

인터넷 멀티캐스트 라우팅에서 지연시간에 대한 QoS를 지원하는 방법

(A QoS of Delay supporting Scheme for IP Multicast Routing)

박 세 훈 * 안 상 현 **
(Se-Hun Park) (Sanghyun Ahn)

요 약 인터넷 멀티캐스트 라우팅을 위해 PIM, CBT, DVMRP 등의 여러 가지 프로토콜이 연구되고 있다. 본 논문에서는 이와 같은 멀티캐스트 라우팅에서 지연시간에 대한 QoS를 보장하기 위한 새로운 방법 DPIM(Delay-bounded PIM)을 제시한다. 이 방법은 기존의 PIM-SM을 개선한 형태로, 동적인 멤버십에 의해 주어진 지연시간 상한값을 만족하지 않는 상황이 발생하면 PIM-SM 트리의 RP(Rendezvous Point)를 다시 선택하고 트리를 재구성함으로써 지연시간에 대한 QoS를 보장한다. 기존에 제안된 다른 프로토콜과의 비교를 통해 지연시간에 대한 이 방법의 유효성을 확인하였다.

Abstract For the Internet multicast routing, there are some protocols researched. This paper presents a new protocol called DPIM (Delay-bounded PIM) to support QoS of delay in multicast routing. This protocol is based on PIM-SM. When the specified delay bound is not satisfied by dynamic membership, the RP (Rendezvous Point) of shared tree is changed and the shared tree is reconstructed. Therefore, the QoS of delay is always supported in this protocol. Through performance analysis and comparison with other protocol, we certify that this protocol is pertinent.

1. 서 론

인터넷에서 멀티캐스트 라우팅을 지원하기 위한 연구는 두 가지 접근법을 통해 진행되고 있다. 첫 번째는 밀집 방식(dense mode) 접근법으로 멀티캐스트 그룹의 구성원들이 밀집되어 분포되어 있고 대역폭이 충분한 경우에 적절한 접근법이다. 그룹 내의 각 송신자에 대해 송신자를 루트로 하는 별개의 트리를 설정하기 때문에 송신자-기반(source-based) 트리 접근법, 또는 data-driven 접근법이라고도 불린다. 이 접근법의 대표적인 프로토콜로는 DVMRP(Distance Vector Multicast Routing Protocol)[1], MOSPF (Multicast Open Shortest Path First), PIM-DM(Protocol Independent Multicast-Dense Mode)[2] 등이 있다.

두 번째는 희소 방식(sparse mode) 접근법으로 멀티캐스트 그룹의 구성원들이 희소하게 분포되어 있고 대역폭이 충분하지 않은 경우에 적절한 접근법이다. 각 그룹들이 오직 하나의 공유 트리를 가지고, 그 그룹의 각 수신자들이 공유 트리에 Join하는 방법으로 트리를 설정하기 때문에 공유 트리 접근법, 또는 수신자-기반(receiver-initiated) 접근법이라고도 한다. 이 방법의 대표적인 프로토콜로는 CBT(Core-Based Tree)[3], PIM-SM(Protocol Independent Protocol-Sparse Mode)[4] 등이 있다.

이 두 가지 접근법 중 현재는 희소 방식 접근법에 연구가 집중되고 있는데, 이는 밀집 방식 접근법이 가지는 단점들 때문이다. 밀집 방식 접근법은 각 송신자마다 개별적인 트리가 존재하므로 구성원이 되는 각각의 라우터들이 가져야 할 정보가 많고, 또한 트리를 구성하기 위해 플러딩 혹은 브로드캐스팅에 의존한다. 따라서 그룹의 구성원이 많아지고, 망이 커질수록 비효율적이다. 그러나 희소 방식 접근법은 하나의 공유 트리만 유지하므로 라우터가 유지해야 할 정보가 적고, 또한 트리를

* 비 회 원 : 한국 IBM 소프트웨어연구소 연구원
aduksiny@dreamwiz.com

** 종신회원 : 서울시립대학교 전산통계학과 교수
(corresponding author)
ahn@venus.uos.ac.kr

논문접수 : 2000년 3월 28일

심사완료 : 2001년 8월 20일

구성하는 데 플러딩과 같이 망 자원을 많이 쓰는 방법을 쓰지 않는다. 따라서 보다 확장이 용이(scalable)하고 현재의 인터넷 환경에 적절하다.

그러나 최소 방식 접근법도 데이터 전달이 항상 공유 트리의 루트를 지나가야 하므로 루트가 병목으로 작용하게 되는 단점이 있다.

인터넷 멀티캐스트 라우팅에서 최근의 주제는 QoS를 지원하는 것이다. 다른 인터넷 응용들과 마찬가지로 대부분의 인터넷 멀티캐스트 라우팅의 응용들이 QoS 지원을 요구하고 있다. 예를 들어 화상회의와 같은 응용에서는 모든 종단간의 최대 지연시간이 어느 수준 이하이기를 원한다.

따라서 본 논문은 인터넷 멀티캐스트 라우팅의 두 가지 접근법 중 최소 방식 접근법을 기반으로 한 프로토콜에서 QoS를 지원하는 방법을 제시한다. PIM-SM 프로토콜을 기반으로 하여 종단간 지연시간(end-to-end delay)에 대한 QoS 요청을 해결하는 개선된 프로토콜 DPIM(Delay-bounded PIM)을 제시한다. DPIM은 QoS 요청에 대해 능동적으로 대처함으로써 QoS 요청을 반드시 해결하는 것을 목표로 한다.

2장에서는 관련된 연구들을 개괄적으로 설명하고, 3장에서는 본 논문의 핵심적인 제안들을 상세하게 설명한다. 4장에서는 성능 분석 및 기존 프로토콜들과 본 논문에서 제시하는 DPIM을 비교하고 5장에서 결론을 맺는다.

2. 관련 연구

공유 트리 접근법의 대표적인 프로토콜인 PIM-SM과 CBT는 QoS에 대한 고려 없이 개발된 프로토콜이다.

