Using Anycast for Improving Peer-to-Peer Overlay Networks

Le Hai Dao^{O} and JongWon Kim
Networked Media Lab., Department Information and Communications,
Gwangju Institute of Science and Technology (GIST), Gwangju, Korea
E-mail : {daolehai^{O}, jongwon}@gist.ac.kr

Abstract

Peer-to-Peer (P2P) overlay networks have drawn much research interest in the past few years because they provide a good substrate for large-scale applications in the Internet. In this paper, we introduce the use of anycast, a new "one-to-one-of-many" communication method in the Internet, to solve the following common problems of P2P overlay networks: load-balancing, topology-awareness, system partitioning, and multi-overlay interconnection. We also give an analysis of the features and limitations of the recently deployed anycast infrastructures in the Internet for supporting P2P overlay networks.

1. Introduction

Anycast [1] is a new "one-to-one-of-many" communication method in the Internet that is designed to deliver a request to the nearest one of many hosts in the same anycast group. It has been included explicitly in the IPv6 protocol definition and there has been a lot of research work on retrofitting anycast in IPv4 as well. There are a number of research proposals on utilizing anycast for improving the current and future Internet applications such as providing robust and efficient service discovery [2], creating DDoS sinkholes [3], building efficient multicast tree [4], etc. However, applying anycast for improving the performance of peer-to-peer (P2P) overlay networks is still an open research area.

P2P overlay networks were born to solve the critical issue of scalability in client-server architecture. In a P2P overlay network, nodes are self-organized in a distributed system. Each node has roles of both client and server. P2P overlay networks are very suitable for services that require fault-tolerance, self-organization and massive scalability properties. Therefore, P2P overlay networks have been an important substrate of the Internet. However, there are a lot of limitations in P2P overlay networks that are impacting their development. A number of approaches have been proposed to date for improving P2P overlay networks; however, most of them are specific for one or a limited number of P2P overlay systems.

In this paper, we analyze several critical common problems of P2P overlay networks and propose the use of anycast to solve them. We also give an analysis of the features and limitations of the recently deployed anycast infrastructures in the Internet in supporting P2P overlay networks. In our approach, each overlay network has one anycast address and nodes in that network are also in the same anycast group. The following goals can be achieved:

- Node initialization and load-balancing. Anycast can help nodes to join an overlay network more easily by just sending a request to the anycast address of the overlay. This technique also can be used to distribute the bootstrapping load to a large number of nodes existing in the overlay but not only to some specific nodes, known as well-known nodes, as in traditional P2P overlay networks.
- Topology-awareness. Anycast can be used for nodes to discover their nearby nodes in the underlying network. From this knowledge, nodes then organize the overlay network topology to closely map the underlying network.
- System partitioning. After network failures, an overlay network may be partitioned into multiple isolated islands. Using anycast, we introduce a mechanism for nodes in those islands to automatically link to each other to hint the overlay network.
- Multi-overlay interconnection. The use of P2P overlay networks will be expanded in the future and there may have a need for linking multiple overlay networks together, and anycast can be used for this purpose.

The remaining of this paper is structured as follows. Section 2 gives a background on anycast and introduces about its current implementations on the Internet. Section 3 introduces an overview of P2P overlay networks and their classification. In Section 4, we analyze the above problems and present the use of anycast for improving P2P overlay network performance in details. Section 5 gives our analysis on the currently deployed anycast infrastructures in supporting P2P overlay networks. We conclude this paper in Section 6.

2. Anycast and its Implementations in the Internet

The original definition of anycast in RFC1546 [1] is: "A host transmits a datagram to an anycast address and the inter-network 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."

In short, anycast is a point-to-point flow of packets between a single client and the nearest destination server identified by an anycast address. However, this communication is not stable because the condition of the underlying network can change and then the destination can be changed. The basic idea behind anycast is that a client would like to send packets to a server offering a particular service or application, but it is not important which server is chosen. To accomplish this, a single anycast address is assigned to one or more servers within a so-called anycast group. A client sends packets to the server by using the anycast address. Just as with a unicast flow, the client and server are unaware that anycast is used. An anycast service, when implemented at the network layer, is called *Network-layer* or simply *IP anycast.* In IP anycast, after receiving an anycast request, the network of routers will then attempt to deliver the packet to the closest server in the destination anycast group, which will then reply.

