A Scalable Hybrid Overlay Multicast Adopting Host Group Model for Subnet-Dense Receivers

Dongkyun Kim1[†] and Ki-Sung Yu2[†,]

Korea Institute of Science and Technology Information, Daejeon, South Korea

Summary
Currently multimedia services for customers are rapidly being deployed (e.g. IPTV), and scalable multicasting is required for the needs of the services, e.g. robustness, security, and QoS. While native IP multicast is considered a good solution for the multimedia services, variety of overlay multicast mechanisms have been suggested to remove the barriers that block deployment of IP multicast on Internet. However, most of overlay multicast mechanisms do not consider host group model, which is not a problematic portion, but an advantageous feature of traditional IP multicast in terms of scalability, robustness, and security. In this paper, we propose a hybrid and hierarchical scheme to take advantages of both IP multicast and overlay multicast, based on host group model to gain performance efficiency for many group members on subnet dense mode. Overall network performance enhancement is shown in the performance analysis of our scheme, regarding low latency, small stress, and optimal stretch for multimedia receivers in subnet dense mode.

Key words:
Multimedia, Overly multicast, Host group model, Subnet dense mode.

1. Introduction

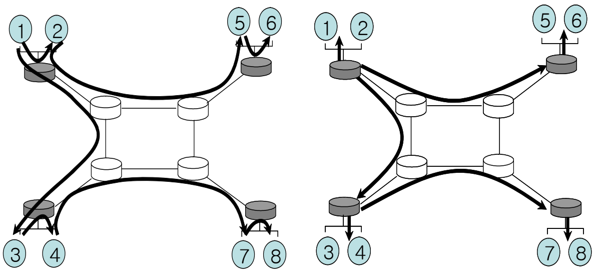
Multimedia services such as television, video, audio, text, graphics, data delivered over IP based networks need to be managed to provide the required level of QoS/QoE, security, interactivity, and reliability [2]. It means there should be a mechanism for IP group communications between multimedia receivers, guaranteeing a certain level of quality as well as gaining reliability, robustness, security, and more importantly, scalability. Traditional IP multicast has long been known as efficient data delivery mechanism in terms of network resource usage for group-oriented communication with many group participants. However, IP multicast is yet to take off on Internet since there have been many barriers against its deployment, for example inter-domain routing problem, forwarding state scalability on core networks, full router dependency, and so on [3]. Several alternatives for IP multicast have also been proposed, e.g. SSM (Source Specific Multicast) [4], XCAST (eXplicit Multicast) [5], XCAST+ (XCAST Extension) [6]. While these newly proposed schemes make improvements for efficient data delivery in some points, they are still insufficient to remove deployment hurdles, mainly due to their full router dependency as in traditional IP multicast. As another alternative, Overlay multicast has been suggested to complement those problems of IP multicast and above alternatives. In overlay multicast, a virtual infrastructure is built to form an overlay network over IP network topology, and data dissemination is achieved by packet relay in application layer, rather than network layer. Thus, there is no network scalability problem on core network. Each intermediate router needs not to keep track of specific multicast state information or exchange multicast routing table information, and protocol complexity can be drastically reduced as well. Furthermore, overlay multicast can operate well through different network domains that each of them is under single administrative control, since most overlay multicast mechanism is designed for end-host without router dependency.

In spite of above advantages, proposed various overlay multicast mechanisms based on end-hosts still need to solve some inefficiency problem in deployment and performance aspects. First, most overlay multicast architectures do not consider host group model of IP multicast, and they simply move multicast function to application layer of host from IP layer of router. However, indicated problems of IP multicast exit not in host to router operation, but in router to router operation. Even Xcast+ [6] adopts host group model of IP multicast, so that it can provide subnet-dense mode for many end-users on shared media network. Most overlay multicast mechanisms overlooked the advantages of host group model, e.g. 1-hop scalability, transparency, and robustness. Second, in case of previously proposed overlay multicast schemes, inherent performance penalty is not avoided because deployment for end-host multicast should be performed by end systems. Therefore end-host multicast is inherently less efficient and unreliable [9]. In our scheme, HOME (Host group based Overly Multicast Environment), we adopt host group model for scalable overlay multicast architecture required by multimedia receivers in subnet dense mode. Figure 1 shows how HOME generates data delivery tree, compared to ordinal overlay multicast mechanisms.