PIM-SM은 RP(Rendezvous Point)를 루트로 하는 공유 트리를 이용해 멀티캐스트 데이터를 배포한다. PIM-SM의 동작에 대한 간단한 개념도는 [그림 1]과 같다. 멀티캐스트 그룹에 새로이 참여하는 구성원(실제로는 그룹의 구성원이 연결된 서버망의 DR(Designated

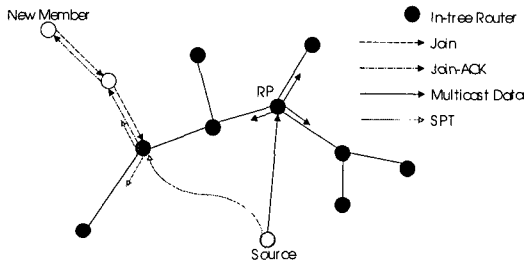


그림 1 PIM-SM의 동작

Router)은 미리 알려진 RP에게 Join 메시지를 보내어 참여를 알린다. Join 메시지는 유니캐스트 메시지로 전달되며, RP로 전달되는 도중에 이미 공유 트리의 일원인 라우터에 도달하면 더 이상 전달되지 않는다.

Join 메시지를 받은 라우터는 Join에 대한 승인메시지를 되돌려 보내고 이를 새로이 Join하는 구성원이 받음으로써 Join 과정은 완성된다. 멀티캐스트 데이터를 보내려는 송신자는 그 데이터를 유니캐스트 메시지로 캡슐화하여 RP에 보내고, 이를 받은 RP는 데이터를 역캡슐화하여 트리 전체에 전달한다. 이때 만약 어떤 수신자가 단위 시간 동안 받는 데이터 양이 일정 수준보다 커진다면 송신자에서 RP를 거치지 않고 직접 데이터를 보내는 송신자-기반 SPT를 구성하기도 한다.

위와 같은 과정에서 PIM-SM은 QoS에 대한 고려를 전혀 하지 않는다. 본 논문은 이러한 PIM-SM을 QoS 제공을 위해 개선하는 방법을 제안한다.

CBT는 PIM-SM과 비슷하게 RP 대신 core라 부르는 루트를 중심으로 공유 트리를 구성한다. 구성원의 Join/Leave는 PIM-SM과 거의 동일하다. 그러나, 송신자가 공유 트리의 구성원일 필요가 없는 PIM-SM과는 달리 송신자는 항상 공유 트리의 구성원이며, 송신자가 보내는 멀티캐스트 데이터는 core에서부터 배포되는 것이 아니라 송신자에서 core로 가는 도중에 만나는 다른 트리 내 라우터(in-tree router)들의 하위 트리에도 전달된다. 즉 PIM-SM은 단방향 트리이며 CBT는 양방향 트리이다. PIM-SM과 CBT의 멀티캐스트 데이터 배포의 차이는 [그림 2]에 잘 나타나 있다. 또한 CBT는 PIM-SM과 달리 데이터 전송량이 많아져도 SPT를 사용하지 않는다.

본 논문과는 다른 방식으로 인터넷 멀티캐스트 라우

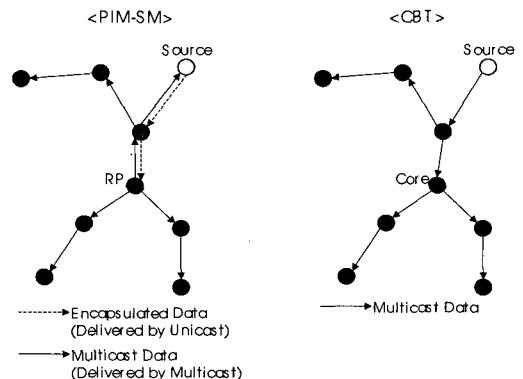


그림 2 PIM-SM과 CBT의 차이

팅에서 QoS를 지원하기 위해 제안된 프로토콜이 몇 가지 있다. 이들은 모두 공유 트리 접근법을 사용한다.

'QoS MIC(Quality of Service sensitive Multicast Internet protoCol)' [5]는 루트가 없는 공유 트리를 사용한다. 따라서 멀티캐스트 데이터는 CBT와 비슷하게 양방향 트리에 의해 전달된다. 그러나 PIM-SM의 RP나 CBT의 core 같은 루트 역할의 라우터가 없는 대신 관리자 라우터(Manager router)가 있어서 그룹을 관리하고 Join에 관여한다. QoS 지원을 위해 각 연결에 대해 대체 경로를 제공하고, 이를 위해 두 가지의 Join 방법을 제공하나 Join 과정의 오버헤드가 크고 Join에 걸리는 시간이 길다는 단점이 있다. 'QoS MIC'은 필요할 경우 특정 송신자에 대해서 PIM-SM에서처럼 송신자-기반의 트리로 전환도 가능하다.

'QoS Extension to CBT' [6]는 QoS 문제를 지원할 수 있도록 기존의 CBT에서 구성원들의 Join과 Leave 과정 및 상태(state) 갱신 과정들을 확장한 것이다. 종단간 지연시간, 최대 패킷 손실율, 최소 대역폭 등과 같은 다양한 QoS 요청들을 처리할 수 있고, 공유 트리 방식을 사용하는 다른 멀티캐스트 라우팅 프로토콜에도 적용 가능하다. 그러나 Join시에 주어진 QoS를 만족하지 못하는 경우 무조건 Join을 거부하며, 하나의 트리 내 라우터는 각 QoS 조건당 (하향 인터페이스의 개수)+1의 상태들을 추가로 요구하므로, 라우터가 가지고 있어야 할 상태가 많아 기존의 CBT에 비해 확장성이 떨어진다.

4장에서는 위에서 기술한 프로토콜들을 큰 논문에서 제시하는 DPIM과 비교하여 DPIM이 기존 프로토콜보다 어떤 면에서 나은지를 설명할 것이다.

3. DPIM

본 논문에서 고려하는 QoS 요구사항은 종단간 지연 시간이다. 화상 회의와 같은 멀티캐스트의 추된 응용들에서 중요한 것은 한 송신자의 영상 및 음성을 수신하는 수신자들이 거의 비슷한 시간대에 보고 듣는다는 것이다. 이때 다른 수신자보다 좀 더 빨리 수신하는 것은 별 문제가 되지 않으나, 다른 수신자들에 비해 훨씬 나중에 수신하는 것은 문제가 된다. 즉 송신자가 데이터를 전송할 때, 모든 수신자에게 특정 시간 내에 모두 전송해야 하는 것이다. 이 문제에 있어서 QoS의 다른 요소들인 대역폭이나 패킷 손실율 등도 영향을 끼치기는 하지만, 직접적인 영향을 끼치는 것은 지연시간이다. 따라서 본 논문은 종단간 지연시간에 대한 QoS 요구사항을

처리하는 것을 목적으로 한다. 결과적으로 DPIM의 목표는 다음과 같다.

