Enhanced Cores Based Tree for Many-to-Many IP Multicasting

Seok-Joo Koh · Shin-Gak Kang · Ki-Shik Park

In this paper, we propose a simple and practical scheme for many-to-many IP multicasting. The proposed scheme is based on the core based tree (CBT) protocol, and designed to enhance the CBT in terms of tree cost and traffic concentration. In the scheme, each group user is simply connected to the nearest core router in the network. Each core router forwards the source traffic to the network via the pre-configured backbone core tree spanning all the core routers in the network. To ensure that the backbone core tree keeps only the core routers with active group users, the core routers that have no downstream users are removed from the backbone core tree. The experimental results show that the proposed scheme significantly improves the existing CBT scheme in terms of tree cost and traffic concentration.

I. INTRODUCTION

The IP multicast routing schemes that have been proposed so far can be classified into three categories: source-based tree[1]~[5], core-based shared tree [6]~[8], and Quality of Service (QoS) based tree [9]~[11] approaches.

The source-based tree approach induces lower end-to-end delay than the other schemes. However, the core-based shared tree approach has advantages over the source-based trees in terms of scalability. The source-based tree approach scales $O(S*G)$, where S is the number of active sources and G is the number of group members. On the other hand, the shared tree approach eliminates the source-scaling factor S since all sources share the same tree. Accordingly, CBT scales $O(G)$. The QoS-based trees provide the efficient optimization natures in terms of end-to-end delay and bandwidth utilization, which are based on the full information on overall network topology and available link capacity, and thus the QoS-based tree approaches are not likely to be widely deployed in real-world Internet in near future.

This paper focuses on the construction and maintenance of a shared tree for many-to-many multicasting, which is based on the CBT protocol. The source-based trees and QoS-based trees are beyond the scope of this paper. In particular, we propose a simple and practical scheme to overcome the drawbacks of the existing CBT protocol, which are described below.

The core based tree (CBT) protocol [6] is regarded as a promising practical solution for many-to-many multicasting, but it has suffered from the following drawbacks:

- *Core router selection* :
  It is not easy to select an optimal core router for a given group, because the group membership information is not given a priori. In CBT, a core is a randomly assigned to the group by using a hash function, without using any knowledge of group membership or distribution information. This results in high tree cost.
- *Traffic concentration at the core router* :
  In CBT, all the users send and receive the data packets via a single core, which induces severe traffic concentration at the core. Thus a larger amount of link resources and processing capacity near the core router are required.

Seok-Joo Koh, Shin-Gak Kang, Ki-Shik Park : Electronics Telecommunications Research Institute (ETRI)

Table 1. Multicast Routing Protocols

Acronym	Multicast Routing Protocols	Category
DVMRP	Distance Vector Multicast Routing Protocol	Source-based
PIM-DM	Protocol Independent Multicast-Dense Mode	Source-based
MOSPF	Multicast extensions to OSPF	Source-based
SSM	Source Specific Multicast	Source-based
CBT	Core Based Tree	Core-based
PIM-SM	Protocol Independent Multicast-Sparse Mode	Core-based
SMP	Simple Multicast Protocol	Core-based
QoSMIC	QoS sensitive Multicast Internet protocol	QoS-based
DDMC	Destination Driven Multicast	QoS-based
QDMR	QoS Dependent Multicast Routing	QoS-based

In this paper, an efficient scheme for the many-to-many IP multicasting is proposed. The proposed scheme is based on the CBT, but designed to address the drawbacks described above. For each incoming group user, the proposed scheme constructs a tree in the fashion that the user is simply connected to the nearest core router in the network. Thus, one or more core routers may be involved in the multicast tree. In the multicast data transmission, each core router forwards the multicast packets of a source to the network via a pre-configured backbone core tree spanning all the active core routers. A tree generated by the proposed scheme has a low tree cost and alleviates the traffic concentration, compared to the CBT. The experimental results show that the proposed scheme provides the tree cost saving of 20~40%. We also note that traffic concentration can be alleviated by the protocol scheme.

This paper is organized as follows. Section II reviews the existing multicast routing schemes. In Section III, we propose a new scheme to build a bi-directional shared tree and to maintain the tree, together with the detailed extensions from the CBT. In Section IV, The performance of the proposed scheme is compared with the existing CBT protocol by simulations. Section V concludes this paper.

