If yes, the corresponding entry of the anycast routing table is updated and a tunnel is built to replace the old tunnel. The SR may keep more than one routing entry to a routing group for consideration of fault tolerant or supporting different routing metrics. The discovery mechanism is triggered only when a router starts up. After the discovery, a router will runs a routing protocol, e.g., BGP, to maintain the anycast routing table for anycast addresses of the same routing group. In addition, keep alive messages are periodically sent to next-hop routers to maintain the tunnels.

In summary, after the neighbor discovery procedure, an IAS router knows several neighboring IAS routers of its own routing group and at least one neighboring IAS router of each of other routing groups. Because the discovery mechanism is based on TTL-scoped expanding-ring searching, neighboring IAS routers found should be close to the SR in distance of hops.

4.2 Address registration

An IAS router needs to register all anycast addresses that belong to its interfaces. For each anycast address, if the address is mapped to the same routing group of the router (based on the global hash function), the router simply adds an entry to its anycast routing table with cost of zero and exchanges this new routing information with its neighbors on the overlay via the routing protocol later on. If the address belongs to other group, the router sends a registration message to a router of that group. (Recall that, via the anycast routing table, each router knows at least one router of each of other groups) The router that receives the registration message will add a routing entry for this anycast address. It is also responsible for forwarding packets destined to this anycast address to the originating router.

Anycast group membership management is another research issue. The interested readers are referred to the work by IETF Multicast and Anycast Group Membership Announcement (MAGMA) working group [14] which tries to extend group management protocols to support anycast.

4.3 Routing table exchange

Routers of the same routing group will run the distributed Bellman-Ford algorithm or BGP's path vector routing algorithm to exchange their anycast routing information such that anycast packets destined to this group should be able to be routed successfully. Neighboring routers found in the neighbor discovery stage are viewed as the router's routing peers (or BGP peers). Therefore, routers of the same routing group will form an overlay network.

In summary, let us present a complete routing example in IAS. A packet destined to an anycast address is received by an IAS router. The router either knows how to route it, if the destination address belongs to the same group of the router, or forwards it to a nearby IAS router that belongs to the routing group of the destination address. As the packet reaches the final stop, i.e., the router that this address registered with, the router either sends the packet to one of its directly connected subnets or forward the packet to the router that originally sends the registration message of this anycast address.

During the transition period, the Internet will consist of IAS-enabled routers as well as non-IAS-enabled routers. The coexist of both kinds of routers only affect the efficiency (quality) of routing, will not ruin the operation of IAS. Consider a network with both kinds of routers, an anycast packet is sent to a non-IAS router. Since the anycast address is within the address space of unicast, the non-IAS router will route it as if it were a normal unicast packet, i.e., will route it to the home network of the anycast address. As long as there exists an IAS-enabled router on the routing path, the packet will be routed to the nearest anycast server as the IAS-enabled router will notice that it is an anycast packet and will then route it using its anycast routing table. In the case where there is no IAS-router on the routing path, the packet will be routed to the anycast server at the home network. Since the IAS-enabled routers of each routing group will form an overlay network, the routing IAS will be successful as long as at least one neighboring router is found for each routing group during the neighbor discovery stage.

5. SIMULATION RESULTS

This section describes the simulation results for evaluating the performance of IAS. Two experiments are designed to evaluate the performance of IAS from various aspects. The first experiment evaluates the scalability of IAS by measuring the size of anycast routing table under various anycast group sizes. Meanwhile, the second experiment studies the efficiency of the IAS's routing path in terms of hop stretch. The hop stretch refers to the ratio of average routing hop-count on the IAS routing path to the shortest path to the nearest server on the underlying IP network. Note that although hop count is used as the routing metric in our simulations, other routing metrics, e.g., average latency, network bandwidth, can also be adopted in IAS.

The network topology simulated was the AS topology is generated by the BRITE topology [15] using the Waxman model where alpha and beta are set to 0.15 and 0.2, respectively. In addition, HS (size of one side of the plane) is set to 1000 and LS (size of one side of a high-level square) is set to 100. Totally, the AS topology consists of 10,000 nodes.