목표: 멀티캐스트 공유 트리에서 모든 두 구성원 사이의 지연시간은 항상 미리 결정된 (응용에 의해 주어진) 상한값 내에 있다.

상한값은 수신자가 기다릴 수 있는 최대의 시간으로, 이 상한값을 넘는 구성원 쌍이 발견되면 멀티캐스트 트리를 수정함으로써 문제를 해결한다.

3.1 가정

멀티캐스트 응용들 중에는 모든 구성원이 수신자이자 송신자인 경우도 있지만, 그렇지 않고 송신자이긴 하지만 수신자는 아닌 경우도 있다. [그림 1]에서처럼 기존의 PIM-SM에서는 꼭 송신자가 수신자일 필요는 없으나, 송신자가 공유 트리의 구성원이 아닌 경우 종단간 최대 지연시간을 트리 내에서 구하기가 어려워진다. 따라서 본 논문에서는 모든 송신자는 곧 수신자로서 공유 트리의 구성원임을 가정하였다.

지연시간의 척도로는 경로의 길이, 즉 경로의 홉(hop) 수를 사용한다. 실제 망에서는 각 링크마다 대역폭이 다르고 시간에 따라 사용량이 변화하므로 경로의 길이가 지연시간을 정확하게 반영한다고 보기는 어려우나, 본 논문에서는 구현의 편이를 위해 경로의 길이를 사용한다. 만약 각 링크의 지연시간을 구하는 방법이 망에서 제공된다면 쉽게 실제 지연시간을 사용하도록 수정할 수 있다. 따라서 본 논문에서는 지연시간과 경로의 길이를 거의 같은 의미로 사용하며 실제 지연시간을 사용하는 경우에 대하여 따로 설명할 것이다.

2장의 [그림 2]에서 설명한 것처럼 공유 트리 기반의 멀티캐스트에서 데이터를 전달하는 방법은 두 가지가 있다. 본 논문은 PIM-SM을 기반으로 하므로 모든 멀티캐스트 데이터는 RP에서 배포된다. 따라서 지연시간에 대한 상한값을 초과하는 경우를 검사하는 위치는 항상 루트, 즉 RP가 된다.

RP나 core는 전체 공유 트리의 관리자로서 기능하며 또한 데이터 전달의 중심으로서 병목으로 작용할 수 있으므로 일반적인 라우터보다 좋은 성능과 추가적인 기능이 필요하다. 따라서 RP가 될 수 있는 후보 라우터(Candidate-RP; C-RP)가 망에 다수 존재함을 전제로 한다. 또한 모든 라우터는 이 C-RP의 목록을 가지고 있다고 가정한다. 기존의 PIM-SM은 RP가 손상되었을

때를 대비한 대체 라우터로서 C-RP를 사용하나, DPIM은 RP 손상의 대체 라우터 목적과 함께, 종단간 지연시간에 대한 QoS 요구를 만족하기 위해 멀티캐스트 트리를 수정하게 될 경우 RP를 바꾸기 위해 C-RP들을 유지한다.

3.2 DPIM의 작동방법

그룹에 참여할 구성원들의 위치와 전체 망의 구조를 사전에 알 수 있다면 RP 선정시에 QoS 요구사항에 적절한 라우터를 RP로 선정함으로써 QoS 문제를 해결할 수 있을 것이다. 그러나 동적 멤버십 하에서는 사전에 정보를 알기가 어려우므로 본 논문은 상황에 따라 RP를 다시 선택하여 트리를 재구성하는 방법을 선택하였다. 따라서 동적 멤버십 하에서 공유 트리의 구성에 따른 지연시간 정보를 알아내고, 지연시간 상한값을 만족하지 못하는 경우를 발견할 수 있는 방법이 필요하다.

기본적인 틀은 트리에 속한 각 라우터들이 자신과 하위 단말 노드까지의 지연시간 값들 중 최대의 값을 유지하는 것이다. 따라서 새로운 구성원의 Join/Leave 시에 각 라우터들의 지연시간 값은 갱신되고, 만약 이 갱신된 값의 조합이 상한값 D 를 벗어나게 되면 트리를 재구성한다.

3.2.1 각 트리 내 라우터들이 유지해야 하는 상태

트리 내 라우터 u 는 PIM-SM에서 필요로 하는 상태 외에 하나의 상태를 추가로 유지한다.

$d_{max}(u)$: u 에서 하향 인터페이스로 도달 가능한 가장 먼 단말 라우터까지의 지연시간

단 RP는 $d_{max}(root)$ 대신 다음의 상태들을 유지한다.

$d_{max,i}(root)$: 하향 인터페이스 i 로 도달 가능한 단말 라우터 중 가장 먼 것까지의 지연시간

$d_{max}(u)$ 는 u 를 루트로 하는 서브트리에 새로운 구성원이 Join하거나 기존 구성원이 Leave할 때마다 갱신된다. 물론 링크가 손실되는 등의 경우에도 갱신되어야 한다.

위 상태 값들을 얻기 위해 Join 시에 Join하는 구성원과 트리 내 라우터 간의 지연시간 값을 얻는 방법이 필요하다. 본 논문에서는 지연시간으로 경로의 길이를 사용하므로, 가장 간단한 방법으로 IP 헤더에 있는 TTL(Time To Live field)을 이용하는 방법을 생각할 수 있다. 예를 들어, 모든 Join 메시지는 TTL을 상한값 D 로 하도록 미리 결정해 둔다면, Join 메시지를 받은 트리 내 라우터는 받았을 때의 TTL 값과 D 의 차로부터 Join하는 구성원까지의 경로 길이를 알 수 있다.

만약 지연시간의 척도로서 경로 길이 대신 실제 지연시간 값을 반영하도록 한다면 TTL 대신 망에서 지원하는 다른 수단을 이용하는 방법이 고안되어야 할 것이다.

