인터넷 멀티캐스트 라우팅 프로토콜 분석

Analysis of Internet Multicast Routing Protocols

고석주(S.J. Koh) 선행표준연구팀 선임연구원
박정수(J.S. Park) 선행표준연구팀 연구원
김광수(G.S. Kim) 선행표준연구팀 선임연구원
김용진(Y.J. Kim) 선행표준연구팀 선임연구원, 팀장

본 논문은 지금까지 제안되어 온 인터넷 멀티캐스트 라우팅 프로토콜에 대하여 살펴본다. 지금까지 제안되어 온 방식들은 크게 SBT(Source Based Tree) 및 CBT(Center Based Tree) 방식으로 분류할 수 있다. SBT 방식은 하나의 송신자에서 각 수신자에 이르는 최단 경로 트리를 구하여, 데이터 전송지연 시간 측면에서 효율적인 반면에, 불필요한 제어 메시지의 수가 많이 발생하여 확장성(scalability) 문제를 갖는다. 반면에 CBT 방식은 여러 송신자가 멀티캐스트 트리를 공유하는 공유형(shared) 트리 구성 방식이며, 송신자와 수신자 모두 센터(center) 노드에 이르는 최단 경로를 통해 데이터를 주고 받는다. 이 방식은 SBT 방식에 비해 확장성 측면에서 효율적인 반면에 전송지연 시간이 길다. 향후의 관련 연구에서는 확장성과 전송지연 및 네트워크 자원의 효율적인 이용을 고려한 멀티캐스트 라우팅 기법이 개발되어야 할 것이다.

I. 서 론

본 논문은 인터넷 멀티캐스트 트리 구성 방법에 관한 것이다. 기존의 인터넷 환경은 느린 속도, 낮은 대역폭 및 전송 품질 등으로 인하여 초고속 멀티미디어 서비스의 제공이 어렵다. 최근에 개발중인 여러 가지 차세대 인터넷 기술들은 이러한 인터넷의 단점을 보완하여 화상회의 등의 멀티미디어 서비스를 효율적으로 제공하는 데에 목적을 두고 있다. 이러한 차세대 인터넷 기술들에는 자원예약 프로토콜(Resource Reservation Protocol: RSVP), 실시간 프로토콜(Real Time Protocol: RTP) 등이 있다[1].

최근에 IP 멀티캐스트 트리 구성 방식에 대한 몇 가지 기법들이 제안되어 왔다. 그 중 대표적인 방식들에는 DVMRP, MOSPF, PIM 및 CBT 등이 있다. 이에 대하여 간략히 살펴보면 다음과 같다.

먼저, DVMRP(Distance Vector Multicast Routing Protocol)[2] 기법은 멀티캐스트 데이터그램 포워딩(forwarding) 기능을 지원하기 위해 설계된 프로토콜이다. DVMRP에서는 RPB(Reverse Path Broadcasting) 알고리즘의 변형을 사용하여 소스 기반 멀티캐스트 트리를 구성한다. DVMRP는 현재 대부분의 MBONE(Multicast Backbone) 라우터들에 의해서 사용되고 있다.

한편, MOSPF(Multicast Extensions to OSPF)[3] 기법에서는 OSPF(Open Shortest Path First) 유니캐스트 프로토콜을 개선시킨 것으로, 라우터간에 link state 정보를 교환하여 네트워크 토폴로지에 대한 정보를 파악한다. 또한 MOSPF는 OS

PF version 2를 토대로 운용됨으로, 비멀티캐스트(non-multicast) 라우터와 상호호환성을 유지할 수 있다.

PIM(Protocol Independent Multicast)[4] 방식은 이름 그대로 특정 유니캐스트 라우팅 프로토콜 메커니즘에 의존하지 않는다. PIM은 dense 환경을 위한 PIM-DM 방식과 sparse 환경을 위한 PIM-SM 방식으로 구분할 수 있다. Dense 방식은 그룹 멤버가 상대적으로 밀집되어 있고 대역폭이 풍부한 환경에서 주로 사용되며, sparse 방식은 그룹 멤버가 상대적으로 다양한 지역에서 널리 분포되어 있고, 대역폭이 풍부하지 않은 환경에서 운용된다.

CBT(Core Based Tree)[5]는 가장 최근에 제안된 방식으로 PIM-SM와 운용방식이 비슷하다. 하지만 단방향(uni-directional) 트리를 구성하는 PIM-SM 방식과는 달리 양방향(bi-directional) 트리를 구성한다. CBT 방식은 확장성(scalability) 측면에서 SBT 방식에 비해 우수한 특성을 갖는다. 멀티캐스트 트래픽을 전송하는 소스의 수를 S라 하고 그룹 멤버의 수를 G라 하면, 소스 기반 멀티캐스트 트리는 O(S*G)의 규모성을 갖는다. CBT 등의 공유방식 트리는 S의 효과를 없앤 O(G)의 규모성을 갖는다. 모든 멤버들이 같은 트리를 공유하기 때문이다. 만약 비디오 게임이나 화상회의와 같은 협동형 서비스에서는 소스의 수가 많아지므로, 소스 기반 트리를 사용하는 경우 확장성 측면에서 매우 좋지 않다.

본 논문에서는 이러한 각 방식들에 대한 구체적인 운용 방식 및 장단점 등을 분석한다.

Ⅱ. DVMRP

DVMRP는 네트워크 사이에 멀티캐스트 데이터그램(datagram) 포워딩 기능을 지원하기 위해 설계된 라우팅 프로토콜이다. DVMRP는 RPB 알고리즘의 변형을 사용하여 소스 기반(source-rooted) 멀티캐스트 트리를 구성한다. DVMRP는 현재 대부분의 MBONE 라우터들에 의해서 사용되고 있다[2].