We propose HOME scheme adopting host group model to make scalable, robust, and secure overlay multicast

Manuscript received November 5, 2007
Manuscript revised November 20, 2007

architecture by moving overlay multicast functionalities to DR (Designated Router) from end-host. Thus, traditional IP multicast mechanism, e.g. IGMP (Internet Group Management Protocol) or MLD (Multicast Listener Discovery), is applied between DR and hosts, and overlay multicast mechanism is used between DRs. The advantages of this architecture are as follows: (1) Transparency is guaranteed. Existing hosts do not need any modifications for application level multicast because IP multicast scheme is used on their subnet. Hosts just need to support IP multicast protocols without considering additional addressing scheme for overlay multicast and regarding to what algorithm is used on network level. (2) By applying IP multicast mechanism between DR and hosts (host group model), performance efficiency increased in terms of 1-hop scalability, higher robustness, reduced end-to-end delay and security for multimedia receivers, particularly when the receivers are densely clustered in each subnet. Therefore, these advantages gained help our scheme, HOME, achieve a certain level of network quality required by multimedia services.



End-host based overlay multicast HOME

Fig. 1.Comparison of Data Delivery Tree

The remainder of this paper is organized as follows. Section II describes the design of our system in terms of overall architecture, group operations, and data delivery. Section III evaluates the performance, and compares our approach with other works. Finally, we conclude this paper in Section IV.

2. System Design

2.1 Overall Architecture

Previously proposed overlay multicast (or ALM) mechanisms are basically based on host relay system for all cases. Thus, each hosts need to take in overlay multicast function to apply existing overlay multicast mechanism. But, this architecture is not scalable as well as cost-efficient. For example, suppose that a host tries to join three different group sessions. One group session makes use of End system multicast [3], another session is supported by Yoid [10], and other adopts scattercast. In this case, a host has to support three different protocols. So, a host may suffer from protocol complexity, overhead to compute each path and to relay multicast data. Moreover, some members do not want to relay data packet in security aspects. If such situation happens, the multicast service cannot be naturally provided.

For above reason, it is required to design common and scalable network architecture for overlay multicast deployment without dependency in ALM algorithms. To achieve this, we propose to use hierarchical architecture using host group model. In the shared-media layer, IP multicast scheme is adopted as usual, where IGMP or MLD is used as a control protocol for data dissemination on subnet. Therefore, it is expected to need any specific requirements on hosts to enhance performance. Higher layers consist of various overlay multicast protocols between DRs across networks. In practical, it can accelerate to deploy group-oriented communication, making use of transparency, scalability, robustness and cost-efficiency.

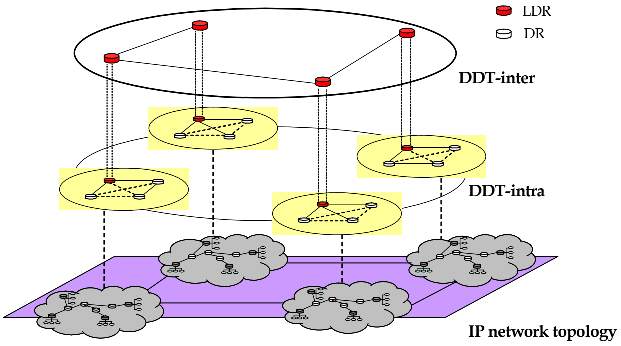


Fig. 2 Overall Architecture of HOME

Figure 2 shows a logical topology which represents a hierarchical architecture of HOME. As you can see in Fig. 2, the data delivery tree in intra-domain (DDT-intra) consists of DRs who are connected group participants to serve end hosts on its subnet. One of DRs in a data delivery tree is selected as a leader DR (LDR) to construct the inter-domain data delivery tree (DDT-inter). The hierarchical architecture allows group members to be clustered and helps them gain scalable group communications.

