Making QoS Aware Multicast Scalable in Terms of Link State Advertisement

Toshihiko Kato†*, Seiji Ueno‡, Shigeki Mukaiyama‡ and Kenji Suzuki†
†: KDDI R&D Laboratories, Inc. ‡: University of Electro-Communications
*: to-kato@kddilabs.jp

Abstract

Recently routing protocols for QoS aware multicast are actively studied, but there are few studies focusing on the scalability of link state advertisement when the available bandwidth of a link is updated along with the QoS aware multicast tree construction. This paper proposes a new QoS aware multicast routing protocol that is scalable in terms of the link state advertisement exchange. Our protocol has the following features; (1) A multicast network is divided into domains, and the advertisement of information on links within a domain is limited within the domain. (2) Among the border multicast routers, only the link state information of inter-domain links is advertised. As a result, the number of link state advertisement messages will be drastically reduced. (3) When a multicast tree spreads over multiple domains, the tree construction needs to be performed without information on links in other domains, and it is possible that the construction may fail. In order to cope with this problem, the crank back mechanism of a tree construction is introduced. This paper describes the detailed procedures and the message formats of our protocol. It also describes the evaluation of the number of exchanged link state advertisement messages and shows that our protocol can reduce the number by the order of 1/(number of domains) comparing with the conventional protocols.

1. Introduction

Resulting from the wide spread of audio and video data delivery over the Internet, the IP multicast technologies have become important. However, the current version of IP multicast function is based on the best effort, and therefore, it is expected that the bandwidth reservation is necessary to support QoS (Quality of Service) for individual multicast traffic.

Recently, there are active researches on QoS aware multicast [1-5]. The approach adopted by those researches consists of the following procedures.

- *QoS aware multicast tree construction and resource reservation*: At the handling of a message initiating a specific multicast traffic, such as a Join message of PIM-SM (Protocol Independent Multicast – Sparse Mode) [6], a QoS aware multicast tree is constructed using links which have enough bandwidth for this traffic. At the same time, the resource will be reserved on individual links along the route.
- *QoS aware forwarding*: Multicast datagrams of a multicast group will be forwarded along with the QoS aware multicast tree for this traffic.
- *Link state advertisement and QoS routing information update*: The multicast routers maintain routing information including network topology and available bandwidth of links. If any QoS aware multicast tree is constructed and some amount of bandwidth is reserved, the update information on individual link is advertised to all other routers.

The researches so far [1-5] are intensively studying on the first two procedures, i.e. the multicast tree construction and forwarding. For example, the research described in [1] is based on PIM-SM, and the QoS aware multicast tree is constructed when a PIM-SM Join message changing multicast tree from *RPT* (*Rendezvous Point Tree*) to *SPT* (*Shortest Path Tree*) is transferred. The researches described in [2] and [3] are based on RSVP [7] and propose a multicast protocol based on RSVP like signaling. The research in [5], on the other hand, is focusing on the inter-domain multicasting where the worldwide multicast network is divided into domains and is extending BGMP (Border Gateway Multicast Protocol) [8] for QoS support.

However, these researches do not address the scalability of link state advertisement. When one QoS aware multicast tree is constructed, the available bandwidth of links comprising this tree is changed and the link state information needs to be advertised. Therefore, the link state advertisement will be done much more frequently than the point-to-point normal routing protocol. Moreover, if the network structure is flat for multicast, the link state advertisement will be exchanged to all multicast routers all over the world. The researches described in [1], [2] and [3] mention that an example of link state advertisement protocol is MOSPF (Multicast Extension to OSPF) [9]. This means that large number of messages for link state advertisement will be exchanged over the world every time a new QoS aware multicast tree is constructed. Although the research in [5] is focusing on multi domain network, there are no considerations on how the number of messages for link state advertisement can be reduced.

In this paper, we proposes a new QoS aware multicast routing protocol which is scalable in terms of the exchange of link state advertisement messages. In our approach, a multicast network is divided into domains.

1092-1658/01 $17.00 © 2001 IEEE

The advertisement of information on links within a domain is limited within that domain. Among the border multicast routers, only the link state information of inter-domain links is advertised. As a result, the number of link state advertisement messages will be drastically reduced. However, when a multicast tree spread over multiple domains, the tree construction needs to be performed without information on links in other domains, and it is possible that the construction may fail. Therefore, we introduce the crank back mechanism in the forwarding of a tree construction message. This paper also describes the evaluation of our approach in terms of the reduction of number of necessary link state advertisement messages.

In the rest of the paper, sections 2 and 3 describe the design principles and the detailed design of our QoS aware multicast routing protocol, respectively. Section 4 describes the evaluation of our protocol.

2. Design Principles

We have adopted the following principles in order to design our QoS aware multicast routing protocol.