II. PREVIOUS WORKS

Many IP multicast routing protocols and algorithms have been proposed so far. Those can be categorized into the source-based tree, core-based shared tree and Quality of Service (QoS) based tree approaches. Table 1 shows the acronyms of those protocols.

A source-based tree is a source-rooted shortest path tree from a source to all receivers, which is employed in DVMRP [1], PIM-DM [2], MOSPF [3], and SSM [4], [5]. These protocols are slightly different from each other in terms of a detailed tree construction mechanism.

DVMRP and PIM-DM are based on the so-called broadcast-and-prune mechanism, in which the source broadcast data to all the routers in the network, regardless of being active receivers or not in the downstream. The router that has no downstream receivers sends a 'prune' message to its upstream routers. These broadcast and prune procedures are repeated periodically, which is not desirable to be scalable to large networks. The only difference is that DVMRP uses its own distance vector protocol, while PIM-DM can employ any unicast protocol.

MOSPF is based on the unicast OSPF protocol. Like an OSPF, each router broadcasts the link state advertisement (LSA) messages into the network. The LSA may contain information on link cost, hop distance and available link capacity. The LSA is used for each router to identify the overall network topology. In MOSPF, additional 'group membership LSA' messages are delivered together with LSA messages. Based on the network topology and group membership information, the source calculates the shortest path tree spanning all the active receivers. The broadcasting of LSA and group-membership LSA in the network induces a large amount of control traffic, which makes MOSPF difficult to deploy in large networks.

SSM is the most recently proposed protocol. The SSM uses an explicit join mechanism like PIM-SM and CBT. In the explicit join mechanism, each receiver sends a join message to the source by using unicast routing protocol, during which a shortest path tree is constructed. It is agreed that SSM is a simple and efficient protocol to construct a source-based tree for one-to-many multicasting.

A core based shared tree is a single delivery tree that is shared by all users of a group. All active senders in a group share a common tree for many-to-many multicasting. Typically a shared tree is constructed by using a core or rendezvous. This approach is employed in CBT [6], PIM-SM [7], and SMP [8].

In CBT and PIM-SM, a multicast tree is built by choosing a suitable core router for a group. For selection of the core router, the so-called 'bootstrap mechanism' is used. In the network or domain, there is a bootstrap router (BSR) that informs all the CBT or PIM-SM routers in the network of the list of candidate core routers by broadcast. To do this, each candidate core router must send a periodic 'keep-alive' message to the BSR. Each router, which has a receiver for a group G, selects a core router by using hash function that maps a group address G to one of candidate core routers. Then the CBT or PIM-SM router joins the tree by sending special join messages toward the core. The routers along the path keep state about which ports are in the tree. The result is a tree of the shortest paths from the center to all members.

The only difference between PIM-SM and CBT is that the forwarding state created by PIM-SM is unidirectional in that it only allows traffic to flow away from the core, not toward it. The CBT however builds a bi-directional tree.

SMP was proposed to improve the complexity of the bootstrap mechanism in CBT and PIM-SM. The basic idea in SMP is that a multicast group is identified with a pair of a pre-designated core router and a multicast address. Thus the bootstrap mechanism is not required. However, SMP requires an IP packet header to contain the IP address of the core router as well as group destination address. This induces an additional extension IP header, which has not been fully agreed yet.

The third category for the multicast routing scheme is the QoS based tree. In this approach, the QoS metrics such as delay from the source and the required bandwidth for the applications are considered in the tree construction. The QoS-based tree construction schemes include QoSMIC [9], DDMC [10], and QDMR [11].

QoSMIC constructs a tree under the constraint of the required bandwidth. Each new receiver is connected to the closest branch of the existing tree, by using only the links with available link capacity. This requires each receiver or router to collect the information on the overall network topology including the available link capacity.

DDMC was proposed to improve total tree cost of the source-based tree. To do this, the Steiner spanning tree is calculated instead of the shortest path tree. Note that the Steiner tree problem is known as NP-complete. They proposed a fast heuristic to obtain a Steiner tree spanning a given set of a source and receivers.

QDMR generates a delay-constrained low cost tree. In the scheme, the delay requirement of the application is considered as a constraint. Under the delay constraint, a feasible tree is constructed such the tree cost is minimized.