DVMRP는 RFC-1075에 의해서 처음으로 제안되었다. 초기의 스펙은 RIP(Routing Information Protocol)을 토대로 작성되었으며, TRPB (Truncated Reverse Path Broadcasting) 알고리즘을 사용하고 있다. RIP와 DVMRP의 가장 큰 차이점은, RIP는 도착지(destination)까지 이르는 최단 경로(shortest path)를 구성한다는 점이고, DVMRP는 거꾸로 출발점(source)까지 이르는 최단 경로를 구성한다는 점이다. 최근에 개발되고 있는 MBONE 라우터들은 초기의 RFC에서 약간 변형된 DVMRP 프로토콜을 구현하고 있다. 즉, 패킷 포맷(packet format), 터널 포맷(tunnel format) 및 패킷 형태(type) 등이 초기의 RFC 스펙과는 약간 다른 형태로 실제 라우터들이 개발되고 있다.

1. 물리적 인터페이스

DVMRP 라우터의 포트는 서브네트워크(sub-network)에 직접적으로 연결되는 물리적 인터페이스일 수도 있고, 혹은 또 다른 멀티캐스트 라우터에 이르는 터널(tunnel) 인터페이스일 수도 있다. 모든 인터페이스는 각 포트의 비용에 대한 데이터(metric)와 멀티캐스트 전송 범위를 제한하는 TTL(Time to Live) 기준치(threshold) 정보를 가지게 된다. 또한 터널 인터페이스의 경우 추가적으로 로컬(local) 라우터의 인터페이스에 대한 IP 어드레스와 원격(remote) 라우터 인터페이스에 대한 어드레스 정보를 포함하게 된다.

2. 기본적인 동작원리

DVMRP는 RPM(Reverse Path Multicasting) 알고리즘을 구현한다. RPM에 따르면, 소스-그룹(source-group)에 대한 첫번째 데이터그램이 기준치에 허용되는 TTL과 라우터 인터페이스 범위

에 따라 포워딩된다. 일단 초기 데이터그램은 모든 leaf 라우터들에게 전달되며, 만약 수신 라우터와 직접적으로 연결되어 있는 서브네트워크에서 그룹멤버가 존재하지 않을 경우에는, leaf 라우터가 소스를 향하여 절단(prune) 메시지를 전달하게 된다. 결과적으로 이러한 절단 메시지를 통하여 모든 그룹멤버를 연결시켜주는 하나의 소스 기반(source specific) 최단 경로 트리가 구성된다. 일정기간이 경과한 후에 이러한 절단 메시지 전달과정이 반복되며 새로운 그룹 멤버를 포함하는 새로운 트리가 구성될 수 있다.

DVMRP는 이전에 절단되었던 가지(branch) 라우터를 필요할 경우에 재빠르게 접목(graft)시켜 주는 메커니즘을 구현하고 있다. 즉, 이전에는 그룹멤버를 보유하지 않아 멀티캐스트 트리에서 제외되어, 소스 라우터에게 절단 메시지를 전달했던 라우터가, 갑자기 새로운 그룹 멤버를 보유하게 되었을 때, 신속하게 인접한 라우터에게 graft 메시지를 전달하여 멀티캐스트 트리를 구성하도록 한다. 업스트림(upstream)에 있는 라우터들은 graft 메시지를 받는 즉시 이전에 받았던 절단 메시지를 버리게 된다.

3. DVMRP 라우터 기능

서브네트워크에 하나 이상의 DVMRP 라우터가 존재하는 경우, DR(Dominant Router)가 IGM P '호스트 멤버쉽 조회(Host Membership Query)' 메시지를 주기적으로 전송하는 일을 맡게 된다. 초기화 단계에서 각 라우터는 그 자신이 DR의 역할을 수행하게 되지만, 자신보다 더 낮은(lower) IP 주소를 지니는 인접 라우터로부터 '호스트 멤버쉽 조회' 메시지를 받는 순간 DR의 역할을 종료하게 된다.

4. DVMRP 라우팅 테이블

DVMRP는 유니캐스트(unicast)가 아닌 멀티캐스트 트래픽 전송을 위해 개발되었으므로, 라우터는 유니캐스트뿐만 아니라 멀티캐스트 처리 능

<표 1> DVMRP 라우팅 테이블

Source Subnet	Subnet Mask	From Gateway	Metric	Status	TTL
128.1.0.0	255.255.0.0	128.7.5.2	3	Up	200
128.2.0.0	255.255.0.0	128.7.5.2	5	Up	150
128.3.0.0	255.255.0.0	128.6.3.1	2	Up	150
128.3.0.0	255.255.0.0	128.6.3.1	4	Up	200

력을 보유하고 있어야 한다. DVMRP는 인접한 멀티캐스트 라우터들과 주기적으로 '라우팅 테이블 갱신 메시지'를 주고 받는다. 이러한 메시지 갱신은 유니캐스트 라우팅을 지원하는 내부 게이트웨이 프로토콜(Interior Gateway Protocol)과는 독립적으로 일어난다.

DVMRP의 라우팅 테이블 샘플은 <표 1>과 같다.

RIP 등의 유니캐스트 라우팅 프로토콜과는 달리 DVMRP 라우팅 테이블은 목적지 및 Next-Hop 게이트웨이 대신에 Source Subnet과 From Gateway를 포함한다. 라우팅 테이블은 네트워크에서 멀티캐스트 서비스에 참여하는 모든 서브넷을 포함하는 소스 기반 트리를 나타내게 된다. 이러한 트리는 기본적으로 RPB 트리에 해당한다. DVMRP 라우팅 테이블에는 그룹 멤버쉽이나 절단 메시지 등은 고려하지 않는다.

DVMRP 라우팅 테이블은 구체적으로 다음과 같은 항목을 포함한다.

- 소스 서브넷: 멀티캐스트 데이터그램을 발생시키는 호스트를 포함하는 서브넷
- 서브넷 마스크: 소스 서브넷에 할당된 서브넷 마스크
- 게이트웨이: 소스 서브넷으로 가는 경로 상에 있는 라우터 중 바로 인접한 라우터
- TTL: 테이블에서 엔트리를 보유하는 시간(초 단위)

5. DVMRP 포워딩 테이블

DVMRP 라우팅 테이블은 그룹 멤버쉽 정보를 포함하지 않기 때문에, 별도의 포워딩 테이블을

<표 2> DVMRP 포워딩 테이블

Source Subnet	멀티캐스트 그룹	TTL	입력포트	출력포트
128.1.0.0	224.1.1.1	200	1 Pr	2p3p
	224.2.2.2	100	1	2p3
	224.3.3.3	250	1	2
128.2.0.0	224.1.1.1	150	2	2p3

구성하여 로컬 라우터가 각 소스-그룹 쌍(pair)에 대하여 소스 기반 최단 경로 트리의 정보를 인식하도록 한다.