(1) The worldwide multicast network is divided into domains and the hierarchy of *intra-domain* and *inter-domain* is introduced. The range of link state advertisement exchange is limited for intra-domain and inter-domain levels.

(2) For the intra-domain multicasting, we use PIM-SM as an intra-domain multicast routing protocol, and OSPF [10] for the purpose of intra-domain link state advertisement, respectively, by giving a QoS extension to both of them. Similarly with [1], we use SPT (Shortest Path Tree) as a QoS aware multicast tree, and therefore, we handle the first Join message constructing an SPT as a *QoS Join message*. When an SPT is newly constructed, the required bandwidth of links comprising this SPT is reserved and the link state information such as updated available bandwidth is advertised. *The range of link state advertisement is limited within one domain.*

(3) For the inter-domain multicasting, we use BGMP as an inter-domain multicast routing protocol, and BGP-4 [11] for the inter-domain link state advertisement, respectively, with a QoS extension. Similarly with (2), an SPT is used as a QoS aware multicast tree and the first Join message creating an SPT is handed as a *QoS Join message. Among the border multicast routers, only the link state information of inter-domain links is advertised.*

(4) When an SPT (QoS aware multicast tree) spreads over multiple domains, in response to a QoS Join message, multicast routers of one domain try to construct it using only the information on available bandwidth of links within this domain and the inter-domain links, and forward the QoS Join message to the next domain. The routers in the next domain repeat the similar procedures.

(5) It is possible that, in these procedures, the SPT construction fails due to the bandwidth limitation of intra-domain links in some domain. In order to cope with such a situation, we introduce a procedure in which a failure of tree construction is reported back to the original domain of QoS Join message and another trial of SPT construction is performed through different domains. We call this procedure the *crank back mechanism.* This mechanism is done by introducing a *Join NACK message* to PIM-SM and BGMP.

3. Detailed Design

In this section, we describe the procedure and message format of our multicast routing protocol in details. We use the network configuration shown in Fig. 1 as an example. In this network configuration, a sender of a multicast group is located in domain *Ds* and domain *Droot* works as the root domain for this multicast group, according to BGMP. There is a receiver (a group member) for this multicast traffic in domain *Dr*. In the beginning, the multicast traffic is conveyed through

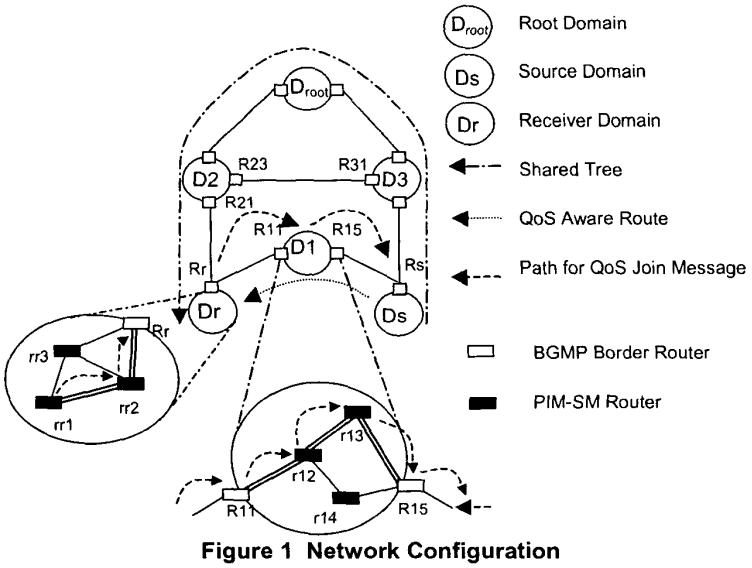


Figure 1 Network Configuration

Table 1 Example of Available Bandwidth Information in PIM-SM Routers in Domain *Dr* (M bit/s)

		From			
		rr1	rr2	rr3	Rr
To	rr1	-	10	1	-
	rr2	10	-	5	20
	rr3	1	10	-	6
	Rr	-	8	2	-

BGMP Shared Tree whose root is domain *Droot*, i.e., via domains *Ds*, *D3*, *Droot*, *D2*, and *Dr*. In each domain, multicast traffic is transferred according to PIM-SM. For example, in domain *Dr*, router *Rr*, which is a BGMP border router, works as the sender and the rendezvous point, and the multicast traffic is transferred via routers *Rr*, *rr2* and *rr1*. In the rest of this section, we describe the protocol procedure for QoS aware tree construction and that for crank back mechanism.