위 방법을 이용하여 $d_{max}(u)$ 는 트리 내 라우터 u 에 새로운 구성원 v 가 Join할 때, 기존의 값보다 Join 메시지에서 얻은 값 $d(v,u)$ 가 클 경우 새로운 값으로 갱신된다.

여기서 $d(v,u)$ 는 라우터 v 에서 라우터 u 까지 메시지가 전달되는 데 걸린 지연시간이다. DPIM에서 트리 구성 및 재구성에 사용되는 모든 메시지들은 전달되는 도중의 라우터들과 최종 목적지 라우터에게 출발지로부터의 지연시간을 알려줄 수 있어야 한다.

3.2.2 새로운 구성원의 Join

다른 수신자-기반 프로토콜과 마찬가지로 새로이 그룹에 Join하고자 하는 구성원에 연결된 DR(Designated Router) v 는 그 그룹의 RP를 향해 Join 메시지를 보낸다. RP의 주소는 멀티캐스트 응용의 시작 과정에서 미리 알려진다. Join 메시지가 트래 내 라우터 u 에 도달하게 되면 트리 내 라우터는 필요한 처리를 하고 Join-ACK 메시지를 되돌려 보내 Join을 허가한다. 이 과정에서 필요한 처리는 다음과 같다.

트리 내 라우터 u

Join 메시지에서부터 알아낸 v 까지의 지연시간 $d(v,u)$ 과 $d_{max}(u)$ 를 비교하여 $d(v,u)$ 가 더 클 경우 $d_{max}(u)$ 의 값을 $d(v,u)$ 값으로 바꾸고, Join-Update 메시지를 부모 라우터 w 에 전송한다. 이 Join-Update 메시지는 $d_{max}(u)$ 값을 운반한다. 또한 이 메시지는 지연시간 값을 알 수 있는 방법을 제공해야 한다. 마지막으로 v 에 Join-ACK 메시지를 보내면서 $d(v,u)$ 를 함께 전송한다.

v 에서 u 로의 경로 상에 위치한 각 라우터 x

새롭게 트리 내 라우터가 되는 이들 라우터는 Join-ACK 메시지를 받으면 $d_{max}(x)$ 상태를 새로이 생성하고 $d(v,u)$ 에서 $d(v,x)$ 를 뺀 값을 부여한다.

Join하는 라우터 v

Join-ACK 메시지를 받으면 $d_{max}(v)$ 를 생성하고 그 값으로 0을 부여한다.

u 에서 RP까지의 경로에 위치한 트리 내 라우터 w

Join-Update 메시지에서 얻은 $d_{max}(u)+d(u,w)$ 값이 기존의 $d_{max}(w)$ 보다 크다면 $d_{max}(w)$ 값을 $d_{max}(u)+d(u,w)$ 로 바꾸고 Join-Update 메시지를 상향으로 전달한다. 이 과정은 Join-Update 메시지가 루트에 도달하거나 $d_{max}(u)+d(u,w)$ 값이 $d_{max}(w)$ 보다 작거나 같은 라우터에 도달할 때까지 계속된다.

RP

Join-Update 메시지를 받은 하향 인터페이스 i 에 대해 $d_{max,i}(root)$ 를 Join-Update 메시지에서 얻은 $d_{max}(u)+d(u,root)+1$ 로 바꾼다. 자신이 가지고 있는 하향 인터페이스들의 최대 지연시간 값들의 조합이 미리 결정

2)에서부터 다시 시작한다.

- 6) ReSel-Complete 메시지를 받은 구성원 라우터는 새로운 RP에게 Join 메시지를 보낸다. 이 Join은 기존의 Join과 같다. ReSel-Complete 메시지는 범위를 1 감소시키고 계속 전달한다. 만약 범위가 0이 되면 ReSel-Complete를 더 이상 전달하지 않는다.
- 7) 자신을 RP로 지정한 Join 메시지를 받은 C-RP는 자신이 새로운 RP가 되었음을 인식하고, 앞으로 들어올 Join에 대해 RP로서 응답한다. 즉 Join-ACK를 보낸다. 또한 앞으로의 Join을 기다리기 위해 새로운 타이머를 시동한다.
- 8) 타이머가 종료되어 새로운 트리의 구성이 끝나면 새로운 RP는 자신이 RP가 되었음을 플러딩 등의 방법을 통해 망 전체에 알린다.
- 9) Join-ACK를 받은 구성원 라우터는 이전 RP에 대한 동적 갱신(dynamic refresh)을 끝낸다.

위에 서술된 과정에 대한 간략한 예가 [그림 4]에 나타나 있다.

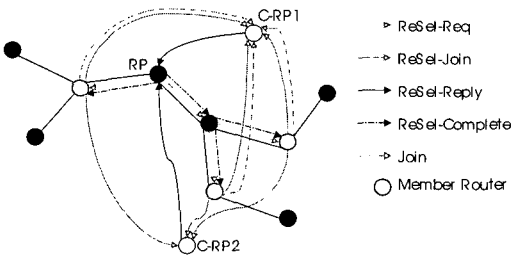


그림 4 RP 재선택 및 트리 재구성

ReSel-Req와 ReSel-Complete 메시지는 범위(scope) 필드를 가진다. 위 과정에 기술된 것처럼, 이 범위를 1씩 증가시키면서 지연시간 상한값을 만족하는 C-RP를 발견할 때까지 ReSel-Req 메시지의 전송이 반복하여 새로운 RP를 찾는다. 따라서 최악의 경우 모든 구성원 라우터들이 다시 새로운 RP에 Join해야 하는 경우가 발생할 수가 있다. 그러나 망 내에 중단간 지연시간에 대한 QoS 요청을 만족시킬 수 있는 C-RP가 존재하는 한 이 방법은 항상 DPIM의 목표를 만족시킨다.

트리 재구성이 끝나게 되면 망 내의 모든 라우터들은 해당 그룹에 대한 새로운 RP를 알고 있어야 한다. 따라서 8)에서처럼 새로운 RP가 자신이 RP가 되었음을 망 전체에 브로드캐스팅 메시지로 알리게 된다.