DVMRP 라우터에 위치하는 전형적인 포워딩 테이블은 <표 2>와 같다.

DVMRP 포워딩 테이블은 구체적으로 다음과 같은 항목을 포함한다.

- 소스 서브넷: 멀티캐스트 데이터그램을 발생시키는 호스트를 포함하는 서브넷이다.
- 멀티캐스트 그룹: 멀티캐스트 데이터그램이 전송되어지는 Class D IP 주소이며, 하나의 서브넷에서 여러 개의 소스가 다른 멀티캐스트 그룹에 참여할 수 있다.
- 입력포트: 해당 그룹에 대한 부모(parent) 포트로, 항목에서 'Pr'의 의미는 업스트림 라우터에게 절단 메시지가 전송되었음을 의미한다.
- 출력포트: 해당 그룹에 대한 자손(child) 포트이며, 항목에서 'p'의 의미는 다운스트림 라우터로부터 절단 메시지를 받았음을 의미한다.

6. 계층적(Hierarchical) DVMRP

급속한 멀티캐스트 수요의 증가로 인하여 MBONE은 빠르게 성장하고 있다. 현재의 DVMRP 버전은 MBONE을 단일 라우팅 영역으로 보고 있다. 즉, MBONE 위에 있는 각 라우터들은 모든 서브 네트워크에 대한 상세한 라우팅 정보를 가지고 있어야 한다. 따라서 서브 네트워크의 수가 증가함에 따라 라우팅 테이블 및 주기적인 갱신 메시지의 수가 기하급수적으로 증가한다. 만약 이러한 부분에 대한 문제점이 해결되지 않는다면 결국 MBONE의 처리능력 및 메모리 용량은 고갈되고 말 것이다.

가. 계층적 멀티캐스트 라우팅의 장점

이러한 잠재적 위험요인을 극복하기 위해 계층적인 DVMRP 버전이 개발중이다. 계층적 라우팅에서 MBONE은 여러 개의 라우팅 영역으로 나뉘어진다. 각 라우팅 영역은 자신의 독자적인 멀티캐스트 라우팅 프로토콜을 수행한다. 또한 각 개별 라우팅 영역 사이의 라우팅을 위해서 또다른 라우팅 프로토콜이 수행된다. 계층적 라우팅에서는 각 라우터가 자신의 영역에 속하는 서브넷에 대한 정보만 간직하고 있으므로 요구되는 라우터 자원을 줄일 수 있다. 하지만 다른 영역에 대해서는 상세한 네트워크 토폴로지 정보 등을 전혀 알 수 없다. 개별 라우팅 영역 사이의 라우팅을 담당하는 라우팅 프로토콜은 개별 라우팅 영역 내부의 상세한 정보를 모른 채 단지 개별 영역의 상호연결성에 대한 정보를 간직하고 있다.

나. 계층적 구조

계층적 DVMRP 구조(architecture)에서 전체 네트워크 영역이 서로 겹치지 않게 개별 지역으로 구분되며, 각 개별 지역을 담당하는 라우터들은 DVMRP, MOSPF, PIM 등의 멀티캐스트 라우팅 프로토콜 중의 하나를 선택하여 멀티캐스트 라우팅 프로토콜로써 사용할 수 있다. 이를 '레벨(level) 1' 프로토콜이라 한다. 각 지역은 적어도 하나의 '경계선(boundary) 라우터'를 가지고 있어야 한다. 경계선 라우터는 다른 지역과의 멀티캐스트 라우팅 수행을 책임진다. 각 경계선 라우터들은 DVMRP를 이용하여 지역간의 멀티캐스트 라우팅을 수행한다. 따라서 DVMRP를 '레벨 2' 프로토콜로써 사용하게 된다.

하나의 지역 안에서 멀티캐스트 패킷이 생성되

면 레벨 1 프로토콜에 의해서 지역 내의 라우터들에게 전달되며, 다른 지역 라우터들에 대한 전송은 레벨 2 프로토콜을 사용하게 된다.

Ⅲ. MOSPF

OSPF(Open Shortest Path First) 라우팅 프로토콜의 버전 2는 RFC-1583에서 제안되었다. 이것은 각 라우터들에게 유니캐스트 토폴로지 정보를 전달하기 위해 설계된 IGP(Interior Gateway Protocol) 이다. OSPF는 최소한의 라우팅 프로토콜 트래픽을 사용하여 신속하게 경로를 계산해 내는 링크상태(link state) 알고리즘에 토대를 두고 있다. 효율적인 경로계산과 함께 OSPF는 계층적 라우팅, 부하균형(load balancing) 및 외부 라우팅 정보의 유입 등의 기능을 지원한다.

MOSPF는 RFC-1584에 의해서 정의된다. MOSPF 라우터는 유니캐스트 OSPF 링크-상태 라우팅 프로토콜을 사용하여 네트워크 토폴로지에 대한 현재 상태 정보를 파악한다. MOSPF는 멀티캐스트 IP 트래픽을 라우팅하는 능력을 제공함으로써 OSPF 프로토콜의 기능을 개선시킨다. 또한 MOSPF는 OSPF 버전 2를 토대로 운용됨으로, 유니캐스트 데이터 트래픽을 전송하는 경우 비멀티캐스트(non-multicast) 라우터와 상호호환성을 유지할 수 있다.

DVMRP와는 달리 MOSPF는 터널에 대한 지원기능이 없다.

1. MOSPF에 의한 인트라 영역 라우팅

인트라 영역(intra-area) 라우팅은 소스와 그룹 멤버들이 모드 같은 OSPF 영역에 위치할 때 멀티캐스트 포워딩 기능을 제공한다.