It is noted that the QoS-based tree approaches provide an optimized tree in terms of tree cost, available link capacity and the delay. However, they require each router in the network to collect the information on overall network topology, including available link capacity and link transmission delay, and the current on-tree routers.

We also note that most of the QoS-based trees consider the delay characteristics as a metric to build the tree. Delay requirement is source-specific, and thus each source may require a different delay bound. Therefore, the delay-based tree approach is not suitable to build a shared tree for many-to-many multicasting.

III. THE PROPOSED SCHEME

In this section, we describe the proposed scheme to build a shared tree based on multiple cores. With the algorithmic sketch, the detailed extensions from the CBT protocol are given.

1. Backbone Core Tree

In the bootstrap mechanism of CBT, each core router informs the bootstrap router (BSR) that it keeps alive, and the BSR advertises the list of active core routers to the multicast routers in the network. Note that BSR as well as core routers are pre-configured by the network administrator.

In the proposed scheme, a further assumption is made that all the active core routers are organized into a backbone core tree (BCT) by the network administrator. The BCT is a bi-directional tree connecting all the core routers in the network. Each core router forwards the multicast data stream of a source to the network via the BCT, if some of the other core routers on BCT have requested the data forwarding, which will be described in the next section.

The configuration of the BCT is relatively easy, since the number of core routers is smaller than the total number of routers in the network. The network considered in this paper typically represented as an Autonomous System (AS), and the core routers in the network will be under

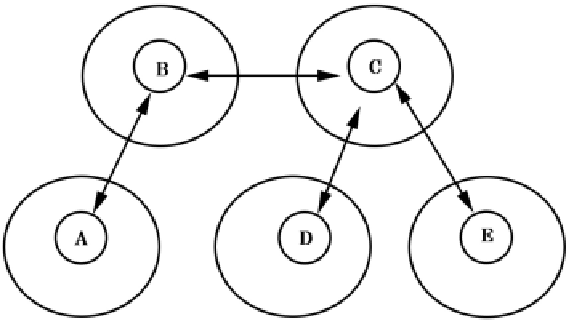

Figure 1. An Example Backbone Core Tree in the Network

control of the administrator. As shown in most of real networks, a network may consist of several sub-regions. In this case, each region may have a core router, which is connected to its upstream core router in BCT. Figure 1 illustrates an example BCT in the network. In the figure each node represents a core router in the region.

The configuration of BCT may depend on several factors including network environments or business strategy of Internet Service Providers (ISP). For an example, a campus network that has its own AS number will usually consist of only one or two regions, where the BCT configuration is relatively easy. On the other hand, a large-scale commercial ISP network will require rather a complicated configuration of BCT with many regions. The configuration of BCT will also be impacted on the location of Gateway routers in the network, which typically provide external interfaces with the other networks. In this case, it is expected that a BCT is configured in the fashion that the core routers in the regions are connected to the Gateway routers. Depending on the locations of Gateway and core routers, it may be desirable to configure a BCT into a star topology, in which one or more Gateway routers will act as a center or a root.

We note that the transmission links between core routers on the BCT have relatively high link capacity, compared to the other network links, so as to encompass the control traffic such as join and prune messages within the BCT routers. The multicast routing and forwarding between core routers on the BCT will benefit from information given by this control traffic.

2. Tree Building and Maintenance Algorithms

In this section, we assume that a BCT has been configured in the network by considering several design factors described in the previous section. Given a BCT, the multicast tree building and maintenance mechanisms are presented for each multicast group.

2.1. Join to a Core Router

Given a BCT in the network, the group join procedure is relatively simple. Each user that wants to participate in a group G just sends a JOIN(G) message to the nearest core router. Differently from the CBT, the core selection process is not performed.

Like the CBT, the bi-directional forwarding states for the given group are established on the routers over the path from the user to the core router

The join algorithm can simply be summarized as follows: *If a user in a subnet wishes to join a group address G, Then the subnet router sends a join message for group G, JOIN(G), to the nearest core router by using any unicast routing protocol.*

2.2. Data Transmission to a Group

Data transmission to a group by a source is done in a similar way as the join mechanism. A source user who wants to transmit data to a group G just sends the multicast packets to the nearest core router.