3.1 Protocol Procedure for QoS Aware Tree Construction

(1) When a group member requires the QoS guarantee for multicast traffic from domain *Ds*, it requires the bandwidth to router *rr1*. The PIM-SM routers in domain *Dr* maintain the bandwidth information of links between them as shown in Table 1. Since the sender of multicast traffic in this domain is BGMP border router *Rr*, router *rr1* determines the route from *Rr* to *rr1* according to this information. In this case, we assume that the route is via *Rr*, *rr2* and *rr1*. Figure 2 shows the communication sequence in this case. Router *rr1* sends a QoS Join message to *rr2* and *rr2* relays it to *Rr*. When router *rr2* receives a QoS Join message, it reserves the required bandwidth, and it sends Link State Update messages reporting the updated available bandwidth to neighbor multicast routers, *rr1*, *rr3* and *Rr*. This is done according to OSPF. When receiving these messages, multicast routers update the bandwidth information of links. Similarly, when router *Rr* receives a relayed QoS Join message, it sends Link State Update messages to neighbor routers. It needs to be mentioned that the Link State Update messages are only exchanged among multicast routers in domain *Dr*, not multicast routers in the whole network.

(2) A QoS Join message exchanged in a domain is defined by extending Join message in PIM-SM. Figure 3 shows its format. A shadowed part is a modified part. We assign a new type value for QoS Join message and add a new parameter, *Required QoS Link Information* containing information such as required bandwidth, at the end of the message. In order to advertise updated available bandwidth, we use OSPF opaque LSA option [12]. Figure 4 shows the format of Link State Update message for bandwidth advertisement. A shadowed part indicates newly defined parameter for reporting updated available bandwidth.