DVMRP와 유사하게 MOSPF 라우터는 각 서브넷에 연결되어 있는 멀티캐스트 그룹 멤버를 파악하기 위해 IGMP(Internet Group Management Protocol)를 사용한다. MOSPF 라우터는 라우터에 직접적으로 접속된 멤버 리스트를 관리하고 멤버들에게 멀티캐스트 데이터그램을 전송하는 기능을 수행한다.

서브넷 영역 안에서 IGMP 호스트 멤버쉽 조회는 단지 DR(Designated Router)에 의해서만 수행된다. 또한 멤버쉽 조회에 관한 보고서 수신 기능은 DR과 BDR(Backup DR)에 의해서 수행된다. MOSPF 및 OSPF 라우터가 모두 존재하는 환경에서는 MOSPF가 DR의 역할을 수행한다. 이것은 비 MOSPF 라우터들에게 "라우터 우선순위" 값을 '0'으로 설정해 줌으로써 가능하다.

DR은 OSPF 영역 안에 있는 모든 라우터들에게 그룹-멤버쉽 LSA(Link State Advertisement)를 전송(flooding)함으로써 그룹 멤버에 대한 정보를 전달한다. 라우터 LSA 및 네트워크 LSA처럼 그룹-멤버쉽 LSA도 단일 영역으로만 전송된다.

데이터그램의 최단 경로 트리는 소스 서브넷에서 각 그룹 멤버들에게 첫번째 멀티캐스트 데이터그램이 전송되면서 구성된다. 즉, 각 소스-그룹간의 최단 경로 트리는 특정 라우터가 첫번째 데이터그램을 수신하는 순간에 'on-demand' 방식으로 구성된다.

초기 데이터그램이 도착할 때에 소스 서브넷은 MOSPF 링크상태 데이터베이스에 위치하게 된다. MOSPF 링크상태 데이터베이스는 OSPF 링크상태 데이터베이스에다 그룹-멤버쉽 LSA 정보만 추가적으로 간직하게 된다. MOSPF 링크상태 데이터베이스 안에 있는 라우터 LSA 및 네트워크 LSA를 토대로, Dijkstra의 알고리즘을 사용하여 소스 기반 최단 경로 트리가 구성된다[1]. 트리가 구성된 후에 그룹-멤버쉽 LSA 정보를 이용하여 개별 멤버를 소유하지 않는 서브넷들은 트리에서 절단된다.

기본적인 MOSPF 라우팅 알고리즘은 다음과 같은 특징을 갖는다.

- 특정 멀티캐스트 데이터그램에 대하여 OSPF 영역 안에 있는 모든 라우터들은 같은 최단 경

로 트리 정보를 갖는다. 동일한 비용을 갖는 경로가 존재하는 경우에 일정한 기준에 의해서 하나가 선택된다. 참고로 OSPF에서는 동일 비용을 갖는 다중경로(multipath) 라우팅이 허용되지만 MOSPF에서는 허용되지 않는다.

- 그룹-멤버쉽 LSA를 포함하는 동기화된(synchronized) 링크상태 데이터베이스를 이용하여 RPM(Reverse Path Multicasting) 방식을 구현한다. 즉 DVMRP처럼 첫번째 데이터그램이 영역에 있는 모든 라우터에게 전송되는 것이 아니라, 멤버가 접속되어 있는 라우터에 대해서만 데이터그램 송수신이 이루어진다.
- 최단 경로 트리를 'on-demand' 방식으로 구성하기 때문에, 멤버의 그룹참여 여부를 트리 구성에 신속하게 반영할 수 있다.

각 MOSPF 라우터들은 포워딩캐시(forwarding cache)의 내용을 토대로 포워딩 결정을 수행한다. 포워딩 캐시는 각 그룹의 트리 정보와 라우터의 그룹 데이터베이스 정보를 토대로 만들어진다. 라우터가 자신이 멀티캐스트 트리에 속하게 된 것을 파악하게 되면, 업스트림 노드와 다운스트림 노드를 포함하는 포워딩 캐시 엔트리 정보가 구축된다. 일단 포워딩 캐시가 구축되면 최단 경로 트리 구축에 소요되는 네트워크 자원은 필요없게 된다. 이후로 라우터에 도착하는 멀티캐스트 데이터그램은 포워딩 캐시 정보에 따라 다운스트림 노드로 전달된다. MOSPF 포워딩 캐시는 <표 3>과 같은 정보를 포함한다.

<표 3>의 항목은 구체적으로 다음과 같은 의미를 갖는다.

- 목적지: 데이터그램이 포워딩 되어야 하는 목적지 그룹 주소
- 소스: 데이터그램의 소스 서브넷
- 업스트림: 데이터그램이 수신된 인터페이스
- 다운스트림: 데이터그램이 목적지 주소로 가기 위해 포워딩 되어야 하는 인터페이스

<표 3> MOSPF 포워딩 캐시

목적지	소스	업스트림	다운스트림	TTL
224.1.1.1	128.1.0.2	11	12, 13	5
224.1.1.1	128.4.1.2	11	12, 13	2
224.1.1.1	128.5.2.2	11	12, 13	3
224.2.2.2	128.2.0.3	12	11	7

- TTL: 데이터그램이 목적지 멤버에 도착하기까지 경유할 수 있는 최소한의 홉(hop) 개수

이러한 포워딩 캐시의 정보는 주기적으로 갱신되는 것이 아니라, 메모리 등의 시스템 자원이 이용 가능한 한 지속적으로 유지된다. 일반적으로 포워딩 캐시의 정보는 다음과 같은 경우에 변경된다. 즉, OSPF 네트워크의 토폴로지가 변화한 경우나 또는 개별 멤버의 그룹 가입여부가 변화하여 그룹-멤버쉽 LSA에 변화가 발생한 경우에 포워딩 캐시가 변경된다.

2. MOSPF에 의한 인터 영역

라우팅 인터 영역(inter-area) 라우팅은 소스와 멤버들의 일부가 서로 다른 OSPF 영역에 속해 있는 경우에 적용된다. 로컬 그룹 데이터베이스에 의하여 멀티캐스트 데이터그램이 포워딩되는 점은 인트라 영역 라우팅과 동일하지만, 그룹 멤버쉽 정보가 전달되는 방법 및 인터 영역 최단 경로 트리가 구성되는 방법 등이 인트라 영역 라우팅 방식과 다르다.