There have been a lot of proposed anycast versions to date; however, few of them have been deployed globally such as IP-anycast for DNS root-servers [5], Proxy IP Anycast Service (PIAS) [7], Internet Indirection Infrastructure (i3) [6] and Overlay-based Anycast Service Infrastructure (OASIS) [8]. In the following sub-sections, we will briefly introduce about them.

2.1. IP-anycast for DNS root-servers

In [5], a technique is introduced to provide anycast service for distributing loads and reducing the network distance to the users of the DNS root-server systems. Note that, the current Internet is based mainly on IPv4 with no specific support for anycast. Each server has two IP addresses, one is for management and other none-service traffic, and the other is the service address that is known globally as the anycast address of the service. The objective here is to distribute the anycast address very widely, rather than over a single subnet. To do so, a BGP announcement that covers the anycast address is sourced in different locations by a set of allocated hosts and network components which are capable of autonomously providing the service. The idea is that BGP will see the different paths towards the different anycast instances, and route the requested datagrams from clients to the best server. This approach can be applied for other services as well.

2.2. PIAS

PIAS [7] has been proposed to overcome the limitations of the traditional IP anycast (scalability, global implement ability and flexibility) while adding new features to support a wide range of applications. PIAS is deployed as an overlay infrastructure with a system of anycast proxies located over the Internet of different ISPs. The proxy system acts as the core of global anycast routing. It memorizes all of anycast group members' addresses and advertises its known anycast addresses into the routing fabric (BGP, IGP) in border routers. The anycast address can be an IP anycast or a combination of an IP and a Layer 4 port number. Requests from clients reach a proxy through native IP anycast and they are redirected to the best destination over a unicast tunnel made by that proxy.

2.3. i3

i3 [6] is an overlay anycast infrastructure that provides a way to address packets using IDs but not their actual destinations, the IP-addresses. To map from IDs to IP addresses, a system of servers is deployed globally to provide a translation fabric. When a sender wishes to contact a receiver, it sends the packet to the i3 network along with the destination ID. The i3 network is then responsible for determining the IP address of the final receiver and delivering the packet.

i3 also supports for inexact matching so that the destination ID of an incoming packet needs only match *k* over *m* bits of a mapped ID to be considered a complete match. This technique could be used to provide anycast in i3. The members of an anycast group choose IDs with a common prefix with length *k*. An incoming packet is then delivered to the anycast destination whose ID is most closely matches the packet's destination ID. For network proximity, the global network is divided into multiple landmarks; each landmark is assigned a prefix in the remaining bits (less than *m-k* bits in length), and the i3 server in this landmark will deliver the packets to the best matched destination.

2.4. OASIS

OASIS [8] is an overlay-based anycast service infrastructure that helps clients to find the best server of a service on the Internet. It is designed based on a set of core nodes whose locations are distributed over the Internet. This overlay anycast uses a set of techniques to measure the entire Internet in advance with low overhead introduced to the underlying network. In contrast to the above systems, OASIS uses a URL for addressing a service.

For redirection purposes, each core node keeps information about all registered services and their live servers, the coordinate of those servers in a geographic coordinate, and the locality information of all known prefixes. A live server that is closest to the requesting client in virtual space will be returned.

All servers of a registered service run a short OASIS-specific code in order to communicate with the OASIS core nodes. The core nodes then measure and map all of those servers into locations in a coordinate space with the other services. Requests from a client can be redirected to a suitable server by using DNS or HTTP redirection.

3. P2P Overlay Networks - An Overview

A P2P overlay network is a distributed system self-organized by two or more peers to collaborate spontaneously in a network of equals (peers) without the need for central coordination. P2P overlay networks are designed to provide a mix of various features that the IP-network unable or difficult to provide, such as robust wide-area routing architecture, efficient search for data items, selection of nearby peers, massive scalability, permanence, hierarchical naming, trust and authentication, redundant storage, anonymity and fault-tolerance. In contrast to the client-server systems, peers in P2P overlay systems have symmetry in roles of client and server. The P2P overlay networks can be classified into two main categories: Unstructured and Structured.

In unstructured P2P systems (for example Freenet [9], Gnutella [10], KaZaA [11]) peers join the network with some loose rules, without any prior knowledge of the topology. In general, flooding mechanism is used for peers to send queries across the overlay with a limited scope. When a peer receives the flood query, it sends a list of all matched contents to the source peer. While flooding-based techniques are effective for locating highly replicated items and are resilient to peers joining and leaving the system, they are poorly suited for locating rare items. Moreover, this class of P2P overlay network has the scalability problem since the load on each peer grows linearly with the total number of queries and the system size.