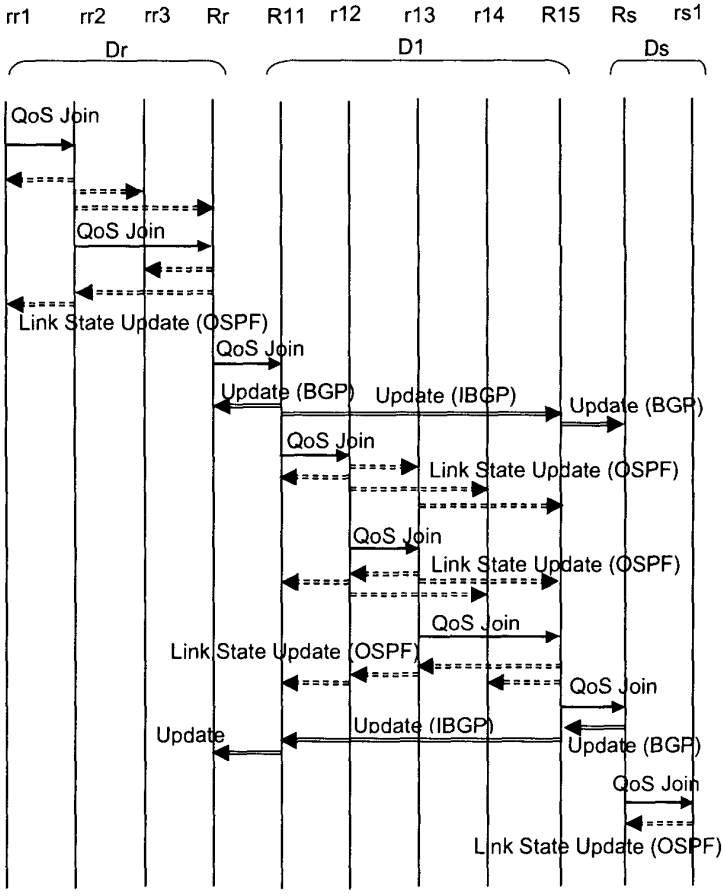


Figure 2 Communication Sequence of QoS Aware Multicast Tree

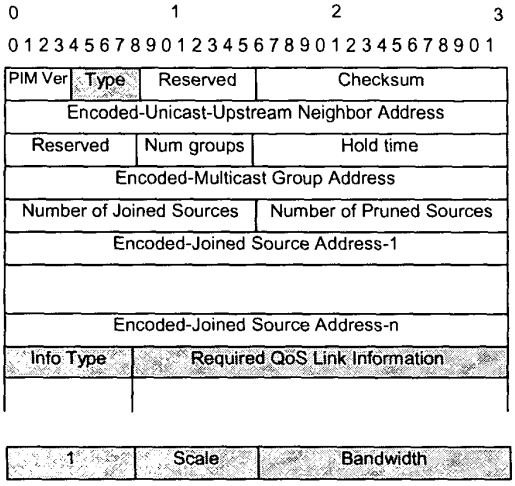


Figure 3 Format of QoS Join Message for PIM-SM

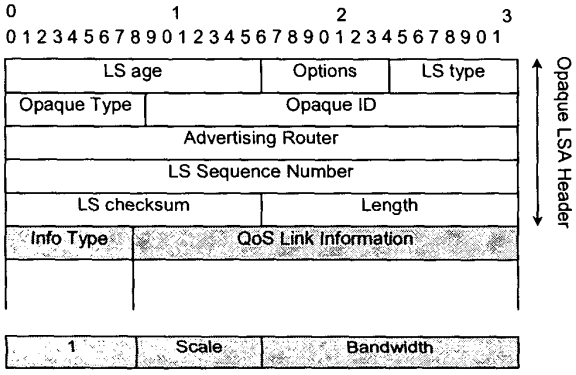


Figure 4 Format of Link State Update for Bandwidth Report

(3) When BGMP border router *Rr* receives a QoS Join message, it tries to relay it to one of the adjacent domains. The BGMP border routers maintain information on the available bandwidth through inter-domain links and the number of hops to the destination domain. Table 2 shows an example of such information for router *Rr*. Based on this information, *Rr* decides to forward the received QoS Join message to domain *D1*, i.e. to border router *R11*. When router *R11* receives a QoS Join message, it reserves the required bandwidth over the inter-domain link from *R11* to *Rr*, and sends a BGP Update message to BGMP border routers. In Fig. 2, a BGP Update message is sent to *Rr* and an IBGP Update message is sent to *R15*, which is another border router in domain *D1*. Then *R15* sends a BGP Update message to border router *Rs*. It needs to be

Table 2 Example of Available Bandwidth Information in BGMP Border Router *Rr*

Destination Domain	Next Hop Domain	Available Bandwidth (M bit/s)	Hops
Ds	D1	20	2
Ds	D2	10	3

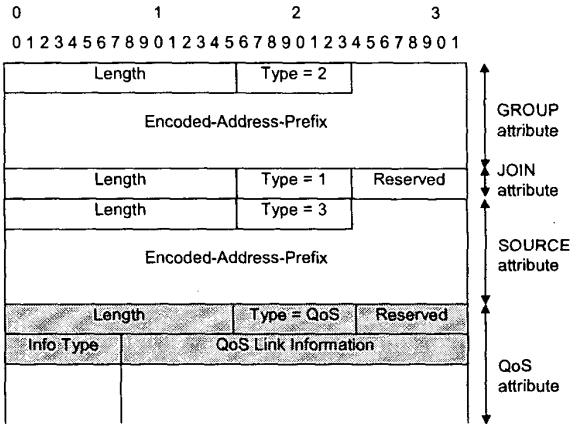


Figure 5 Format of QoS Join Message for BGMP

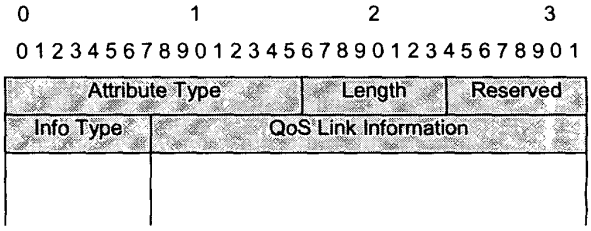


Figure 6 Format of QoS Attribute for BGP Update Message

noted that although Fig. 2 shows the communication sequence among border routers related in the focused QoS aware multicast tree, BGP Update messages are sent to all of border routers in the whole network. It also needs to be noted that even if Update messages are sent to all border routers, the number of exchanged link state advertisement messages is much smaller than the case that such messages are flooded to all routers in the whole network.

(4) A QoS Join message exchanged between BGMP border routers is defined by extending BGMP Join message. Figure 5 shows its format, where a shadowed part is a modified part. We introduce a new attribute, *QoS Attribute*, indicating required bandwidth. In order to advertise updated available bandwidth, we extend BGP Update message as

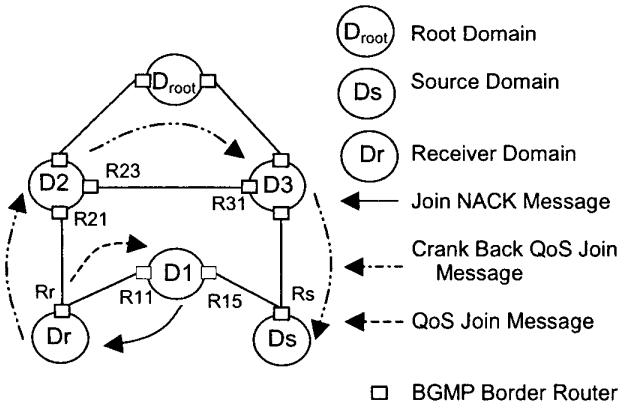


Figure 7 Crank Back Mechanism

shown in Fig. 6. A new attribute, *QoS Attribute*, is also introduced in BGP Update message.