MOSPF에서는 ABR(Area Border Router)들이 인터 영역 멀티캐스트 Forwarder의 기능을 수행한다. 인터 영역 멀티캐스트 Forwarder는 영역간의 그룹 멤버쉽 정보 및 멀티캐스트 데이터그램의 포워딩 기능을 수행한다.

인터 영역 멀티캐스트 라우팅의 경우에 완전한 최단 경로 트리를 구하기가 어렵다. 각 OSPF에 대한 토폴로지 정보 및 상세한 그룹 멤버쉽 정보가 다른 OSPF 영역으로 전달되지 않기 때문이다. 이러한 단점을 극복하기 위해 와일드 카드 리시버

(wild card receiver)와 OSPF Summary Link LSA 등을 이용하여 토폴로지가 추정된다.

Ⅳ. PIM

PIM이라는 이름은 그것이 특정 유니캐스트 라우팅 프로토콜이 제공하는 메커니즘에 의존하지 않음을 보여준다. 하지만 PIM을 기반으로 하는 시스템 구현은 네트워크 토폴로지 변화에 효과적으로 적응하는 유니캐스트 라우팅 프로토콜을 필요로 한다.

PIM은 dense 환경을 위한 멀티캐스트 라우팅 프로토콜과 sparse 환경을 위한 멀티캐스트 라우팅 프로토콜간에 엄격한 구분을 한다. Dense 방식(mode)은 그룹 멤버가 상대적으로 밀집되어 있고 대역폭이 풍부한 환경에서 작동되는 프로토콜을 의미한다. Sparse 방식은 그룹 멤버가 상대적으로 인터넷 상에서 다양한 지역에서 널리 분포되어 있고, 대역폭 이용이 그리 풍부하지 않은 환경에서 작동되는 프로토콜이다. Sparse 방식이 멤버의 수가 적다는 것을 의미하지는 않는다. 단지 멤버가 널리 분포되어 있음을 의미한다.

PIM의 설계자들은 DVMRP와 MOSPF는 그룹 멤버들이 매우 밀집되어 있는 환경에 적합한 환경을 위해서 개발되었다고 주장한다. 반면에 PIM은 멤버들이 대규모 네트워크에서 널리 분포되어 있는 환경을 목표로 하여 개발되었음을 강조한다. DVMRP와 MOSPF는 그러한 환경에서 효율적인 멀티캐스트 서비스를 제공하지 못한다. DVMRP의 경우 멤버를 소유하지 않은 라우터들에게도 멀티캐스트 패킷을 전송하게 되며, MOSPF의 경우 트리 구성 및 유지를 위해 소스-그룹 멤버와 관련 없는 많은 링크들에게까지도 그룹-멤버쉽 정보가 전송된다.

1. PIM-DM

PIM은 본래 Scalable sparse 모드 트리를 제공하기 위해 제안되었지만, 기존의 DVMRP 및 MOSPF 등을 대신하는 새로운 dense 모드 프로토콜도 정의하고 있다. PIM-DM(PIM-Dense Mode)은 자원이 풍부하고 멤버들이 특정 장소에 밀집되어 있는 캠퍼스 LAN 환경을 주요 대상으로 한다.

PIM-DM은 RPM 알고리즘을 사용한다는 측면에서 DVMRP와 비슷하다. 하지만 DVMRP와 PIM-DM 사이에는 몇 가지 차이점이 있다.

- PIM-DM은 토폴로지 변화에 적응할 수 있는 기존 유니캐스트 라우팅 프로토콜을 필요로 한다. 하지만 구체적인 유니캐스트 라우팅 프로토콜 메커니즘과는 독립적이다. 반면에, DVMRP는 요구되는 유니캐스트 라우팅 정보를 계산하기 위하여 고유의 RIP 기반 프로토콜이 요구된다. MOSPF는 OSPF 링크 상태 데이터베이스 안에 포함된 정보를 사용한다. 하지만 MOSPF는 단지 OSPF 유니캐스트 라우팅 프로토콜만을 사용해야 한다.
- DVMRP는 각 그룹에 대하여 child 인터페이스를 계산하는 데에 비하여, PIM-DM은 모든 다운스트림 인터페이스에 멀티캐스트 데이터를 포워딩하며, 명백한 절단 메시지가 수신될 때까지 포워딩이 계속된다. PIM-DM은 또한 사용되는 라우팅 프로토콜과의 독립성을 위해 패킷 복제(duplication)를 허용한다.

DVMRP처럼 PIM-DM도 이미 절단된 가지 라우터에 갑자기 새로운 멤버가 접속되는 것에 대비하여 접목 메시지를 사용하여, 신속하게 절단된 라우터를 멀티캐스트 트리에 붙일 수 있도록 한다.

2. PIM-SM

PIM-SM(PIM-Sparse Mode)은 대규모 네트워크에서 다양하게 분포되어 있는 멤버들에게 멀티캐스트 서비스를 제공하기 위하여 설계되었다. 기존 멀티캐스트 라우팅 방식에서는 멀티캐스트 서비스에 참여하는 몇몇 호스트가 전체 네트워크

를 대상으로 주기적인 멀티캐스트 트래픽을 전송하는 경우 심각한 scaling 문제가 발생하게 된다. 이러한 scaling 문제를 해결하기 위하여 PIM-SM은 단지 트래픽 수신에 관심이 있는 라우터에 대해서만 멀티캐스트 트래픽을 전송하는 방식을 이용한다.

PIM-SM은 기존의 dense 모드 멀티캐스트 알고리즘과 다음 두 가지 측면에서 다르다.