위 5)에서 RP가 C-RP들 중에서 어떻게 가장 좋은 것을 택할 것인가에 대해선 여러 가지 방법이 있을 수

있다. 본 논문에서 채택한 기준은 다음과 같다.

RP 재선택시 선정 기준

- 1) 각 구성원 라우터와 C-RP 간의 지연시간들의 조합이 항상 지연시간 상한값을 만족하는 C-RP들을 택한다.
- 2) 1)번에서 결정된 C-RP가 두 개 이상일 경우 지연시간들의 조합의 최대값이 가장 작은 C-RP를 택한다. 하나 뿐이면 그 C-RP를 새로운 RP로 택한다.
- 3) 2)번에서 결정되지 않을 경우, 지연시간의 총합이 가장 작은 것을 택한다.
- 4) 3)번에서도 결정되지 않을 경우, 임의의 C-RP를 택한다.

C-RP는 ReSel-Join 메시지를 보낸 구성원 라우터의 수를 알지 못하므로 ReSel-Join 메시지의 전송을 계속 기다리게 될 것이다. 이를 막기 위해 타이머를 둔다. 타이머는 ReSel-Join 메시지뿐만 아니라 Join 메시지들을 받을 때에도 필요하다. 그런데 Join 메시지는 트리 내 라우터를 만나게 되면 더 이상의 상황 전송이 중단되므로 보내진 모든 Join이 새로운 RP에게 전해지리라는 보장이 없다. 또한 Join 메시지는 트리 재구성 과정뿐만 아니라 동적 멤버십에 의해 항상 발생할 수 있다. 따라서 ReSel-Join 시와는 다른 타이머를 사용하여 타이머가 종료된 후에야 자신이 새로운 RP가 되었음을 전체 망에 알린다. RP는 C-RP의 개수를 알고 있으므로 ReSel-Reply 메시지에 대해 타이머를 사용할 필요는 없다.

RP 재선택 및 트리 재구성 과정을 수행하기 위해서 C-RP 및 RP가 취해야 하는 행동이 몇 가지 있다. 모든 C-RP는 항상 ReSel-Join에 대한 대비를 하고 있어야 하고, 또한 ReSel-Join을 받은 후에는 Join에 대한 대비를 하고 있어야 한다. 이전 RP가 구성원 라우터인 경우에는 이전 RP도 ReSel-Join 및 Join을 보내어 재선택 과정에 구성원 라우터로서 참여한다.

RP 재선택 과정에서 앞서 제시한 DPIM의 목표를 만족하는 C-RP가 없을 경우 끝없이 2)~6)의 순환을 반복하게 된다. 이를 막기 위해 ReSel-Join에는 이 메시지를 보낸 구성원 라우터가 ReSel-Req의 전달을 중단시켰음을 알리는 플래그를 둔다. C-RP가 한 순환동안 받은 ReSel-Join 내의 모든 플래그가 unset이 되면 2)~6)의 순환을 중지한다. 이 경우 새로운 RP의 선정을 위해 위 RP 재선택시 선정 기준을 1)번을 제외하고 적용하여 새로운 RP를 선택한다. 이렇게 되면 중단간 지연시간에 대한 QoS 요구를 만족하지 못하는 경우가 발생할 것이다.

위 과정이 진행되는 동안 데이터 손실을 막기 위해 기

존의 공유 트리는 그대로 유지된다. 기존의 공유 트리는 트리 재구성 과정에서 9)번의 과정에 의해 제거된다. 기존의 PIM-SM이나 CBT와 같은 공유 트리 접근법에서 모든 트리 내 라우터들은 동적 갱신을 주기적으로 계속 함으로써 트리를 유지한다. 즉 트리 내 라우터의 하향 인터페이스에 동적 갱신을 위한 메시지가 더 이상 전달 되지 않으면 라우터는 해당 인터페이스에 연결된 링크가 손실된 것으로 간주하고 관계된 상태들을 삭제한다. 따라서 트리 재구성에 의해 새로운 RP로부터 Join-ACK 메시지를 받은 라우터들은 새로운 트리에 동적 갱신을 계속 수행하고 이전 트리에는 하지 않음으로써 자연스럽게 이전 트리를 제거한다. 이를 위해서 동적 갱신의 주기는 망 직경의 2배보다 커야 한다. 이는 이전 RP에서 가장 멀리 있는 구성원 라우터가 새로운 RP에 대해서도 가장 멀리 있다고 가정할 때 ReSel-Complete, Join, Join-ACK 메시지가 전달되는 경로 길이의 총합이 망 직경의 2배를 넘지 않을 것이기 때문이다.

앞서 제시한 DPIM의 목표를 만족하는 C-RP가 망 내에 존재한다 해도 재선택 과정 중에는 종단간 최대 지연시간의 상한값을 만족하지 못하는 경우가 생기게 된다. 이는 오직 새로이 Join한 구성원이 한 종단이 되는 경우뿐이다.

RP 재선택 및 트리 재구성 과정 중에 새로운 구성원이 Join하려는 경우가 발생할 수 있다. 이는 RP 및 ReSel-Req 메시지를 받은 라우터들이 더 이상의 Join을 거부함으로써 해결한다. 만약 ReSel-Req 메시지를 미처 받지 못한 라우터가 Join을 허가했다 하더라도 이미 $d_{max}()$ 값은 갱신되어 있고 새로운 구성원에 대한 상태를 만들어 둔 이후이므로 RP 재선택 시의 조건이 약간 변화할 뿐 트리 재구성에 영향을 끼치지 않는다.

3.2.5 실제 지연시간의 적용

경로의 길이 대신 경로의 실제 지연시간을 적용하는 것은 망이 경로의 지연시간을 알 수 있는 방법을 제공한다면 간단하게 해결된다. 즉 Join, Join-Update, Leave, Delay-Reply, Leave-Update, ReSel-Join 메시지가 전달되면서 경로의 길이 대신 실제 지연시간 정보를 구하게 하면 된다. 그런데 실제의 망에서 지연시간은 회선 사용량에 따라 달라지게 되므로 시간이 따라 변화하게 된다. 따라서 시간에 따른 지연시간 변화를 반영할 방법이 필요하게 되는 데, 이는 PIM-SM의 동적 갱신을 이용하여 해결할 수 있다. 즉 동적 갱신에 쓰이는 메시지가 전달될 때, 실제 지연시간 정보를 상위 라우터에 전달하도록 하면 된다. 이때 동적 갱신은 주기적으로 일어나므로 시간에 따른 지연시간의 변화를 지속적으로

반영할 수 없는 문제가 아직 남아 있다. 그러나 이 문제는 본 논문의 영역을 벗어나는 문제이다.