(5) After that, a QoS Join message is transferred through domains *D1* and *Ds* in a similar way. That is, it is sent from *R11* to *r12*, and Link State Update messages are sent to multicast routers in *D1*. Then it is sent from *r12* to *r13*, and follows the flooding of Link State Update messages reporting updated available bandwidth of the link from *r13* to *r12*, and so on. When it arrives at BGMP border router *R15*, a BGMP QoS Join message is sent to *Rs*, next border router, and BGP and IBGP Update messages are exchanged among border routers in the whole network. As a result, QoS aware path from router *rs1* to router *rr1* is constructed.

3.2 Protocol Procedure for Crank Back Mechanism

In the procedure described above, BGMP border router *Rr* selects the next domain based on only the information of inter-domain links. In the example above, *Rr* selected domain *D1* as the next hop domain. However, since *Rr* does not have any information on the available bandwidth of links within *D1*, it is possible that there are no paths in *D1* with enough bandwidth from *R15* to *R11*. In such a case, it is considered that Rr needed to select *D2* as the next domain to domain *Ds*. This is due to limiting the exchanging range of link state advertisement messages. In order to cope with such a problem, we introduce the crank back mechanism which allows a border router to retry a tree construction using other next domains. As shown in Fig. 7, border router *R11* sends a nack message, a *Join NACK message*, to *Rr*, and *Rr* tries to construct QoS aware multicast tree by forwarding a QoS Join message via domains *D2*, *D3* and *Ds*. The following are the detailed procedure and the format of messages.

(1) When *R11* receives a QoS Join message and finds that the QoS aware path to domain *Ds* cannot be prepared in domain *D1*, it sends Join NACK message to *Rr* as shown in Fig. 8. It needs to be noticed that there are no Update messages at this step, because *R11* does not reserve any bandwidth for this communication.

(2) A Join NACK message exchanged between BGMP border routers is defined according to BGMP message format. Figure 9 shows its format, containing a new attribute, *Join NACK Attribute*, indicating reason code. Although not shown in Fig. 8, it is possible that a Join NACK message is returned through a path constructed within a domain according to PIM-SM. Therefore, we also defined a Join NACK message exchanged between PIM-SM routers

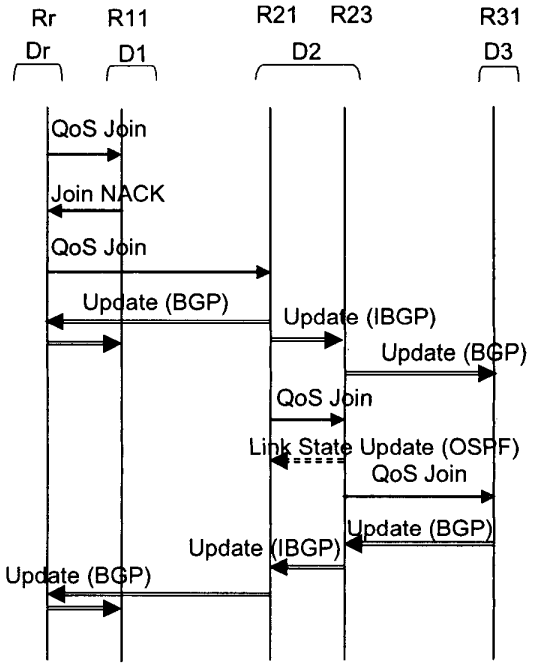


Figure 8 Communication Sequence of Crank Back Mechanism

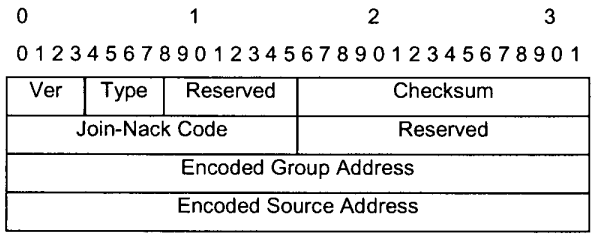


Figure 9 Format of Join NACK Message for BGMP

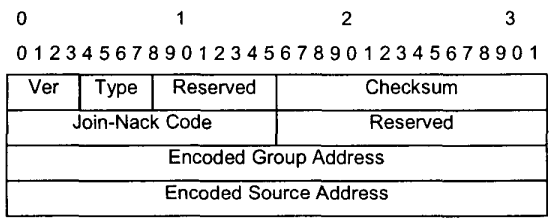


Figure 10 Format of Join NACK Message for PIM-SM

according to PIM-SM message format. Figure 10 shows the format of a PIM-SM based Join NACK message.