- 직접적으로 멤버가 접속되어 있는 라우터는 멀티캐스트 트리에 참여하기 위해 'join' 메시지를 전송한다. 만약 이미 해당 라우터가 멀티캐스트 트리에 참여하고 있었다면, 추가적인 멀티캐스트 트래픽 수신은 이루어지지 않을 것이다. 반대로 dense 모드에서는 지속적으로 멀티캐스트 트래픽을 전송하면서 다운스트림 링크로부터 명백한 절단 메시지가 도착하지 않는 한 멀티캐스트 트래픽 송신이 지속된다. 다른 dense 모드 멀티캐스트 라우팅 방식— DVMRP, MOS PF— 에서는 기본적으로 트래픽을 포워딩한다. 반면에 sparse 모드 방식에서는 멤버의 요구가 있는 경우에만 트래픽을 포워딩하게 된다.
- PIM-SM은 RP(Rendezvous Point)를 모든 수신 라우터들이 공유한다는 측면에서 CBT (Core Based Tree) 방식과 비슷하다. 각 멀티캐스트 그룹의 초기 실행자(initiator)는 주요 RP를 선택하고 RP-lists를 이용하여 대안 RP들의 목록을 유지하고 있다. 각 멀티캐스트 그룹에 대하여 단 하나의 RP만 가동된다. 멀티캐스트 그룹에 참여를 원하는 각 수신 라우터는 RP에게 참여 메시지를 전송한다. 각 소스는 RP를 사용하여 각 멤버들에게 이르는 경로를 구한다. 이러한 모형은 각 라우터가 RP-list와 같은 일종의 상태 정보를 유지하도록 요구한다. 반대로 dense 모드에서는 첫번째 데이터 패킷이 도착한 후에 멀티캐스트 트리가 구성되므로, 이러한 상태정보가 요구되지 않는다.

LAN에 연결되어 있는 PIM 라우터가 하나 이상 존재하는 경우에는, 가장 높은 IP 주소를 가지는 라우터가 해당 LAN에 대한 DR(Designated Router)로 사용된다. DR은 IGMP 호스트 조회 메시지, RP로의 참여/절단 메시지 전송, 작동중인 RP의 상태 정보 등에 대한 정보를 관리한다.

DM과 SM 사이의 차별화를 위해, Class D 주소 중의 일부가 SM 그룹을 위해 예약되는 중이다. DR이 새로운 그룹에 대한 IGMP 메시지를 수신할 때에, DR은 해당 그룹이 RP 기반 멀티캐스트 그룹인지 아닌지를 그룹 주소를 보고 판별할 수 있다. 만약 해당 그룹이 SM 그룹인 경우, DR은 관련 그룹의 RP-list를 조회(lookup)한 다음 그룹에 대한 주요 RP를 선택한다. 만약 주요 RP에 접속불가 상황이 되면, RP-list 중의 다른 RP로 대체된다.

호스트가 그룹으로 멀티캐스트 패킷을 전송할 때에 호스트의 DR은 주요 RP로 데이터그램을 포워딩해야 한다. DR은 멀티캐스트 패킷을 PIM-SM-Register 패킷 형태로 캡슐화하여 주요 RP로 유니캐스트 전송을 수행한다. PIM-SM-Register 패킷은 RP에게 새로운 소스의 존재 사실을 알리며, 활성화된 RP는 PIM-join 메시지를 소스의 DR로 되돌려 보낸다. DR과 RP 사이에 위치한 라우터들은 PIM-SM-Register 패킷을 포워딩하면서 상태 정보를 저장해 둔다. 이를 토대로 나중에 소스에서 RP까지 전송되는 멀티캐스트 패킷을 포워딩하게 된다. 소스로부터 멀티캐스트 패킷을 전송 받은 RP는 다운스트림에 있는 멤버들에게 RP-shared 트리를 통하여 데이터그램을 전송하게 된다.

RP-shared 트리는 그룹 멤버간의 연결성은 제공하지만, 인터넷 상의 전송 경로를 최적화하지는 않는다. PIM-SM은 각 수신 라우터가 지속적으로 RP-shared 트리를 이용하든지 아니면 소스 기반 최단 경로 트리를 이용할 수 있는 선택을 허용한다. 소스 기반 최단 경로는 소스와 멤버간의 전송지연을 최소화하는 특징이 있다.

<표 4> CBT와 DVMRP의 규모성 비교

그룹 수	10			100			1,000		
그룹당 멤버 수	20			40			60		
그룹당 소스 수	10 %	50 %	100 %	10 %	50 %	100 %	10 %	50 %	100 %
DVMRP 라우터 수	20	100	200	400	2,000	4,000	6,000	30,000	60,000
CBT 라우터 수	10			100			1,000		

V. CBT

CBT는 가장 최근에 제안된 멀티캐스트 포워딩 알고리즘이다. 기존의 소스 기반 최단 경로 트리와는 달리, CBT는 그룹의 모든 멤버에 의해서 공유되는 단일 트리를 구성한다. CBT는 각 그룹마다 다른 트리가 구성된다는 점을 제외하고는 spanning tree 알고리즘과 유사하다[5].

1. CBT의 필요성

CBT 알고리즘은 확장성 측면에서 위에서 기술된 소스 기반 멀티캐스트 라우팅 방식에 비해 우수한 특성을 갖는다. 멀티캐스트 트래픽을 전송하는 소스의 수를 S라 하고 그룹 멤버의 수를 G라 하면, 소스 기반 멀티캐스트 트리는 O(S*G)의 규모성을 갖는다. CBT 등의 공유방식 트리는 S의 효과를 없앤 O(G)의 규모성을 갖는다. 모든 멤버들이 같은 트리를 공유하기 때문이다. 만약 비디오 게임이나 화상회의와 같은 협동형 서비스에서는 소스의 수가 많아지므로, 소스 기반 트리를 사용하는 경우 규모성 측면에서 매우 좋지 않다.

<표 4>는 다양한 소스의 수와 다양한 그룹 크기에 따른 CBT 및 DVMRP의 규모성을 보여준다.

<표 4>에서 우리는 CBT 방식이 라우터 자원의 이용 효율성 측면에서 DVMRP 방식보다 효율적임을 알 수 있다.

2. 운용원리

CBT 프로토콜은 단지 멀티캐스트 서비스에 직접적으로 관련되는 라우터 혹은 네트워크들만을 포함하는 공유 멀티캐스트 트리를 구성하고 유지하기 위해 제안된 방식이다.