Figure 3 shows the instance of network architecture for group-oriented communication. IP multicast is basically used on shared-media network. Hosts use IGMP to join group, and broadcast data in its subnet area. On the other hand, overlay multicast mechanism is applied between

DRs that participate in a multicast group. Various ALM mechanisms can be applied in order to establish data transmission topologies, and build a structure to relay data using the topologies. Service for each group members will be described later in detail in this paper. To support multicast over our network architecture, functionality of DR should be extended. In short, DR plays a role of gateway as a border node between IP multicast and overlay multicast. The protocol stack and component of DR is defined as shown in Figure 4.

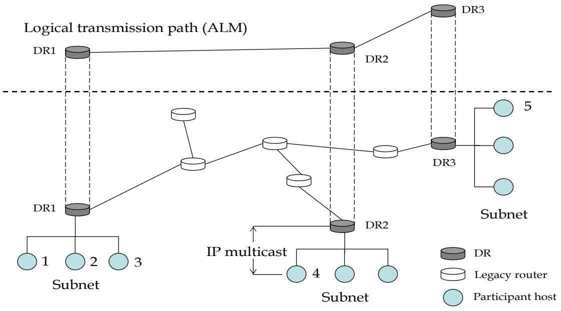


Fig. 3. Example of network architecture.

DR participating in the group is equipped with two protocols to support overlay multicast and IP multicast as shown in Figure 4. ALM protocol needs to be added between application layer and transport layer. The most basic function of ALM layer is to transmit data by constructing relay system between DRs participating in the group. In addition, ALM protocol generally has two important roles, group management using control messages and packet relay. Each mechanism in ALM layer uses control protocol to find the parent DR that transmits data to itself, and the next DR to which data should be sent. Results acquired by control protocol are stored in group forwarding table to maintain the parent and next DR information recognized by group identifier.

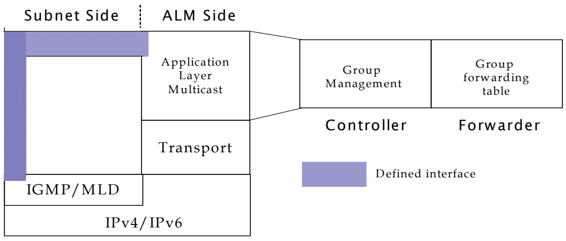


Fig. 4. Protocol stack and components of DR

In the proposed design of DR, an interface exists between ALM layer and IP layer (subnet), and another interface is newly defined between ALM and IGMP. Using the interface between ALM layer and IP layer, IP multicast datagram delivered from a subnet is transmitted directly to the overlay multicast layer without looking up general multicast routing table. To the contrary, overlay multicast layer uses the interface to transmit original IP multicast datagram extracted from ALM packets that are sent from neighbor DRs in the group, directly to IP layer at subnet side. IP datagram here includes IP header which contains a multicast group address as destination address. Also, the interface between IGMP and ALM should be newly defined for hosts to notify their join and leave to DR using IGMP. DR needs to deliver these IGMP messages to ALM when it receives new group join or leave messages. In other worlds, the interface between application layer and IP layer is designed for data transmission, while the newly defined interface between ALM and IGMP is a control interface for group management. Our mechanism is designed to provide common architecture for adopting already proposed ALM mechanisms for those steps in order to make best use of various well-defined ALM algorithms [9-14]. In addition, we suggest group operations for member join and leave, as indicated in Section B and C.