(3) When *Rr* receives a Join NACK message, it selects an alternative for next hop domain to domain *Ds*. It uses the information such as shown in Table 2. In this case, domain *D2* will be an alternative. So, it send a QoS Join message to *R21*, border router in *D2*. If *R21* can construct a path with required bandwidth in domain *D2*, it reserves the bandwidth at the inter-domain link from *R21* to *Rr*, and then it send BGP and IBGP Update messages to border routers. After that, this QoS Join message is relayed via *R23*, *R31* and so on. With this procedure, a QoS aware multicast tree is constructed by using a secondary route from domain *Ds* to *Dr*.

4. Evaluation

In order to evaluate the QoS aware multicast routing protocol we propose, we perform a simulation evaluating the number of exchanged link state advertisement messages (OSPF Link State Update messages and BGP Update messages) when multiple senders or receives exist in a multicast network consisting of multiple domains.

4.1 Simulation Conditions

(1) We assume the network whose configuration is shown in Fig. 11. It consists of 13 domains and each domain has a similar network configuration shown in Fig. 11 (a). That is, each domain has 24 multicast routers, and the routers shown in Fig. 11 (b) work as BGMP border routers.

(2) Only the leaf routers in a domain (routers 1, 3, 5, 12, 13, 14, 15, 18, 19, 23, 24) initiate a QoS Join attempt (send a QoS Join message initially), if they are not working as border routers.

(3) All of the links have enough bandwidth for all of QoS Join attempts. That is, all QoS Join messages will succeed.

(4) All of the links have the same costs, and therefore, QoS aware multicast tree within a domain is constructed using the shortest path.

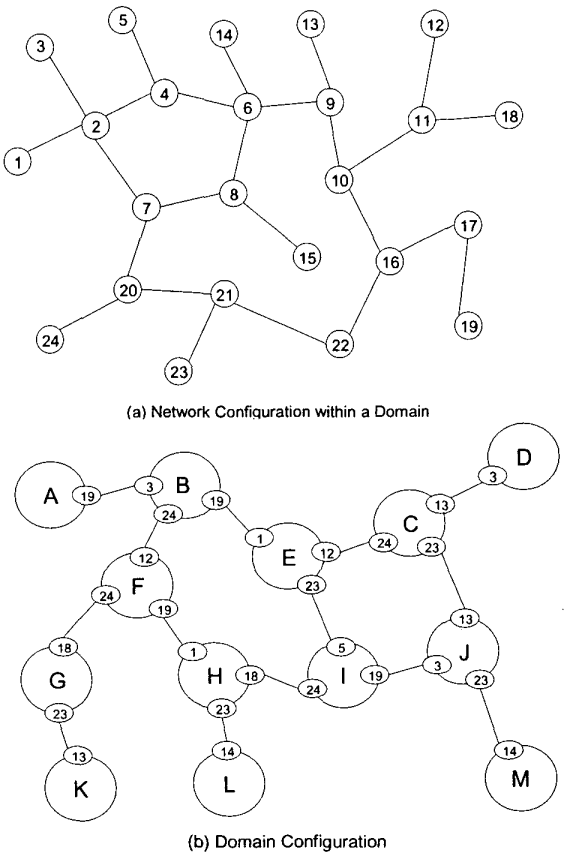


Figure 11 Network Configuration for Simulation

(5) Each QoS Join attempt requires the same bandwidth, and therefore, a QoS Join message will be forwarded until the router that has been already receiving the multicast traffic which this QoS Join message is requesting.

(6) When a QoS Join message is sent in a domain, OSPF Link State Update message is delivered to each router in the same domain. We assume that one Link State Update message is used to be delivered to one router.

(7) When a QoS Join message is sent over an inter-domain link, BGP or IBGP Update message is delivered to each border router in the whole network. Similarly we assume that one Update message is used to be delivered to one border router.

(8) We also evaluate the case of flat network configuration. That is, 312 routers (13 times 24 routers) in the network exchange OSPF Link State Update messages if the bandwidth of a link is newly reserved, that is, a QoS Join message is transferred over the link.

(9) We have performed the following two kinds of simulation runs.

- *Simulation 1*:
  There is only one sender in the whole network. It is located under router 13 in domain D. It is sending a multicast traffic of a specific multicast group. In this situation, receivers located under the leaf routes initiate QoS Join one by one. The order of QoS Join attempts is randomly selected from the leaf routers in the whole network.
- *Simulation 2*:
  There is one sender in each domain. The location of sender is selected randomly in each domain. A sender in one domain is sending a multicast traffic of a specific multicast group. That is, each sender is sending traffic with different group. Similarly to simulation 1, receivers located under the leaf routes initiate QoS Join one by one. The order of QoS Join attempts is randomly selected. There may be one receiver for a specific group under an individual leaf router. This means that there may be 13 receivers under one leaf router.