Structured P2P overlay systems, such as CAN [12], Chord [13], and Pastry [14] use the Distributed Hash Table (DHT) as a substrate to facilitate an efficient searching process. Each node acts as a server for a subset of data items. The operation *lookup(key)* is supported, which returns the node ID storing the data item with that key. The values of the node could be data items or pointers to where the data items are stored. Each data item is associated with a key through a hashing function. Nodes have identifiers, taken from the same space as the keys. Each node maintains a routing table consisting of a small subset of nodes in the system. In this way, an overlay network is constructed that captures logical connections between nodes. Usually, the logical network is a regular network such as a ring, tree, mesh, or hypercube. Lookup queries or message routing are forwarded across overlay paths to the peers in a progressive manner, with the node IDs that are closer to the key in the identifier space. Typically, a node sends one query for each received lookup request. In theory, DHT-based systems can guarantee the lookup path length of *O(logN)* overlay hops on the average where *N* is the number of peers in the system.

4. The Possible Applications of Anycast for P2P Overlay Networks

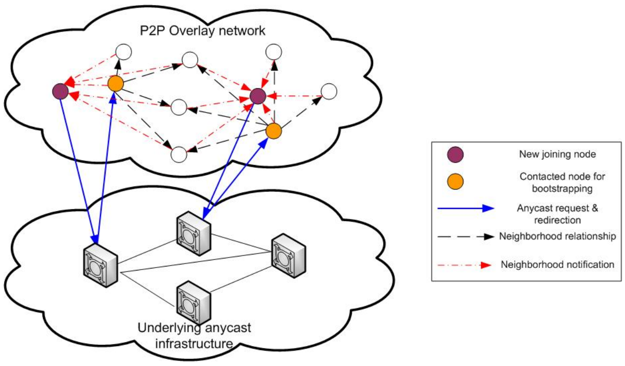
P2P Overlay network
Underlying anycast infrastructure
New joining node
Contacted node for bootstrapping
Anycast request & redirection
Neighborhood relationship
Neighborhood notification

Figure 1: With the support of anycast, new joining nodes can easily discover nearby nodes in the overlay.

In this section, we introduce several possible applications that anycast can bring to P2P overlay networks. We assume a global anycast infrastructure in the Internet that is highly resilient to network failures. The general design is: all nodes in an overlay network are also in the same anycast group. This anycast address is unique and known in the anycast infrastructure, and then any request from a node outside will be redirected to a node existing in the overlay network.

4.1. Node Initialization and Load-Balancing

In a P2P overlay network, when a new node wants to join the overlay, it has to know at least one node existing in the overlay. Normally, there are some well-known nodes (or bootstrapping nodes), which are highly available and powerful, are either listed for users to manually choose or integrated in the program running in the users' machines. In a very large network with millions of users, these well-known nodes have to serve for a huge number of requests that can make them overloaded. It is even more serious in the overlay systems that require the bootstrapping node to transfer initial data to the new joining nodes. In another case, DDoS attackers can easily find out the well-known nodes' IP addresses to attack. This problem is also known as the single point of failure problem.

By using anycast, we can balance the bootstrapping load to all existing nodes in the overlay network but not only the limited number of well-known nodes. Because all nodes in an overlay network are also in the same anycast group, when a new node wants to join an overlay, it just sends a join request to the anycast address of the overlay. The anycast infrastructure redirects the request to an existing node in the overlay that will do bootstrapping (as illustrated in Figure 1). When the new node has finished joining the overlay, it joins the anycast group as well. For a specific load-balancing purpose, we can register specific criteria for the redirecting strategy to the anycast infrastructure, e.g. redirect based on the network latency or redirect based on the load of the serving node.

4.2. Topology-Awareness

Topology-awareness is one of the great drawbacks of P2P overlay networks, i.e. the P2P overlay topology does not match the underlying IP network topology. This may cause the lookup query to travel significantly far in the underlying network before reaching the destination. This problem is more critical in structured P2P overlay networks because of the two main reasons. First, since nodes in a structured P2P overlay network are located in a fixed map based on the random ID assignment mechanism that does not take into account the underlying network topology, their neighborhood relationship is random as well. In contrast, nodes in an unstructured P2P overlay network can freely change their neighborhood relationships with other nodes that have better metrics, e.g. network latency. Second, because unstructured P2P overlay networks use flooding technique to solve a lookup query then there may be multiple destinations for the same content, thus the lookup source can receive the response very fast. In structured P2P overlay networks, there is only one or some fixed path for requests of a node to travel to the destination of a lookup; therefore, even though the lookup path length is *O(logN)* hops in average, the lookup query may travel significantly far in the underlying network.