In our simulations, members (routers) of an anycast group are randomly selected from the simulated network. In some cases, size of anycast group is set proportional to the network size. When the group size is set to 0.02% in the AS topology, the average group size is 2. All members of the same anycast group will share the same anycast address. The TTL value of neighbor discovery mechanism is set to 3 in all simulations. In following figures, each data plotted is the average of 100 runs.

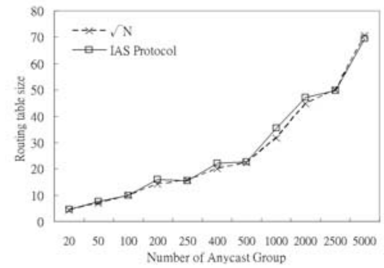

Figure 4: Routing table size under various number of anycast group.

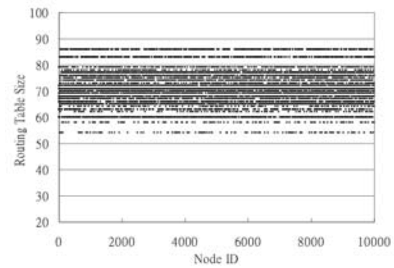

Figure 5: Average routing table size of each router with 73 routing groups.

5.1 Routing table size

We first measure the size of the IAS anycast routing table under various number of anycast groups. In the traditional anycast routing approach, the anycast routing table grows in proportional to the number anycast groups. In IAS, we expect the routing table size can be bounded by O($\sqrt{N}$), where N is the number of anycast groups. In Figure 4, we compare the measured anycast routing table size (only counts entries for the same group) with $\sqrt{N}$. As we can observe from Figure 4, the routing table size is approximated by $\sqrt{N}$ very well. To look further into detail, Figure 5 shows the routing table size of each router in the AS network topology with 73 routing groups, 5,000 anycast groups, each group consists of 2 members. The table size varies from 55 to 86 with an average of 70. The variation comes from randomness and non-perfect hash.

In summary, each IAS router maintains about $\sqrt{N}$ routing entries for anycast ad-dresses of the same routing group and ($\sqrt{N}$-1) routing entries to routers of other routing groups.

5.2 Routing overhead

In the second experiment, we evaluated the efficiency of the IAS routing path by comparing it against the optimal path, where the term 'optimal path' refers to the shortest path to the nearest server using priori knowledge of the nearest server. As aforementioned, routing path found by IAS may be longer or not to the nearest server, due to the overlay structure.

Figure 6 (a) shows the stretch, i.e., the ratio of the average hop-count of IAS route to that of the optimal routing path, under various numbers of routing groups in the AS network topology where the anycast group size is set to 0.02% (the average anycast group member is 2). The average stretches

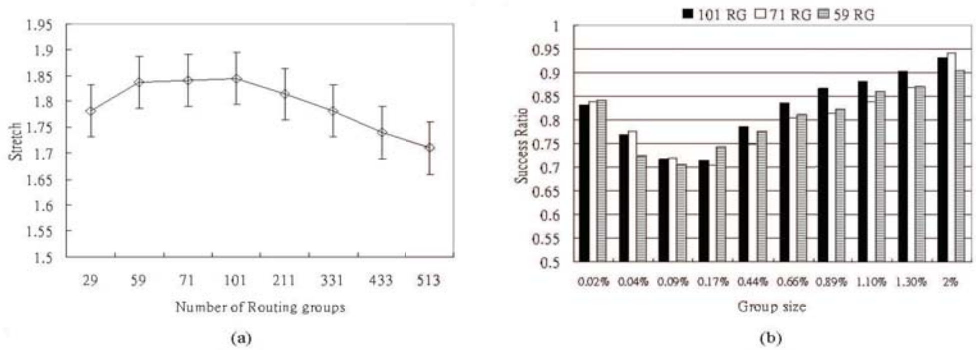

Figure 6: (a) Average stretch under various numbers of routing group (the average anycast group member is 2). (b) The probability a routing finds the nearest anycast member under various anycast group sizes.