(10) For the evaluation, we measure the average of the number of link state advertisement messages which one router handles (sends, relays or receives). In the case of flat network configuration, we count the number of OSFP Link State Update message which individual routers handed and calculate its average. In the case of network configuration with domains, we perform similar calculation for OSPF Link State Update messages and BGP Update messages.

4.2 Simulation Results

Figures 12 and 13 show the results for simulation 1. In the flat network configuration, the number of link state advertisement messages (OSPF Link State Update messages) is rapidly increasing as the number of receivers increases. On the other hand, in the network configuration with domains, the number of messages (OSPF Link State Update messages for intra-domain and BGP Update message for inter-domain) does not increase so rapidly as the case of flat configuration. As a result, it is considered that our protocol can decrease the number of link advertisement messages by the order of 1/(number of domains) compared with the conventional approaches.

Figures 14 and 15 show the results for simulation 2. In this simulation, we obtained similar results with simulation 1. It is considered again that our protocol can reduce the messages by the order of 1/(number of domains) of the conventional approaches.

5. Conclusions

This paper described the proposal of a new QoS aware multicast routing protocol that is scalable in terms of the number of exchanged link state advertisement messages. In our approach, a multicast network is divided into domains. For the intra-domain multicasting, we used PIM-SM and OSPF by adding some QoS functions. When a shortest path tree is requested, a *QoS Join message* is exchanged and the QoS aware multicast tree is constructed. Here, the range of link state advertisement is limited within one domain. For the inter-domain multicasting, we used BGMP and BGP-4 with some extensions. By use of QoS Join message, a shortest path tree is constructed with satisfying QoS requirement. Among BGMP border multicast routers, only the link state information of inter-domain links is advertised. As a result, the number of exchanged link state advertisement messages is drastically reduced. However, when a multicast tree spread over multiple domains, the tree construction needs to be performed without information on links in other domains, and it is possible to tree construction may fail. In order to cope with such a problem, we introduce the crank back mechanism using a *Join NACK message* introduced to PIM-SM and BGMP.

This paper described the detailed procedures and the message format of the proposed protocol. It also described the evaluation of the number of exchanged link state advertisement messages and showed that our protocol can reduce the number by the order of 1/(number of domains) comparing with the conventional protocols using a flat network configuration.

Reference

[1]: S. Biswas, et al., "A QoS-Aware Routing Framework for PIM-SM Based IP-Multicast," work in progress, <draft-biswas-pim-sm-qos-00.txt>, June 1999.

[2]: K. Fujikawa and K. Ikeda, "RSVP Integrated Multicast (RIM)," in Proc. of INET '99, June 1999.

[3]: K. Fujikawa, et al., "Integration of Multicast Routing and QoS Routing," in Proc. of INET 2000, July 2000.

[4]: W. Moh and B. Nguyen, "Minimizing Multiple-Shared Trees for QoS-Guaranteed Multicast Routing," in Proc. of ICCCN '99, October 1999.

[5]: W. Moh, et al., "Extending BGMP for QoS-based Inter-Domain Multicasting over the Internet," in Proc of ICC2000, June.2000.

[6]: D. Estrin, et al., "Protocol Independent Multicast – Sparse Mode (PIM-SM): Protocol Specification," RFC 2362, June 1998.

[7]: R. Braden, et al., "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification," RFC 2205, September 1997.

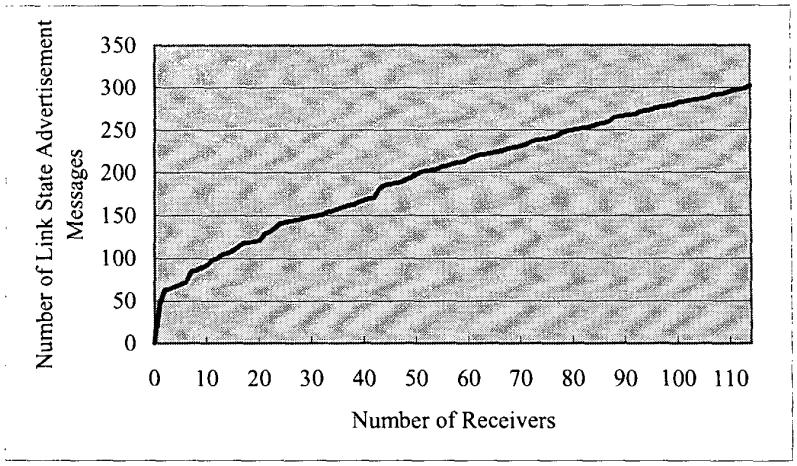
[8]: D. Thaler, et al., "Border Gateway Multicast Protocol (BGMP): Protocol Specification," work in progress, <draft-ietf-bgmp-spec-01.txt>, March 2000.