2.2 Group Join Operation

This section describes the operations of proposed mechanism for dynamic group participation. In our scheme, we assume each participant can acquire group information from CP (Core Point), a repository system to hold records of group members. CP is different from RP in traditional IP multicast in a point that it is not involved with data forwarding, but just related to building data delivery tree. Therefore, CP can operate with less overhead than RP used for IP multicast does.

When it comes to group join operation of HOME, if an LDR is present, CP replies the requesting DR with other members' addresses belong to the requested (S,G) channel and corresponding to GID using REGISTRATION REPLY (RRep). After receiving REGISTRATION REPLY from CP, the requesting DR tries to join the group using JOIN REQUEST (JReq) to the nearest DR known by IP routing information. The nearest DR sends JOIN REPLY (JRep) as a temporal parent to the joining DR. Consequently the joining DR sends the first KEEPALIVE message to its LDR, and other members are informed the address of the new DR from the LDR After this procedure, each DR computes and establishes a new DDT-intra including the new member. In practical, it is not necessary to define additional control messages for establishing new DDT

Otherwise, if there is no recorded LDR corresponding to GID, it means that it is the first requesting DR in the corresponding domain. In this case, the DR is chosen as an LDR so that the LDR is inserted as one of the member nodes on DDT-inter. In particular, the source DR is always elected as an LDR. Figure 5 elaborates the join procedures.

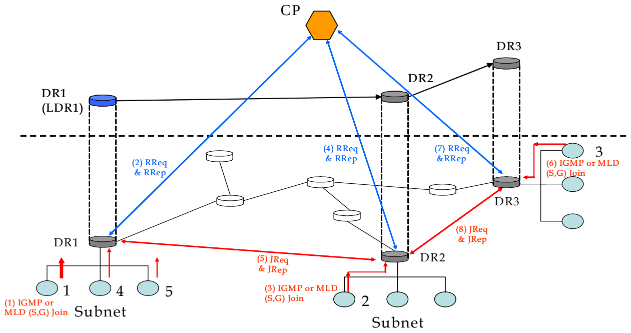


Fig. 5. Group Join Process

2.3 Group Leave Operation

In case of group leave, it is more complex than group join. When there is no more group participant on a subnet, a leaving DR requests IGMP or MLD GROUP LEAVE to its LDR or CP. The included information in GROUP LEAVE message is different from each case depending on the role of a DR, that is, the LDR and non-LDR. Figure 6 and Figure 7 show the procedures of group leave for each case.

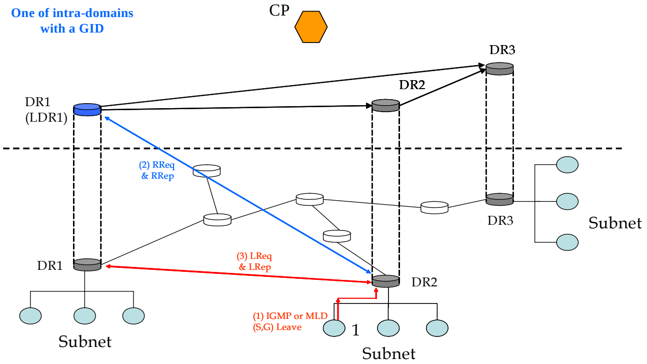


Fig. 6. Group Leave of non-LDR

If a leaving DR is not an LDR for a group in intra-domain, as we expect, the DDT-inter has no effect of the DR's leave. It is handled by rebuilding DDT-intra. The leaving DR sends and receives REGISTRATION REQUEST and REGISTRATION REPLY to and from its LDR respectively. Then, it request group leave to its parent DR using LEAVE REQUEST (LReq) with the information of its children so that the parent can build temporal links to them and deliver multicast datagram continuously. The leaving DR leaves the group after receiving LEAVE REPLY (LRep) from its LDR, and communicates KEEPALIVE REQUEST/REPLY messages (KReq/KRep) with its LDR to rebuild DDT-intra with remain members.