The lookup performance of any P2P overlay networks depends on how nodes build their routing table. If the routing table is built based on the network latency between nodes, the lookup time will be improved. But the question here is "how does a node know who are locating close to it?" This question is more difficult for a new node that has just joined the overlay. Knowing the surrounding nodes will help a node to discover its location in the underlying network. We can utilize the support of an anycast infrastructure to overcome this challenge. Since anycast tends to redirect a request to a nearby, or ideally the nearest, node in the target group (i.e., a P2P overlay network), and assuming that every node in the overlay has a list of nearby nodes in the underlying network, called the neighbor list; a new joining node can easily learn from its bootstrapping node about nearby nodes to build its own neighbor list (see Figure 1). Moreover, some anycast infrastructures such as PIAS and OASIS support for a node in an anycast group to discover its nearby nodes in the same anycast group. When receiving a request from a node in an anycast group, the anycast infrastructure will either redirect the request to or return a list of nearby nodes in the same anycast group to the node. Using this technique, nodes then organize the overlay network topology that is closely mapped the underlying network topology.

4.3. System Partitioning

The third issue in P2P overlay networks is the possibility of an overlay to be partitioned into several isolated islands after network failures. This problem may occur occasionally in practice since most of users in the current Internet are located in stub networks, e.g. networks in an ISP or in an organization. When the network connections to the Internet in a stub network are down in some time, all overlay nodes in this area can be isolated even after the physical connections are recovered. When happens, this problem will seriously affect the applications that running on the overlay network. This is a difficult problem because the overlay nodes do not know what is happening in the underlying network, they just remove their routing entries after timeout.

We utilize the above mentioned neighbor discovery feature of anycast as a mean to reconnect the peers after network recovery. Each node periodically discovers a number of nearby nodes but with a long time period for not to burden too much the anycast infrastructure. However, this is not always enough to solve the problem because a peer may only discover others in the same stub network. We can utilize another feature supported by PIAS, i3 and OASIS that allows peers to find arbitrary other peers in the overlay networks. Once there is at least a link between two nodes in two isolated islands, those islands are merged.

4.4. Multi-Overlay Interconnection

In the current Internet, each P2P overlay network is used for a specific application, e.g. file sharing, voice over IP, distributed storage, etc. In the future, there would be a need for automatically interconnecting multiple overlay networks for supporting users' interests since a P2P overlay network is only suitable for a limited number of services but a user's interests can be many. This new feature will help integrating multiple services in only one application. One of the challenges now is how and where to link different overlay networks together. A trivial solution is for users to manually select some fixed service point to join a new overlay network when they want to have a new service. However, this solution is not suitable in the case of multiple overlay networks are used in multiple service layers.

In our approach, since each overlay network is assigned a unique anycast address, a node needs only to send join requests to the interested overlays to participate in the services that are served by these overlay networks. A node can freely join and leave several overlay networks at a time. This solution is simple, scalable and transparent to the upper layers.

5. An Analysis on the Currently Deployed Anycast Infrastructures for Supporting P2P Overlay Networks

In this section, we give a qualitative analysis and comparison on the features of the current anycast implementations (as presented in Section 2) that are suitable for supporting P2P overlay networks. Table 1 shows our comparison.

The currently deployed IP anycast turns out to be not suitable for supporting P2P overlay networks. The main issue is the difficulty of deploying a dynamic application upon because when joining an anycast group, a node needs to do several tasks in the IP layer such as self-configures an additional IP address for anycasting, manages the two addresses simultaneously, announces a new network to the local router, etc. Moreover, IP anycast is not scale since each anycast group needs one unique IP address and this address cannot be aggregated. IP anycast also does not have neighbor discovery functionality that needed for improving topology-awareness in P2P overlay network because requests from an anycast group member to the same anycast address will be routed to itself.