이를 위해, 호스트는 먼저 IGMP 호스트 멤버쉽 리포트를 전송함으로써 그룹 참여 의사를 표현한다. 이러한 메시지를 받은 지역 라우터는 join_request 메시지를 코어(core) 라우터를 향하여 발송함으로써 '트리 참여 프로세스'를 시작한다.

이러한 참여 메시지는 코어 라우터에 의해서 승인되던지(Join_Ack) 아니면 이미 CBT 트리 상에 위치한 다른 라우터에 의해서 승인된다. 멤버의 참여 과정이 승인되면, 모든 멀티캐스트 데이터그램의 송수신은 코어 라우터를 경유하여 일어난다.

이제 구체적인 서비스 제공을 위한 CBT 멀티캐스트 라우팅 방식을 알아보자. 본 논문에서 고려하는 서비스에서는 각 고객이 실시간으로 차례차례 네트워크에 도착한다고 가정한다. 첫번째 고객은 세션(session)을 개최한다. 일단 네트워크에 진입한 고객은 가장 가까운 라우터에 연결된다. 이 때 이러한 라우터는 멀티캐스트 라우팅 능력을 보유하고 있다고 가정한다. MBONE 네트워크를 고려하는 경우 모든 라우터는 멀티캐스트 라우팅 기능을 수행한다고 가정한다.

인접한 멀티캐스트 라우터에 접속한 고객은 코어 라우터를 향하여 화상회의 세션에 대한 참여 메시지를 전송한다. 이러한 메시지는 유니캐스트 방식으로 고객 라우터에서 코어 라우터까지 전송되며, 코어 라우터는 충분한 링크 및 라우터 등의 네트워크 자원이 확보되어 있는 경우에 참여 승인(Join_Ack) 메시지를 고객에게 전달한다. 본 논문에서는 충분한 네트워크 자원이 확보되어 있다고

가정한다.

만약 고객이 접속한 인접 라우터가 이미 해당 세션에 대한 멀티캐스트 트리 상에 존재하는 경우에, 별도의 트리 구성 절차가 필요 없이 인접 라우터를 통하여 트래픽을 송수신하게 된다.

만약 고객이 접속한 인접 라우터가 멀티캐스트 트리에 포함되어 있지 않은 경우에, 인접 라우터와 코어 노드 사이에 최단 경로가 구해진다. 이렇게 구해진 최단 경로를 따라 도착 고객은 서비스를 이용하게 되며, 기존의 멀티캐스트 트리는 이러한 경로를 포함하도록 갱신된다.

CBT 멀티캐스트 라우팅 알고리즘은 다음과 같이 정리될 수 있다.

[CBT 멀티캐스트 라우팅 알고리즘]

① 현재 서비스중인 멀티캐스트리를 CBT라 하고, 고려하는 멀티캐스트 트리에 대한 코어 노드를 m 이라 하자.

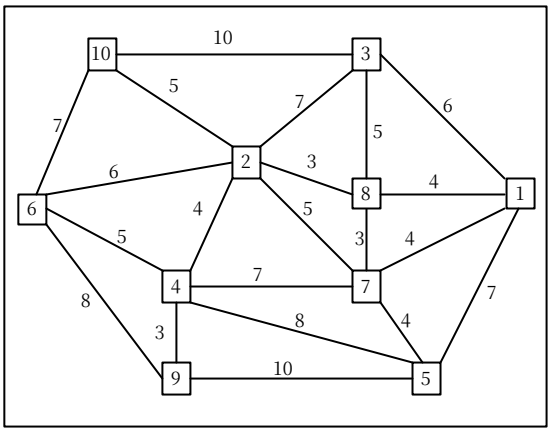
② 새로운 도착 고객이 인접한 라우터 a에 접속된다고 하자.

③ 만약 접속 노드 a가 현재 구성되어 있는 트리 CBT에 포함되어 있으면, 도착 고객은 추가적인 트리 구성의 필요 없이 멀티캐스트 서비스를 이용할 수 있다.

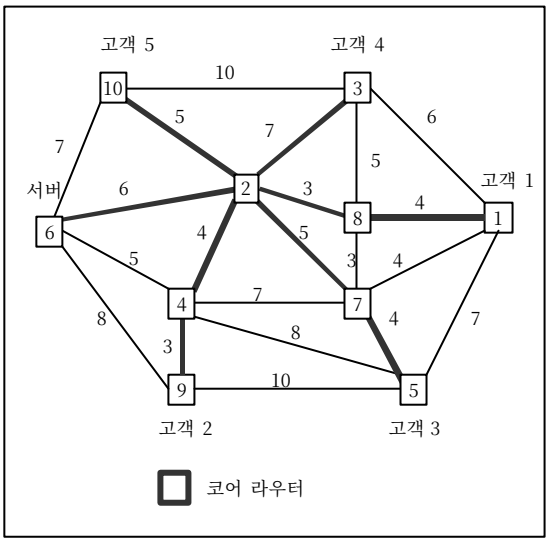
④ 그렇지 않다면, 접속 노드 a에서 코어 노드 m에 이르는 최단 경로를 구한다.

⑤ 구해진 최단 경로를 현재 CBT 트리에 첨가하고, 고객은 새로이 구해진 CBT 트리를 이용하여 멀티캐스트 데이터를 송수신한다.

(그림 1)은 10개의 라우터로 구성된 MBONE 네트워크를 보여준다. (그림 1)에서 각 링크 상의 숫자는 두 노드 사이의 링크비용을 나타낸다. 링크비용은 두 노드간의 전송지연 시간 혹은 단위 링크대역폭 사용 비용으로 볼 수 있다. 또한 이러한 전송지연 시간은 링크 상에 존재하는 hop count 수로 볼 수 있다. 본 논문에서는 이러한 링크비용이 전송지연 시간과 비례하여 주어져 있다고 가정



(그림 1) 10개의 노드를 가지는 네트워크



(그림 2) CBT 알고리즘

한다.