4. 다른 프로토콜과의 비교 및 성능 분석

4.1 PIM-SM과의 차이점

본 논문에서 제시하는 프로토콜 DPIM은 PIM-SM을 QoS를 지원하기 위해 개선한 것이다. 따라서 기존의 PIM-SM과는 다음과 같은 부분들이 다르다.

- 1) PIM-SM은 다양한 QoS 요소 중 어느 것에 대해서도 고려가 없다.
- 2) PIM-SM에서는 송신자가 반드시 공유 트리의 구성원 즉, 수신자일 필요가 없다.
- 3) DPIM은 종단간 최대지연에 대한 QoS를 만족하기 위해 PIM-SM에서 사용하는 Join 및 Leave 메시지의 틀을 유지하되 몇 가지의 요소가 추가되었다. 또한 각 트리 내 라우터들이 유지해야 하는 상태도 추가되었다.
- 4) DPIM은 RP 재선택 및 트리 재구성을 위해 몇 개의 메시지와 타이머가 추가되었다. 타이머는 C-RP 및 새로운 RP가 유지해야 하는 상태 중의 하나이다.
- 5) DPIM은 PIM-SM에서 사용하는 SPT를 고려하지 않고 설계된 것이다. 그러나 SPT를 PIM-SM에서와 마찬가지로 사용한다면 DPIM에도 아무런 변경 없이 삽입될 수 있다.

4.2 성능 분석

4.2.1 RP 재선택 빈도

RP 재선택을 요구하는 상황 즉, 지연시간 상한값을 넘는 지연시간 조합이 RP에서 발견되는 빈도를 조사하기 위해 간단한 시뮬레이션을 수행하였다. 가상의 망을 임의로 만들고 망의 직경을 지연시간 상한값으로 하여 RP 재선택이 필요하게 될 때까지 Join한 구성원의 수를 측정하였다. 시뮬레이션은 C언어로 수행하였으며 자세한 사항은 다음과 같다.

시뮬레이션 방법

- 1) Waxman이 제시한 임의의 그래프 모델[7]을 이용하여 임의의 망을 생성하였다.
- 2) 지연시간의 척도는 경로의 길이로 하고, 지연시간 상한값은 망의 직경으로 정하였다.
- 3) C-RP의 개수는 망 내의 라우터 수의 20%이고, C-RP는 망 내의 라우터 중에서 무작위로 결정하였다.
- 4) RP는 C-RP중 하나를 무작위로 정하였다.
- 5) 망 내의 모든 라우터들 중에서 무작위로 하나씩을

RP에 Join시켰다. 즉 Join 순서는 무작위로 결정 되도록 하였다.

- 6) RP에서 지연시간 상한값을 만족하지 못하는 조합이 발견되면 위 3.2절에서 기술된 방법을 이용하여 새로운 RP를 구하고 트리를 재구성하였다.

임의의 망을 생성하는데 사용한 Waxman의 임의 그래프 모델에서 모든 노드는 직교 좌표계의 정수점에 임의로 지정된다. 두 노드 u 와 v 사이에 링크가 존재할 확률은 다음과 같다.

$$P(\{u,v\}) = \beta \exp \frac{-d(u,v)}{La}$$

여기서 $d(u,v)$ 는 직교 좌표계에서 구한 u 와 v 사이의 거리이고, L 은 가능한 가장 먼 거리이다. α 와 β 는 $0 < \alpha, \beta \leq 1$ 의 범위를 가지는 변수로, α 가 커지면 짧은 링크에 대해 긴 링크의 비율이 증가하고, β 가 커지면 평균 차수가 증가한다. 실험에서는 $\alpha=0.4$, $\beta=0.4$ 로 지정하였다.

시뮬레이션 결과는 다음 [표 1]과 같다. 각 망의 크기와 평균 차수에 대해 망의 생성을 10회 반복하여 시뮬레이션을 수행하였고, 그 평균을 표에 나타내었다. 표의 각 열은 몇 번째의 RP 재선택인가를 나타낸다. 즉 평균 차수 4, 라우터 수 50인 망에서 첫 번째 RP 재선택이 일어난 후 다음 RP 재선택이 요구될 때까지 Join한 구성원 수는 평균 4.5개이다.

표 1 RP 재선택이 일어날 때까지 Join한 라우터의 수

망의 크기/ 평균 차수	각 RP 재선택 사이에 Join한 라우터의 수				
	첫 번째	두 번째	세 번째	네 번째	다섯 번째
20/4	3.5	4.9	6.7	-	-
50/4	3.4	4.5	6.1	7.9	9.2
100/4	2.9	4.3	5.6	7.3	8.6

평균 차수 4, 라우터 수 20인 망에서 세 번째 이후의 결과가 없는 것은 다음 RP 재선택이 요구되지 않고 망 내의 모든 라우터가 Join한 경우가 발생했기 때문이다. 이러한 경우가 발생하는 이유는 지연시간 상한값을 망의 직경으로 지정했기 때문이다. 즉 망의 중심에 RP가 위치하게 됨으로써 모든 지연시간의 조합이 망의 직경보다 적거나 같아졌기 때문이다.

[표 1]의 결과를 보면 최초의 RP 재선택을 요구하는 경우가 2~4회 정도로 상당히 빈번함을 알 수 있다. 또한 새로운 RP를 중심으로 트리를 교체한 후에도 RP 재선택을 요구하는 상황이 빈번하게 발생한다. 그러나 새

로운 구성원의 Join이 계속될수록 RP 재선택이 일어날 때까지 Join한 라우터의 수가 점차 증가함을 알 수 있다. 이는 망의 중심 쪽으로 RP가 이동함에 따라 생기는 현상이다.