Table 1: A comparison of currently deployed anycast infrastructures in supporting P2P overlay networks

	IP-anycast	PIAS	i3	OASIS
Current deployment range	Global	Limited area	Global	Global
Ease of overlay service deployment	Difficult	Easy	Easy	Easy
Proximity accuracy	Fair	Good	Poor	Good
Neighbor discovery	N/A	Good?	Poor	Fair
Random peer discovery	No	Yes	Yes	Yes
Scalability in number of anycast groups	No	Yes	Yes	Yes

The issues of IP anycast can be overcome by proxy-based IP anycast (PIAS) and overlay anycast services such as OASIS or i3. In PIAS, since all anycast requests are served by anycast proxies, then the anycast infrastructure is more flexible to support dynamic systems such as P2P overlay networks (compared to IP anycast). For example, nodes can join and leave an anycast group more easily; the number of supported anycast addresses is much larger by the combination of IP addresses and port numbers for anycast addresses; the target selection criteria can be vary. Also, the functionality of neighbor discovery is promised to be included in PIAS; however, it is not designed yet implemented. One of the biggest challenges of PIAS is global deployment since the proxy system should be able to communicate with the border routers in the Internet which are owned by ISPs to advertise the anycast addresses. Because of this reason, PIAS is now only deployed in some limited areas in the Internet.

In contrast to PIAS, OASIS and i3 provide anycast in the application layer that is easier to deploy in the current Internet. All target anycast addresses (IDs in i3 or URLs in OASIS) are translated into IP addresses and requests are redirected to the destinations. The proximity accuracy of i3 is poor because it does not have a mechanism to measure network distances but divides the Internet into multiple regions. Nodes are in a region will be assigned the same prefix which is used for nearby node discovery. PIAS and OASIS use mechanisms to measure the Internet in advance to achieve good network proximity accuracy. With regard to this metric, IP-anycast was expected to have the best performance; however, the evaluation in [15] shows that in many cases inter-domain routing, designed with unicast path-selection in mind, chooses anycast locations which are not close to the source. In OASIS, neighbor discovery is also included that allows to return a couple of nearby nodes to the requested node in the same anycast group. However, for better supporting topology-awareness in P2P overlay networks, an anycast infrastructure should adapt for several new neighbor discovery requirements from the overlay nodes such as a large number of discovered neighbors, the neighbors in a specified region to the node, etc.

6. Conclusion

In this paper, we introduced a use of global anycast infrastructures to solve several common critical issues of P2P overlay networks such as: load-balancing, topology-awareness, network partitioning, and multi-overlay interconnection. Our solution is simple and feasible in the current and future Internet by the use of one among several currently deployed global anycast infrastructures. We also analyzed these infrastructures for their features and limitations in supporting a P2P overlay network.

Acknowledgement

This work is supported by the basic research project through a grant provided by the Gwangju Institute of Science and Technology in 2007.

References

[1] C. Partridge, T. Mendez, and W. Milliken, “Host Anycasting Service,” IETF RFC 1546, Nov. 1993.

[2] S. Matsunaga, S. Ata, H. Kitamura, and M. Murata, “Applications of IPv6 Anycasting,” draft-ata-ipv6-anycast-app-00, February 2005.

[3] B. Greene and D. McPherson, “ISP Security: Deploying and Using Sinkholes,” June 2003, NANOG TALK, http://www.nanog.org/mtg-0306/sink.html.

[4] D. Katabi, “The Use of IP-Anycast for Building Efficient Multicast Trees,” *in Proc. of Global Internet’99*, Costa Rica, 1999.

[5] T. Hardy, “RFC 3258 - Distributing Authoritative Name Servers via Shared Unicast Addresses,” April 2002.

[6] I. Stoica *et al.*, “Internet Indirection Infrastructure,” *in Proc. of ACM SIGCOMM*, Aug. 2002, pp. 73–88.

[7] H. Ballani and P. Francis, “Towards a global IP anycast service,” *ACM SIGCOMM Computer Communication Review*, v.35 n.4, October 2005.

[8] M. J. Freedman, K. Lakshminarayanan, and D. Mazières, “OASIS: Anycast for Any Service,” *in Proc. of the 3rdt Symposium on Networked System Design and Implementation (*NSDI '06*),* San Jose, CA, May 2006.

[9] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A distributed anonymous information storage and retrieval system,” Freenet White Paper, http://freenetproject.org/freenet.pdf.

[10] http://gnutella.wego.com

[11] http://www.kazaa.com

[12] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, “A scalable content addressable network,” *in Proc. of ACM SIGCOMM*, 2001, pp. 161–172.

[13] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A scalable peer-to-peer lookup protocol for internet applications,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 17–32, 2003.

[14] A. Rowstron and P. Druschel, “Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems,” *in Proc. of the Middleware*, 2001.

[15] H. Ballani and P. Francis, “Understanding IP Anycast,” Cornell CIS Technical Report TR2006-2028, May 2006.