If a leaving DR is an LDR for a group in intra-domain, the group leave procedure is as follows. First, leaving LDR registers its group leave to CP, because CP maintains the list of LDRs for a multicast channel. Second, when registering to CP, the leaving LDR includes the information of a new candidate LDR which is the nearest DR so that CP updates its member list with the new LDR. Third, the LDR sends group leave to the nearest DR with the information of its children both for DDT-inter and DDT-intra and the nearest DR takes the role of an LDR. After this procedure, the CP should address the LDR's change to other LDRs of DDT-inter and neighboring LDRs only update their state information. In this case, the DDT-inter have the same delivery path regardless of fluctuating group membership.

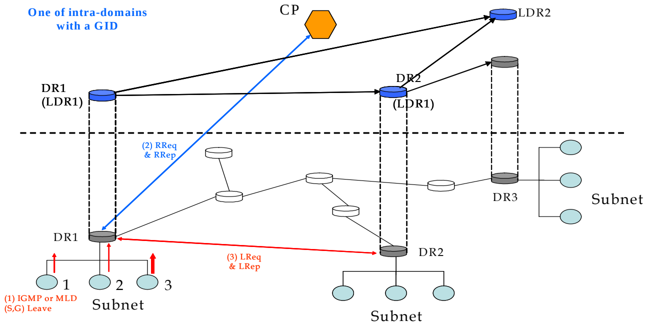


Fig. 7. Group Leave of LDR

2.4 Data Delivery

As described in section B and C, overlay transmission path is configured for data dissemination by dynamic group join and leave operations based on DRs. Data can be delivered right after the overlay path is constructed. For example, suppose that the overlay ALM path for data dissemination is built such as DR1 <-> DR2 <-> DR3 for the group G as shown in Fig 2. Suppose that group participants are host-1 to host-5 and group address is G as well. We assume data delivery from host-5 to the group G. First, host-5 sends DR3 multicast packet that includes the group address G in IP header's destination address field. DR3 can detect that the received packet is for multicast after parsing the destination field in the packet's IP header. Then, DR3 adds ALM header information to the received packet instead of processing procedure for traditional multicast forwarding. This ALM header includes the group address G and sender's source IP address. On the other hand, when DR3 receives IP datagram encapsulated with ALM header from neighbor DRs, ALM layer in DR3 parses the header to identify group address G and sender's source IP address.

To send packets to neighbor DR, ALM encapsulates the received IP datagram from shared-media network, i.e. subnet, with the new ALM header, regarding the IP datagram as payload. It also acquires next DR's address by searching into group forwarding table of forwarder in order to send unicast packet to neighbor DR. Figure 8 shows the packet encapsulation format.

DR3 sends data to DR2 using this packet encapsulation format. When DR2 receives the packet, it performs two kinds of works. First, it removes the header of the packet and sends an original multicast packet toward its subnet. Second, it relays the packet to DR1.

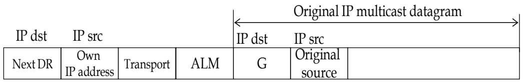


Fig. 8. Packet encapsulation format

In case of the first step, ALM acquires the group address G and sender's IP address by parsing the packet when it receives the packet. Suppose the group address is G and sender's address is DR1 in this example. ALM extracts original multicast packet and sends it through newly defined interface. IP layer that receives this packet sends its subnet the packet using existing multicast mechanism; therefore host-4 (participant in the group G) can receive this data packet.

The second procedure is to relay packets. ALM can find DR1 as the neighbor DR with the group address G and sender's address, host-5, after looking up group forwarding table. Therefore, it sends DR1 encapsulated packet of which payload is original IP multicast packet. After DR1 receives packet, it only performs data transmission for the group participants in its subnet because it has no responsibility for relaying data to any other neighbors. If host-1 wants to send data packet to group members, the opposite direction, DR1->DR-2 ->DR3, can be applied.