(그림 2)는 CBT 방식을 이용하여 멀티캐스트 트리가 구성된 예제이다. (그림 2)에서 서버는 노드 6에 위치하고, 코어 라우터는 노드 2에 위치하고 있다. 5명의 고객은 각각 노드 1, 9, 5, 3, 10을 경유하여 서비스를 요청하고 있다.

(그림 2)는 다음과 같이 CBT 멀티캐스트 라우팅 알고리즘에 의하여 구성된다.

먼저 서버는 서비스를 개최하고 코어 노드 2와 연결된다. 고객 1은 코어 노드 2에 이르는 최단 경로 1-8-2를 경유하여 세션에 참여하게 된다.

나머지 고객들도 접속 라우터들이 모두 현행 멀티캐스트 트리상에 있지 않으므로, (그림 2)와 같은 멀티캐스트 트리가 구성된다. (그림 2)에 의하면 모든 고객들은 코어 노드 2를 경유하여 멀티캐스트 트래픽을 전송하게 된다.

3. CBT 방식의 장점과 단점

CBT 멀티캐스트 라우팅 방식은 다음과 같은 장점 및 단점을 갖는다.

가. 장점

확장성 측면에서 CBT는 기존의 DVMRP 및 MOSPF 방식에 비하여 우수한 특성을 갖는다. 소스 기반 멀티캐스트 트리의 경우 각 소스-그룹 쌍에 대하여 트리 정보를 유지해야 하지만, CBT 방식의 경우 소스에 상관없이 하나의 트리를 모든 멤버가 공유하게 되므로, 멤버의 수나 소스의 수가 늘어난다고 해서 네트워크 정보량이 급격히 증가하지 않는다.

또한 CBT 방식은 모든 라우터에 주기적으로 IGMP 호스트-멤버쉽 정보를 전송하지 않으므로, 링크대역폭 등의 네트워크 자원 측면에서도 우수하다.

나. 단점

이러한 장점에도 불구하고, CBT 방식은 몇 가지 단점을 가지고 있다. (그림 2)에서 알 수 있듯이, 모든 트래픽의 대부분이 코어 노드에 집중되는 이른바 "트래픽 집중현상"을 일으킨다. 따라서 코어 노드 주변에 심각한 병목현상이 일어날 수 있다.

또한 멀티캐스트 라우팅 경로가 코어 노드를 중심으로 구해지므로, 트리 링크 비용 측면에서 비효율적이다. 즉, 구해지는 멀티캐스트 트리의 토폴로지가 트리 링크 비용 측면에서 최적의 해를 제공하지 못한다.

CBT 라우팅 방식의 또 하나의 단점은 바로 코어 라우터의 선택 문제이다. 즉, 화상회의 서비스 이용자가 네트워크의 어느 지점에서 접속할지를 모르는 상황에서, 코어 라우터를 지정해야 하는 문제는 매우 어렵고도 논란의 소지가 많은 문제이다. 실제로 IDMR 등의 인터넷 작업그룹회의에서도 이러한 코어 노드 선택 문제는 논쟁거리가 되고 있다.

VI. 결론 및 향후 연구방향

본 논문에서는 위한 멀티캐스트 라우팅 방식에 대하여 알아보았다. 기존의 IP 멀티캐스트 라우팅 알고리즘 중에서 DVMRP, MOSPF 등의 소스 기반 트리 방식은 주로 클라이언트-서버 응용서비스에 적합한 방식이며, 멤버들이 LAN 등의 좁은 네트워크에 밀집된 상황에 적합한 방식이어서, 화상회의와 같은 협동형 서비스나 대규모 인터넷에서 멤버가 골고루 분산되어 있는 상황에서는 적합하지 않다.

따라서 화상회의 등의 협동형 서비스에는 CBT 기반 PIM-SM 방식이 적합하다. 이 방식에서는 소스 기반 멀티캐스트 트리 방식과는 달리 모든 멤버 혹은 화상회의 고객이 하나의 트리를 공유하여 데이터를 송수신한다. 따라서 소스 기반 트리에 비해서 규모성 측면에서 매우 효율적이다. 하지만 CBT는 몇 가지 단점을 갖는다. 첫째 CBT 방식에서는 코어 라우터를 미리 설정하여 모든 멤버가 코어 라우터를 통하여 멀티캐스트를 수행하는데, 소위 코어 라우터 선택 문제가 발생한다. 서비스 고객이 언제 어느 라우터로 접속할지 모르는 상황에서 코어 라우터를 특정된 하나로 선택, 지정하기란 그 자체가 어려운 문제일 뿐만 아니라, 특정 라우터를 코어 노드로 지정했다고 하더라도, 고객의 도착 분포가 매우 불규칙한 상황에서는 네트워크 자원 이용 측면에서 비 효율적이다. 또한 CBT 방식의 경우 코어 라우터 주변에 트래픽이 집중되

는 트래픽 집중현상이 발생하기 쉽다. 이 경우 코어 라우터에 병목 현상이 발생할 수도 있으며 네트워크 자원 이용률이 매우 불균형 상태에 이른다. 게다가 CBT 방식은 트리 비용 및 전송지연 측면에서 최적화 구조가 아니다.

향후의 관련 연구에서는 확장성과 전송지연 및 네트워크 자원의 효율적인 이용 측면을 고려한 멀티캐스트 트리 구성 기법이 개발되어야 할 것으로 보인다. 특히 CBT 방식의 단점을 보완하여 전송지연 및 트리 링크비용을 최적화하는 멀티캐스트 트리 구조에 관한 연구가 필요하다[6].

참 고 문 헌

[1] D. Kosiur, ***IP Multicasting***, John Wiley & Sons, Inc. Carol Long, 1998.

[2] D. Waitzman ***et al.***, "Distance Vector Multicast Routing Protocol," ***RFC1075***, Nov. 1988.

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

[4] D. Estrin ***et al.***, "Protocol Independent Multicast-Sparse Mode(PIM-SM): Protocol Specification," ***RFC2117***, June 1997.

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

[6] Seok J. Koh ***et al.***, "Non-Core Based Shared Tree Architecture for IP Multicasting," ***IEE Electronics Letters***, Vol. 35, No. 11, May 1999.