2.3. Tree Building for a Group

In Section 2.1, when a core router receives a JOIN(G) message from a user, it broadcasts the JOIN(G) message to all the other core routers along the pre-configured BCT tree. When a core router on BCT receive a JOIN(G) message form the other core routers, it creates and maintains the multicast forwarding state on G. Based on the forwarding state information, the core router will determine the forwarding interfaces (or neighboring core routers) for the multicast data of the group G, when it receives multicast data from a source.

For example, in Figure 1, suppose the core router A receives a JOIN(G) message from a user. The core A broadcasts JOIN(G) to the other cores B, C, D, and E over the BCT. Thus, the other core routers will realize that the core A is involved in the group G. If a source transmits data for a group G by way of the core router C, the data will be delivered to the core A along the forwarding state established on the core routers C and B.

This tree building mechanism ensures that the core routers without group users are involved in the multicast tree on the group. In the example described above, if the

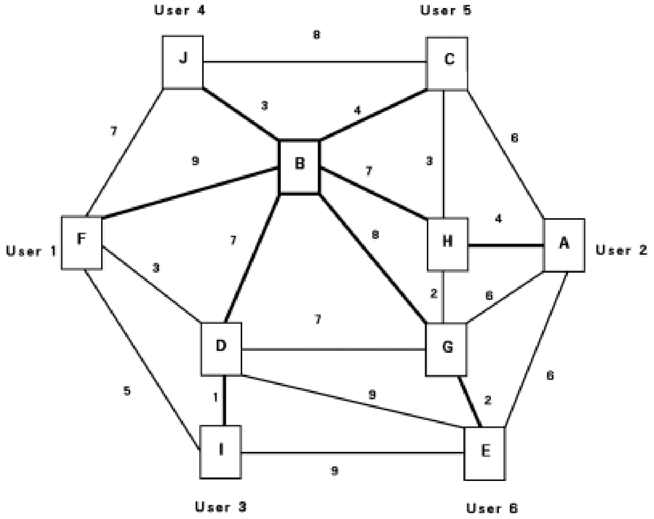

Figure 2. Core Based Tree

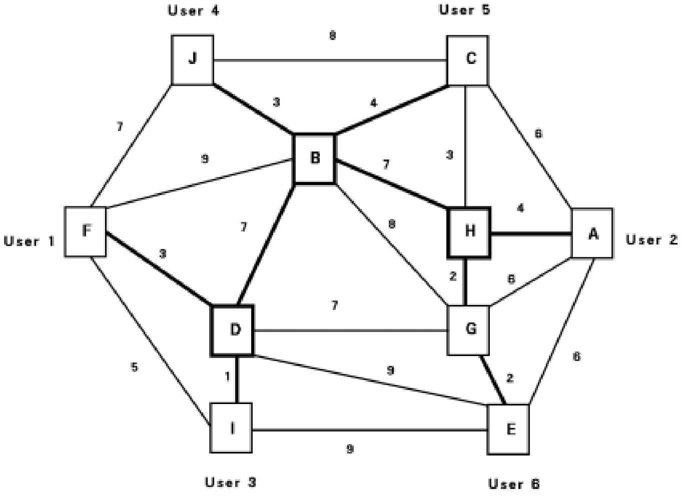

Figure 3. Proposed Scheme

cores D and E have no users for the group G, the core router C including a source will not forward the source data to the core routers D and E.

As another example, if all the group users including sources and receivers are connected to only a core A, then the data will not be delivered to the other cores since they have not sent any join request to the core A. Note that if all the users are connected to single core router, the result tree is the same as that of the CBT. That is, the CBT can be viewed as a special case of the proposed scheme.

2.4. Tree Maintenance for a Group

Each core router distributes the JOIN(G) messages to the other BCT core routers every a periodic time interval, only if the core router still has the users on a group G. Thus, the state information in each core router on BCT will be refreshed every time period. That is, if any forwarding request (or JOIN(G)) does not arrive from the other core routers until the timer expires, the forwarding state on G will be removed from the core router. This mechanism ensures that the core routers without group users are automatically pruned off the tree. In particular, this is very effective for the multicast sessions with a very short duration time.

With the timer-based refreshment mechanism, an explicit PRUNE (G) message can be used so as to minimize the latency of a session leave. If a leaf core router realizes that it has no users for a group, it prunes itself off from the BCT by sending a 'prune' message to the other core routers. For example, in Figure 1, if the core E does not have any attached user for a group, then it sends a prune message to its upstream core C for the group. The core router C then stops forwarding the data to the core E. If the core C as well as the core D has no downstream users, then the core C will send a prune message to the core A. This can be done to reduce the leave latency for the session leave. If a new user arrives at a core router after the pruning process, the core router will again send a JOIN(G) message to the other core routers on BCD, as done in the tree creation process described in Section 2.3.