3. Performance Analysis

Our mechanism is adopting host group model to enhance performance in terms of optimality on some stages. We design and adjust host group model to our scheme and show performance improvement in terms of three factors, end-to-end delay, stress and stretch.

End-to-end delay means the data transmission latency from a source to a receiver. Stress means the number of duplicate packets per link or node. Stretch indicates the ratio of the path length along the overlay from the source to the member to the length of the direct unicast path. We try to compare performance between general end host based overlay multicast (EHOM) and our scheme (HOME) based on the three factors. Figure 9 describes the comparison of general end host multicast mechanisms and our scheme focused on their architectures. In Figure 9, A and C indicate subnets on the overall path S from sender i to the receiver j, M is a list of members, k is the last DR toward j in a subnet, and B indicates overlay network.

As shown in Figure 9, EHOM does not adopt host group model, therefore only does overlay network exist between end hosts. In the mean while, HOME is based on host group model, so overlay network exists between DRs so that each end host can just perform IP multicasting in its subnet. This difference means that EHOM does not have optimality at the same level of IP multicast, while HOME does in subnet. In the worst case particularly, all the nodes are dangled on tree in EHOM, however the nodes of HOME broadcast datagrams in subnet based on host group model. In this context, we can summarize our performance enhancement with three factors described as follows.

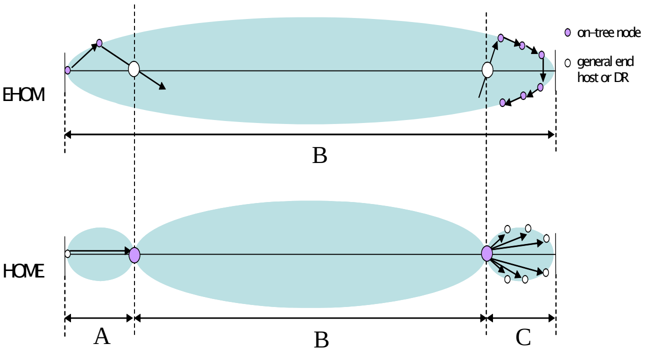


Fig. 9. Comparison of EHOM and HOME

- Delay: the delay from source i to the last DR, k in the subnet having receiver j + the delay from DR, k to receiver j
- Stress: the stress from source i to the last DR, k in the subnet having receiver j + the stress from DR, k to receiver j
- Stretch: the stretch from source i to the last DR, k in the subnet having receiver j + the stretch from DR, k to receiver j

Based on Table 1, it is found that HOME has better performance than EHOM in terms of the optimality in subnet stage. In case of EHOM, if n number of end hosts exist in a subnet, its delay and stress in a subnet are n*delay to each end host and n*stress to each end host, respectively. However the delay and stress in a subnet is always delay to one end host and stress to one end host, respectively in HOME because of host group model adaptation. In addition, stretch of EHOM is bigger than the one of HOME in a subnet since HOME's stretch is always 1. Consequently, as the number of end hosts increases in a subnet, the performance of HOME shows better results in a view that HOME has the same level of optimality as IP multicast on the subnet stage as found in Figure 9 and Table 1.

Table 1. Performance enhancement of HOME

	EHOM	HOME
Delay	$\sum_{m=i}^{k} D_{DR_m} + n \cdot D_{h_i (i=k..j)}$	$\sum_{m=i}^{k} D_{DR_m} + D_h$
Stress	$\sum_{m=i}^{k} S_{DR_m} + n \cdot S_{h_i (i=k..j)}$	$\sum_{m=i}^{k} S_{DR_m} + S_h$
Stretch	$\frac{\sum_{n=i}^{k} O_{DR_n}}{\sum_{n=i}^{k} U_{DR_n}} + \sum_{l=k}^{j} O_{h_l}$	$\frac{\sum_{n=i}^{k} O_{DR_n}}{\sum_{n=i}^{k} U_{DR_n}} + 1$

D : delay to active DR (D_{DR}) or active end host (D_h),

S : stress on the active link to DR (S_{DR}) or end host (S_h),

O : # of overlay path to DR (O_{DR}) or end host (O_h),

U : # of unicast path to DR (U_{DR}) or end host (U_h),

n: # of end-hosts in a subnet.