[9]: J. Moy, "Multicast Extensions to OSPF," RFC 1584, March 1994.

[10]: J. Moy, "OSPF Version 2," RFC 2328, April 1998.

[11]: Y. Rekhter and T. Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 1771, March 1995.

[12]: R. Coltun, "The OSPF Opaque LSA Option," RFC 2370, July 1998.



Figuer 12 Results of Simulation 1 (Flat Configuration)

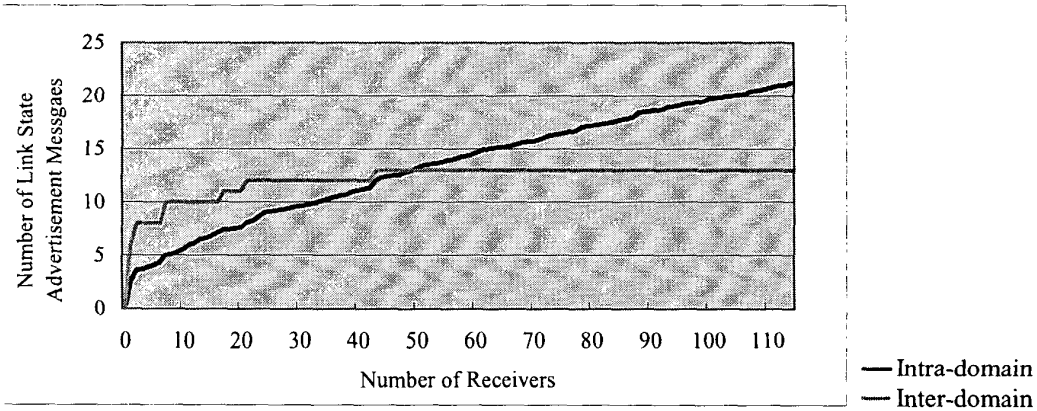


Figure 13 Results of Simulation 1 (Configuration with Domains)

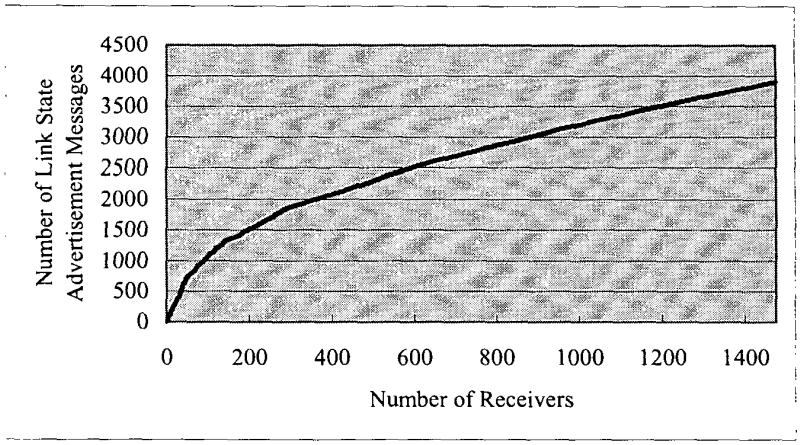


Figure14 Results of Simulation 2 (Flat Configuration)

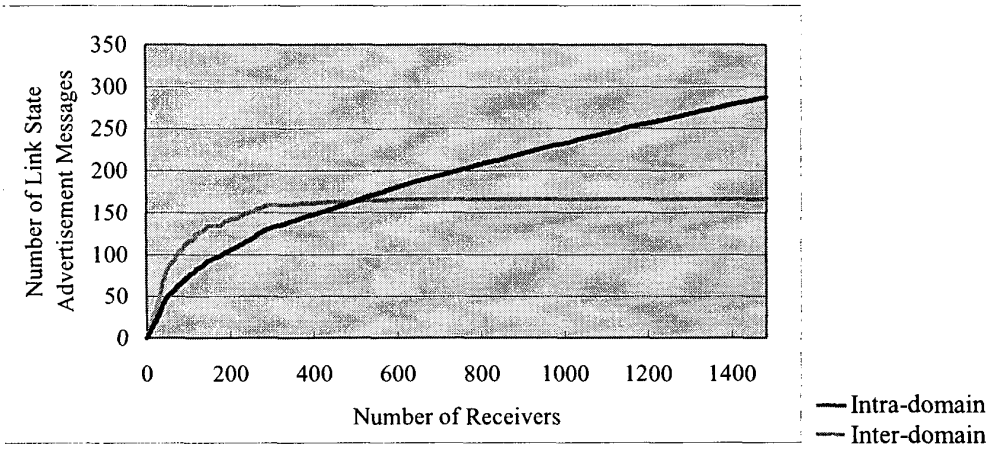


Figure 15 Results of Simulation 2 (Configuration with Domains)