3. Comparison with the CBT Protocol

The proposed scheme is different from the existing CBT protocol in that the core router is not pre-determined for a group. Instead, the nearest core router is chosen for an incoming user. This feature provides the following advantages over the CBT protocol:

1) Each router in the network does not need to run the hash function, which is used to map a group address to a unique core router. In CBT, the selection of a core router by the hash function does not consider any information on the distribution of the group receivers. This tends to generate a high tree cost.
2) The existing CBT protocol incurs traffic concentration near the core, since all the users join and exchange the data via a single core router. In the proposed scheme, the receivers are separately assigned to different core routers, and thus the traffic is dispersed.

Figure 2 and Figure 3 give an illustrative example of the differences described above. In the figures the bold-lined nodes represent a core router, and the others do the multicast routers. The number on the link is the link cost.

Figure 2 illustrates a tree obtained by CBT. In the figure, we assume that node B is selected as the core for the group by using the hash function. User 1 first join the group via node F and the tree is configured between two nodes, F and B. As a similar way, Users 2~6 also send a join message toward the core node B. The result is the core based tree, as shown in the figure. In the figure, the tree cost is 45 and the core node B has the maximum number of tree branches of six.

Figure 3 illustrates a tree obtained by the proposed scheme. We assume that there exist three candidate core nodes B, D and H in the network. In the example, the backbone core tree consists of the two links (B, D) and (B, H). By the proposed scheme, User 1 and User3 are connected to the nearest core node D. Similarly, User 2 and User6 join the core node H, and User 4 and User5 are attached to the core node B. As a result, we see that total tree cost is 33, and the maximum tree branch is four, which is at core node B.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of those algorithms, we employ two kinds of test networks. The first is a real network topology in the MBONE networks[12]. We eliminate routers with only one incident link, since such routers do not affect routing. The final graph has 32 routers, 80 links, and average degree of 2.5. The second is randomly generated networks with 100 nodes. To generate such networks, we employ the Georgia Technology Internetwork Topology Models (GT-IMT) software [13].

The link cost is randomly assigned as an integer number ranged from 1 to 10. The nodes with reasonably many node degrees are selected as core nodes, and the backbone core tree is established by using the shortest path tree among the selected core nodes. We run every experiment 100 times, and the results are averaged.

To compare the performance of test algorithms, we measured the following two metrics; total tree costs, which is the sum of the link costs of the links on the tree, and the maximum number of tree branches at the tree nodes, which represents the degree of traffic concentration.

Figure 4 shows the performance of the proposed and CBT schemes in terms of total tree cost in MBONE networks with 32 routers. The proposed scheme was tested for different number of core nodes, as indicated by C = 2, C = 3, C = 4 and C = 5. In the figure, it is shown that the proposed scheme significantly improves the tree cost, compared to the CBT. In the figure, two schemes provide the same tree costs for the smaller number of users being less than five, since the backbone core tree is maintained by using the pruning mechanism as described in Section III.