4.2.2 다른 프로토콜과의 비교

PIM-SM과 CBT는 QoS에 대한 고려가 전혀 없는 프로토콜이다. 따라서 4.2.1절에서 본 것처럼 종단간 지연시간에 대한 QoS 요청을 만족하지 못하는 경우가 빈번하게 발생해도 이를 해결할 수 없다. 즉 특정 종단이 다른 구성원들에 비해 멀티캐스트 데이터를 항상 늦게 받는 상황이 흔하게 발생함에도 불구하고 이 상황은 해당 종단이 Leave할 때까지 계속된다.

'QoSMIC'은 구성원의 수가 적고 구성원들의 분포가 지역적일수록 쉽게 종단간 지연시간에 대한 QoS 요청을 만족한다. 그러나 새로운 구성원이 Join할 때 기존 트리에서 가장 가까운 부분에 새로운 구성원이 연결되게 되므로 구성원의 수가 점점 늘어남에 따라 길게 늘어진 트리를 가지게 될 수 있다. 즉 새로운 구성원들의 Join 순서에 따라 [그림 5]에서처럼 종단간 지연시간 면에서 좋은 트리를 가지게 될 수도 있고, 길게 늘어진 나쁜 트리를 가지게 될 수도 있다. [그림 5]의 왼쪽 트리는 최대 종단간 지연시간이 6이 되나 오른쪽 트리는 5가 된다. 종단간 지연시간의 상한값이 5라고 하면 왼쪽 트리의 7번째 Join한 구성원은 다른 구성원들보다 항상 늦게 수신하게 된다. 그러나 오른쪽 트리에서는 상한값을 넘는 경우가 발생하지 않는다. 이와 같이 길게 늘어진 트리를 가지게 되는 경우 PIM-SM이나 CBT의 경우와 마찬가지로 특정 종단은 Leave할 때까지 다른 구성원들보다 멀티캐스트 데이터를 항상 늦게 받게 된다.

'QoS Extension to CBT'에서는 새로운 구성원이 Join할 때 지연시간 상한값을 만족하지 못하면 무조건

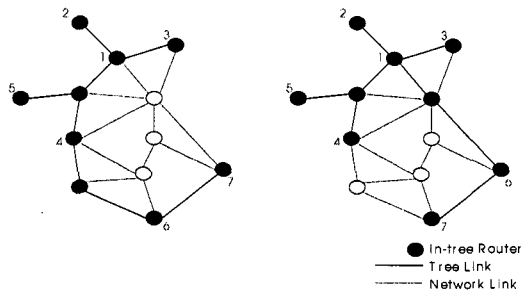


그림 5 QoSMIC에서 Join 순서에 따라 달라지는 트리의 구성의 예

Join을 거부한다. 따라서 Join하는 순서에 따라 나중에 Join하는 구성원은 아예 멀티캐스트 서비스를 받지 못할 가능성이 커진다. 즉 주어진 QoS를 만족하지 못하는 구성원은 멀티캐스트 서비스에서 배제되므로 이 프로토콜은 QoS에 결정적인(critical) 멀티캐스트 응용에만 적절하다고 볼 수 있다.

위 프로토콜들은 모두 특정 종단이 다른 구성원들에 비해 멀티캐스트 데이터를 지연시간 상한값보다 늦게 받는 경우가 일단 발생하면 그 상황은 해당 종단이 Leave할 때까지 지속된다. 'QoS Extension to CBT'의 경우에는 멀티캐스트 데이터를 늦게 받게 될 구성원은 멀티캐스트 서비스에서 배제된다. 그러나 DPIM은 그런 경우가 발생하면 모든 구성원이 지연시간 상한값 이내에 수신할 수 있도록 트리 형태를 개선한다. 따라서 구성원들의 Join 순서에 상관없이 모든 구성원들은 공정하게 서비스를 받을 수 있다.

PIM-SM, CBT, 'QoSMIC' 및 DPIM에 대해 지연시간 상한값을 만족하지 못하는 경우의 빈도를 시뮬레이션을 통하여 측정한 결과가 다음의 [그림 6], [그림 7]이다. 4.2.1절에서 사용한 시뮬레이션 모델을 그대로 사용하였다. 망의 크기는 [그림 6]이 50개, [그림 7]이 100개의 노드를 가지도록 지정하였고, 100개의 망에 대해 각각 10번의 무작위 순서로 Join을 수행하였다. 모든 프로토콜에 대해 경로길이를 지연시간의 척도로 사용하였다.

PIM-SM과 CBT는 QoS에 대한 고려가 전혀 없으므로 어느 정도 이상 그룹이 커지면 항상 상한값을 초과하는 경우가 발생함을 볼 수 있었다. 'QoSMIC'은 QoS에 대한 고려 때문에 보다 완만하긴 하지만, 지연시간 상한값을 초과하는 경우에 대한 대책이 없으므로 결국 그룹의 크기가 어느 정도 이상이 되면 항상 초과하는 경우가 많아짐을 볼 수 있다. DPIM의 경우 비록 망의 직경을 지연시간 상한값으로 하였으나, 항상 지연시간 상한값을 초과하지 않는 트리가 있는 C-RP의 존재를

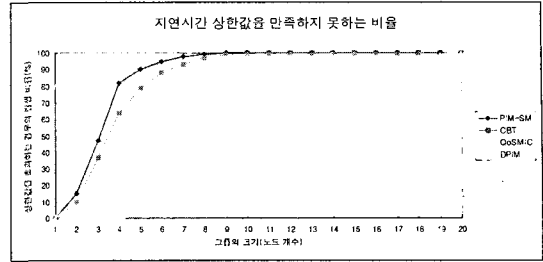


그림 7 망의 크기가 100일 때의 실험 결과

보장할 수 없으므로 위와 같이 그룹의 크기가 커짐에 따라 만족하지 못하는 경우가 발생하였다. 그러나, 그 빈도는 매우 적었다. 'QoS Extension to CBT'는 상한값을 초과하는 경우에는 Join을 거부하므로 이 실험에 포함하지 않았다.