with 95% confidence interval are plotted in Figure 6 (b). The confidence intervals are computed over 100 independent runs. The simulation indicates that the routing overhead is not very significant; only around 1.8. Figure 6 (a) also implies that the number of routing groups is not a significant factor of the routing overhead. In particular, the variation of stretch due to the increase of routing group number is very insignificant. In the next experiment, we measure the probability that an anycast packet is successfully routed to the nearest server. This probability indicates that how often an anycast packet is really routed to the nearest server, not servers that are farther away. Figure 6 (b) shows the successful probability of IAS under different group sizes and number of routing groups. Again, as the group size is large, the successful probability is reasonably high. Most of the successful rates are higher than 0.75.

6. CONCLUSION AND FUTURE WORK

Anycast is a very promising service for future Internet applications. However, scalable and efficient anycast routing is still an open issue. In this paper, we proposed a novel anycast routing framework based on the routing group concept and BGP. The proposed framework follows the IPv6 standard such that anycast addresses are within the address space of unicast. It is scalable since the size of routing table can be bounded by $O(\sqrt{N})$, where N is the number of anycast groups, when the number of routing group is set to $\sqrt{N}$. It is efficient since it has high probability to route anycast packets to the nearest servers and the routing overhead, in terms of stretch, is quite low. It is also easy to deployment since it allows coexist of IAS-enabled routers and non-IAS-enabled routers in the Internet. However, we are at the very early stages of our investigation. We need to better understand how IAS's performance in a sparse IAS deployed network or in the real Internet environment. All of these items are future work.

7. REFERENCES

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

[2] S. Deering and R. Hinden. Internet Protocol, Version 6 (IPv6) Specification. *IETF RFC 2460*, December 1998.

[3] R. Hinden and S. Deering. IP Version 6 Addressing Architecture. *IETF RFC 2373*, July 1998.

[4] R.Engel, V.Peris and D.Saha. Using IP Anycast for Load Distribution and Server Location. *In Proceedings of Global Internet*, December 1998.

[5] Dina Katabi, John Wroclawski. A Framework for Global IP-Anycast (GIA). *In Proceedings of ACM SIGCOMM*, April 2001.

[6] Xuan, W. Jia, W. Zhao, and H. Zhu. A Routing Protocol for Anycast Messages. *IEEE Transactions on Parallel and Distributed Systems*, Volume 11, pp.571-588, Jun. 2000.

[7] Sott Weber and Liang Cheng. A Survey of Anycast in IPv6 Networks. *IEEE Communication Magazine*, Volume 11, pp.571-588, Jun. 2000.

[8] Satoshi Doi, Shigo Ata, Hiroshi Kitamura and Masayuki Murata. IP Anycast point-to-(Any) Point Communication. *IEEE Internet Computing*, Vol. 6, No. 2 , pp. 94-98, March 2002.

[9] S. Bhattacharjee, M. H. Ammar, E. W. Zegura, N. Shah and Z.Fei. Application Layer Anycasting. *In Proceedings of IEEE INFOCOM*, April 1997.

[10] H. Ballani and P. Francis. Towards a Deployable IP Anycast Service. *in Proceedings of ACM SIGCOMM*, 2005.

[11] FIPS 180-1. Secure Hash Standard. U.S. Department of Commerce NIST. *National Technical Information Service*, Springfield, VA, Apr. 1995.

[12] Ion Stoica, Robert Morris, David Liben-Nowell, David R. Karger, M. Frans Kaashoek, Frank Dabek, Hari Balakrishnan. Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications. *IEEE/ACM Transactions on Networking*, Vol. 11, No. 1, pp. 17-32, February 2003.

[13] D. Johnso and S. Deering. Reserved IPv6 Subnet Anycast Addresses. *IETF RFC 2526*, March 1999.

[14] Multicast and Anycast Group Membership charter, IETF. http://www.ietf.org/html.charters/magma-charter.html.

[15] A. Medina, A. Lakhina, I. Matta, and J. Byers. BRITE: An Approach to Universal Topology Generation. *In Proceedings of ACM MASCOTS*, pp. 346, August 2001.