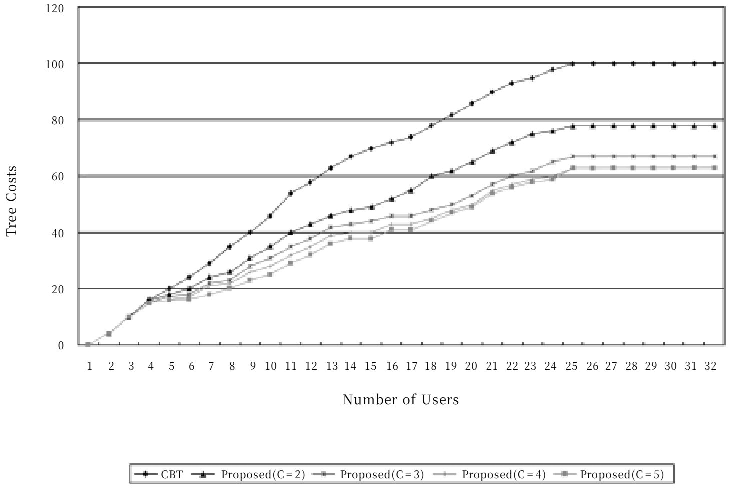

Figure 4. Comparison of Tree Costs

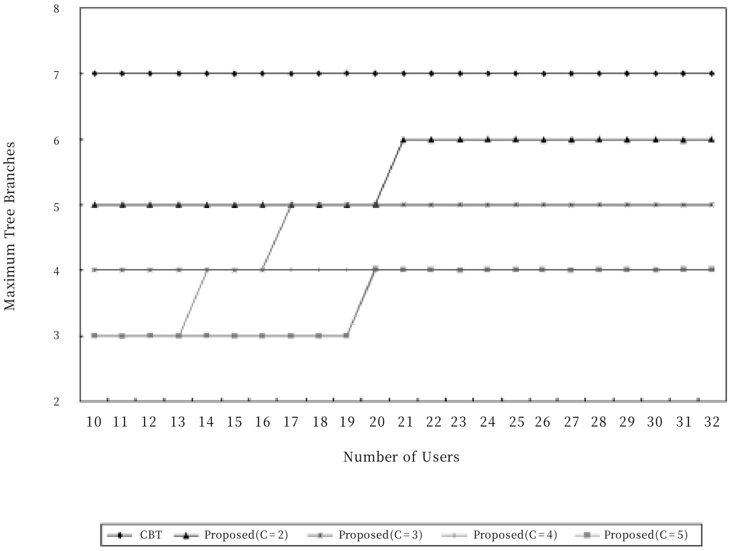

Figure 5. Comparison of Traffic Concentration

Figure 5 shows the performance of the proposed and CBT schemes in terms of traffic concentration at the core routers in MBONE networks with 32 routers. From the figure, we see that the traffic concentration at the core router in CBT is alleviated in the proposed scheme. The maximum number of tree branches is reduced from seven to four. We note that for C = 4 and C = 5 the nearly same performance is shown in terms of tree costs and traffic concentration. The performance gap between the proposed and CBT scheme becomes lower, even though the number of core nodes gets larger. From those results, we can guess that there exists a bound for the number of core routers employed to reduce the tree costs. In this example, the C = 4 seems a suitable choice, but the exact number for such a bound may depends on the network topology and the distribution of group users.

Table 2 shows the experimental results for the randomly generated problems with 100 nodes. The simulation results are very similar to those for the MBONE networks. As the number of session users increases, the performance (including the trend and slope of the lines) is nearly the same as that sown in Figure 4

Table 2. Performance of the Proposed Scheme in Random Networks with 100 Nodes

	CBT	Proposed			
		C = 2	C = 5	C = 7	C = 10
Tree Cost	100	81	72	66	64
Maximum Tree Branches	15	11	9	8	8

and Figure 5. Table 2 gives a summary of the simulation results for the networks with 100 nodes and 100 session users.

In the table, the tree cost in CBT is represented as 100. From the table, it is clear that the proposed scheme provides better performance in terms of tree costs and traffic concentration. Compared to the CBT, the proposed scheme provides the tree cost saving of approximately 20~40%, depending on the number of core nodes in the networks. In terms of traffic concentration, the proposed scheme decreases the maximum number of tree branches in the CBT by half.

V. CONCLUSION AND FURTHER STUDY

This paper proposes an enhancement of the CBT protocol for many-to-many IP multicasting. In the proposed scheme, each user is simply connected to the nearest core router. The core router will forward the multicast packets to the network via the pre-configured backbone core tree. The proposed scheme overcomes the drawbacks of the CBT protocol: high tree cost and traffic concentration. By experiments, it is shown that the proposed scheme provide s the tree cost saving of 20~40%. We also see that traffic concentration can be alleviated by the protocol scheme. The proposed scheme is based on the simple extensions from the CBT protocol, and can easily be deployed in the real Internet.

This study has focused on the performance evaluation on the tree cost and the traffic concentration at core routers. We note that the delay factor is another important metric for performance comparison, which will need some suitable integration of the tree cost and delay metrics. In future study, such an extensional research needs to be made.

[REFERENCES]

[1] D. Waitzman, S. Deering and C. Partridge, "Distance Vector Multicast Routing Protocol," RFC1075, Nov. 1988.