4.2.1절에서 본 것처럼 PIM-SM이나 CBT의 경우 특정 종단이 늦게 수신하는 경우가 빈번하게 발생하므로 이 같은 상황은 서비스 기간의 상당 부분을 차지한다. 따라서 트리 재구성 기간을 제외하고는 특정 종단이 늦게 수신하는 경우가 없는 DPIM이 QoS 면에서 유리하다. 'QoSMIC'에서는 특정 종단이 늦게 수신하는 경우가 발생할 가능성은 그리 크지 않다. 따라서 구성원의 수가 작고 구성원의 분포가 지역적인 그룹의 경우 'QoSMIC'가 보다 유리할 수도 있다. 그러나 그와 같은 그룹에서는 다른 프로토콜들도 종단간 지연시간에 대한 QoS 요청을 만족할 가능성이 커지므로 'QoSMIC'에 비해 다른 프로토콜들이 크게 불리한 것은 아니다. 또한 비록 가능성은 낮지만 특정 종단이 늦게 수신하는 경우가 발생한다면 DPIM이 서비스의 공정성 면에서 보다 유리하다. 'QoS Extension to CBT'는 다양한 QoS 요청들을 처리할 수 있다는 장점은 있지만, 잠재적인 모든 구성원에게 서비스를 제공해야 하는 멀티캐스트 응용에 있어서는 항상 QoS를 만족시킬 수 있는 DPIM이 보다 유리하다.

4.3 PIM-SM에 비해 추가된 오버헤드

기존의 PIM-SM에 메시지와 상태들을 추가함에 따라 오버헤드가 증가하게 되었다. 상태의 추가에 따른 오버헤드는 C-RP가 아닌 라우터에 대한 1개의 상태만이 추가되었고, C-RP나 RP의 경우에도 몇 개에 불과하므로 확장성 면에서 끼치는 영향은 매우 적다. 반면 메시지의 추가에 따른 오버헤드를 생각하면, Join 및 Leave 과정에서 기본적으로 Join과 Leave를 PIM-SM처럼 허용하고 나서 추가적인 메시지가 필요하며 그 빈도도 그리 크지 않다. 그러나 RP 재선택 및 트리 재구성 과정에서

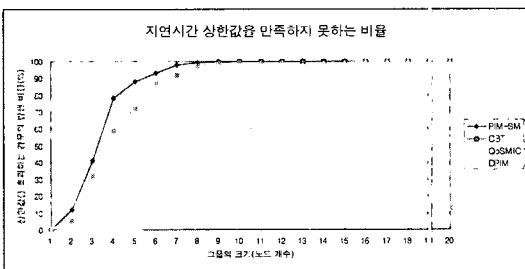


그림 6 망의 크기가 50일 때의 실험 결과

추가된 메시지는 그 양이 많고 과정이 완료되는 데도 시간이 걸린다. 따라서 RP 재선택 및 트리 재구성 과정에서 추가된 메시지의 전달에 따른 오버헤드를 줄이는 것이 필요할 것이다. 특히 DPIM의 목표 면에서 볼 때 가장 중요한 문제는 RP 재선택 및 트리 재구성에 걸리는 시간이다. 이 시간 동안 지연시간 상한값을 초과하는 종단이 존재하게 되므로, 이 시간을 줄일수록 멀티캐스트 서비스가 진행되는 전체 시간 중에서 지연시간 상한값을 만족하지 못하는 상황의 비율이 적게 되기 때문이다. 그런데 서비스의 전체 시간 및 서비스 동안 Join/Leave의 회수는 멀티캐스트 응용에 따라 달라질 것이므로 이 시간 비율은 측정하기 어렵다.

5. 결론

본 논문에서는 멀티캐스트 라우팅의 주요한 QoS 요구사항인 종단간 지연시간에 대한 QoS를 지원하는 새로운 프로토콜 DPIM을 PIM-SM을 기반으로 하여 설계하였다.

수신자간의 수신 속도의 격차는 멀티캐스트 응용에서 중요한 변수이다. 시뮬레이션에서 살펴본 것처럼 PIM-SM과 같이 QoS에 대한 고려 없이 RP를 선정하는 경우에는 지연시간 상한값을 만족하지 않는 조합이 쉽게 발견되어 잦은 트리 재구성을 요구한다. 그러나 사전에 QoS를 고려하여 RP를 선정하기가 어려우므로 본 논문에서는 필요시에 RP를 다시 선택하여 공유 트리를 재구성하는 방법을 제시하였다. 결과적으로 본 논문에서 제시하는 프로토콜 DPIM은 트리 재구성에 따르는 오버헤드를 용인하는 대신, 가능한 한 지연시간 상한값이 만족되는 RP를 갖는 공유 트리를 구성하도록 하여 주어진 지연시간 상한값이 만족되지 않는 경우를 최소화한다.

참 고 문 헌

[1] D. Waitzman, S. Deering, C. Partridge, "Distance Vector Multicast Routing Protocol," RFC 1075, November 1988.
 [2] D. Estrin, D. Farinacci, A. Helmy, V. Jacobson, L. Wei, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification, Proposed Experimental RFC, September 1996.
 [3] A. Ballardie, "Core Based Trees (CBT) Multicast Routing Architecture," RFC 2201, 1997.
 [4] D. Estrin, D. Farinacci, S. Deering, D. Thaler, A. Helmy, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification,"

RFC 2362.

[5] M. Faloutsos, A. Banerjea, and R. Pankaj, "QoS MIC: Quality of Service sensitive Multicast Internet protoCol," ACM SIGCOMM, pp.144-153, 1998.
 [6] J. Hou, H.-Y. Tyan, B. Wang, and Y.-M. Chen, "QoS Extension to CBT," Internet Draft, February 1999.
 [7] B. M. Waxman, "Routing of multipoint connections," IEEE Journal of Selected Areas in Communications, Vol.6, No.9, pp.1617-1622, December 1988.

박 세 훈

1998년 2월 26일 서울대학교 컴퓨터공학과 졸업. 2000년 2월 26일 서울대학교 컴퓨터공학과 대학원 졸업. 현재 한국 IBM 소프트웨어 연구소 재직.



안 상 현

1986년 서울대학교 컴퓨터공학과 졸업. 1988년 서울대학교 컴퓨터공학과 석사. 1993년 University of Minnesota 전산학 박사. 1988년 ~ 1989년 데이콤 사원. 1994년 3월 ~ 1998년 2월 세종대학교 전산학과 교수. 1998년 3월 ~ 현재 서울시립대학교 컴퓨터·통계학과 교수