4. Conclusion and Future Works

This paper suggests scalable and efficient overlay multicast architecture to meet requirements of multimedia receivers in subnet dense mode, taking the advantages of overlay multicast as well as IP multicast. Our performance analysis result indicates that the hybrid approach of overlay multicast based on host group model helps construct efficient overlay multicast network in terms of low end-to-end data delivery latency, small stress and constant stretch in subnet dense mode. Although some research issues (e.g. a single point of failure of CP) still need to be followed up. We expect our scheme, HOME, can be used to deploy multicast on Internet more practically, in order for multimedia services with proper performance provided. For further study, we plan to design more concrete architecture considering other performance factors that multimedia receivers require.

References

[1] D. Kosiur, IP Multicasting : The Complete Guide to Interactive Corporate Networks, John Wiley & Sons, Inc., 1998.

[2] ITU-T FG IPTV, IPTV Service Requirements, January 2006

[3] Y. Chu et al., "A Case for End System Multicast," IEEE Journal on Selected Areas in Communication (JSAC), Special Issue on Networking Support for Multicast, 2002

[4] H. Holbrook and B.Cain, "Source-Specific Multicast IP," IETF Internet-Draft, <draft-ietf-holbrook-ssm-arch-00.txt>, 2000.

[5] R. Boivie et al., "Explicit Multicast (Xcast) Basic Specification," IETF Internet-Draft, <draft-ooms-xcast-basic-spec-02.txt>, 2001.

[6] M. Shin et al., "Explicit Multicast Extension(Xcast+) for Effcient Multicast Packet Delivery," ETRI journal, Vol. 23, No. 4, December 2001.

[7] IETF Reliable Multicast Transport (rmt) Working Group Charter, http://www.ietf.org/html.charters/rmt-charter.html.

[8] S. Deering, "Host Extensions for IP Multicasting," IETF RFC-1112, August 1989.

[9] B. Zhang et al., "Host Multicast: A Framework for Delivering Multicast To End Users," IEEE INFOCOM'02, June 2002.

[10] P. Francis, "Yoid: Extending the Internet Multicast Architecture," ACIRI Technical Report, April 2000.

[11] S. Banerjee et al., "Scalable application layer multicast," ACM SIGCOMM'02, August 2002.

[12] A. Rowstron et al., "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems," IFIP/ACM ICDCP'01, November 2001.

[13] B. Y. Zhao et al., "Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing," Technical Report, UCB/CSD-01-1141, University of California, Berkeley, CA, USA, April 2001.

[14] C. G. Plaxton and A.W.Richa, "Accessing nearby copies of replicated objects in a distributed environment," In ACM Symposium on Parallel Algorithms and Architectures, June 1997.

Dongkyun Kim received the B.S. in Computer Science and Engineering from Hannam University in 1996, and received his M.S. and Ph.D. from Chungnam National University in 1999 and 2005 respectively. During 2006-2007, he stayed in Joint Institute of Computer Science of OakRidge National Laboratory (ORNL) and University of Tennessee (UT) as a guest researcher, to perform global joint research project called GLORIAD. He is now a senior researcher, working at KISTI, South Korea.

Ki-Sung Yu received the B.S. in Computer Science and Engineering from Hannam University, and received his M.S. and Ph.D. from Sungkyunkwan University in 2004 and 2007 respectively. He is now a senior researcher and a head of research networking team at KISTI, leading a national research network project called KREONET.