[2] S. Deering, D. Estrin, D. Farinacci, V. Jacobson, A Helmy, D. Meyer, L. Wei, "Protocol Independent Multicast Dense Mode Specification," Internet-Draft: draft-ietf-pim-v2-dm-01.txt, Nov. 1998.

[3] J. Moy, "Multicast Extensions to OSPF," RFC1584, Mar. 1994.

[4] H. Holbrook and B. Cain, "Source-Specific Multicast for IP," Internet-Draft: draft-holbrook-ssm-00.txt, Mar. 2000.

[5] H. Sandick and B. Cain, "PIM-SM Rules for Support of Single-Source Multicast," Internet-Draft: draft-sandick-pimsm-ssmrules-00.txt, Mar. 2000.

[6] A. J. Ballardie, "Core Based Trees Multicast Routing Architecture," RFC2201, Sept. 1997.

[7] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification," RFC2362, June 1998.

[8] R. Perlman, C. Lee, A. Ballardie, J. Crowcroft, Z. Wang and T. Maufer, "Simple Multicast: A Design for Simple, Low-Overhead Multicast," Internet-Draft: draft-perlman-simple-multicast-01.txt, Nov. 1998.

[9] M. Faloutsis, A. Banerjea and R. Pankaj, "QoSMIC: Quality of Service sensitive Multicast Internet protoCol," ACM SIGCOMM '98, Sept. 1998.

[10] A. Shaikh, et al., "Destination driven routing for low cost multicast," IEEE JSAC Vol. 15, No. 3, Apr. 1997.

[11] L. Guo, et al., "QDMR: an efficient QoS dependent multicast routing algorithm," Proceeding of 5th IEEE real-time technology and application symposium (RTAS '99), Vancouver, Canada, June 1999.

[12] S. Casner, "Major MBONE routers and links," Available from ftp.isi.edu/MBONE, 1994.

[13] E. W. Zegura, K. Calvert and S Bhattacharjee, "How to Model an Internetwork," Proceedings of IEEE Infocom '96, San Francisco, CA, 1996.

Seok-Joo Koh

He received B.S. and M.S. degrees in Management Science from KAIST in 1992 and 1994 respectively. He also received Ph.D. degree in Industrial Engineering from KAIST in 1998. Since 1998, he has been a senior researcher at Protocol Engineering Center in ETRI. His research interests include Internet Multicasting, Design and Planning of Optical Internet.

E-mail:sjkoh@pec.etri.re.kr
Tel:+82-42-860-6218
Fax:+82-42-861-5404

Shin-Gak Kang

He received B.S, M.S, and Ph. D. degrees in Electronics Engineering from Chungnam National University, Daejeon, Korea in 1984, 1987 and 1998 respectively. He joined ETRI in 1984 and is a team leader of Communication Protocol Standardization Team in Protocol Engineering Center (PEC) in ETRI. Current interest research fields include Voice over IP and Reliable Multicast.

E-mail:sgkang@etri.re.kr
Tel:+82-42-860-6117
Fax:+82-42-861-5404

Ki-Shik Park

Dr. Ki-Shik PARK was educated at the Seoul National University in the Rep. of Korea, where he obtained a first class honours degree of B.A. in 1982 in English linguistics & literature. And he got M.A. and Ph.D. Degree in the field of Telecommunications Policy in 1984 and in 1995 respectively.

He joined ETRI in 1984 and he is currently working as Director of Protocol Engineering Center. He has more than 16 year research experience in various fields of ETRI including Standardization, Telecommunication Systems Division, Information & Telecommunications Technology Division, etc. His major research interests include strategic planning for technology development, telecommunication policy including legal and institutional aspects, telecom standardization, Intellectual Property Rights, Internet applications, information systems, etc.

At present, Dr. PARK is a Vice-chairman of ITU-T TSAG (Telecommunications Standardization Advisory Group) and also Chairman of WP3/TSAG. He is also a Member of Advisory Board of ASTAP(APT STAndardization Program) regionally, and Vice-Chairman of Technical Assembly of TTA (Telecommunications Technology Association of Korea) domestically. His research interests include Telecommunications Policy Science, Managpement Information System, Wireless Internet, Technical Regulations and Telecommunication Standardizations System & Policy.

E-mail : kipark@etri.re.kr
Tel:+82-42-860-6041
Fax:+82-42-861-5404