

오버레이 멀티캐스트 트리의 성능향상을 위한 로컬 변환

이형옥*, 손승철*, 남지승*
*전남대학교 전자컴퓨터공학과
e-mail:narcis@freechal.com

Local Transformation for Performance Improvement on Overlay Multicast Tree

Hyung-Ok Lee*, Seung-Chul Son*, Ji-Seung Nam*
*Dept of Computer Engineering, Chonnam National University

요 약

오버레이 멀티캐스트는 하드웨어적인 인프라 구축 없이도 시스템의 자원과 네트워크 대역폭을 효율적으로 사용할 수 있는 기법이며 중간 노드의 이탈이 발생하게 될 경우 멀티캐스트 트리를 재구성한다. 그러나 빈번한 멀티캐스트 트리의 재구성은 심각한 성능 저하를 가져오게 된다. 본 논문에서는 이러한 성능저하를 보완하기 위해 각 자식 노드들로부터 소스 노드에게 주기적으로 피드백 되어오는 정보를 기반으로 트리 성능 최적화 알고리즘을 제안한다. 제안된 모델은 서비스 하는 부모노드가 트리의 성능을 저하 시키는 원인으로 판단되어질 때 수행하는 메커니즘이고, 이 메커니즘을 수행하여 성능 최적화 트리를 구성함으로써 전체적인 서비스 트리의 성능을 향상시켰다.

1. 서론

IP 멀티캐스트가 실제로 사용되기 위해서는 각 라우터에서 IP 멀티캐스트 패킷을 처리 할 수 있도록 구현되어야 한다. 그러나 라우터의 구현 및 확장, 혼잡 제어, 신뢰성 있는 전송 등의 문제로 인해 실제 인터넷에서는 적용되어지지 못하고 있는 실정이다.

IP 멀티캐스트의 대안으로 제시되어지는 오버레이 멀티캐스트(Overlay Multicast)는 기존 라우터들을 변경할 필요 없이 응용 계층에서 라우팅을 실시하며 멀티캐스트의 기능을 각 노드(end-host)에 구현한 것으로서 기존 노드가 자신이 받은 패킷을 자신의 다음 노드에게 전달해주는 방법을 사용한 어플리케이션 소프트웨어의 설치만으로 멀티캐스트를 사용할 수 있는 기법이다.

그러나 오버레이 멀티캐스트는 트리를 구성하고 있는 중간 노드가 멀티캐스트 그룹을 이탈을 하였을 경우 트리를 재구성 해야만 된다는 문제점을 가지고 있다. IP 멀티캐스트는 트리를 구성하고 있는 중간 노드들이 라우터이기 때문에 이와 같은 문제를 고려할 필요가 없지만 오버레이 멀티캐스트는 중간 노드를 포함한 모든 노드들이 언제든지 멀티캐스트 그룹을 자유롭게 이탈할 수 있는 노드(end host)로 구성이 된다는 점에서 새롭게 생겨난 문제점인 것이다.

기존의 오버레이 멀티캐스트에 관한 연구는 효율적인 멀티캐스트 트리 구성을 위해 대부분 라우팅 프로토콜의 설계 분야에서 이루어져왔다.[1-9] 반면 멀티캐스트 트리의

성능 최적화를 위해 트리를 재구성 하는데 중점을 둔 연구는 상대적으로 많이 이루어 지지 않았는데 이는 멀티캐스트 환경에서 하고자 하는 서비스의 종류에 따라 그 비중이 달라지는데도 한 원인이 있다. 멀티캐스트를 이용하여 일반 데이터를 전송하는 경우 최적의 경로를 통해 모든 데이터가 전송되는 것이 일차적인 목표이지만 실시간 방송 서비스와 같은 스트리밍 서비스를 하게 되는 경우에는 중간 노드들의 이탈에 의한 트리 구성의 잦은 변화로 발생하게 되는 성능 저하에 각 노드(end host)들이 최대한 영향을 받지 않도록 QoS (Quality of Service)를 보장 하는 것이 최우선이기 때문에 어느 정도 패킷 손실이 있을 지라도 빠른 시간 안에 성능 최적의 트리를 재구성 해야만 한다.

본 논문에서는 트리 구성 알고리즘으로 전통적인 TBCP와 mTBCP[10] 그리고 HMTP를 사용하여 멀티캐스트 트리를 각각 구성하였다.

트리를 구성한 후 노드의 잦은 이탈로 인한 성능 저하를 보완하는 로컬 변환 메커니즘을 적용하여 성능 최적의 트리를 재구성하는 오버레이 멀티캐스트 성능 최적화 알고리즘을 제안한다.

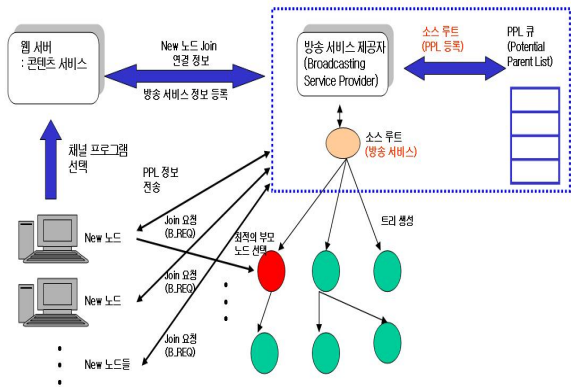
2. 본론

2.1 Overlay Multicast 모델

오버레이 멀티캐스트 개인 방송 시스템은, 효율적으로 디자인 되어 졌으며 3가지 주요 기술(오버레이 기술, 멀티

캐스트 기술, 개인방송 기술)을 이용함으로써 여러 가지 한계를 극복하도록 구현 되었다. (그림 1)은 오버레이 방송 시스템의 프레임워크를 보여주고 있다. (그림 1)에서 방송 서비스를 받고자 하는 New 노드(Join을 원하는 노드)는 웹 서버로부터 서비스 받고자 하는 채널정보를 선택하게 되며 개인 방송 시스템은 방송 서비스를 수행하기 위한 작업을 준비하게 된다. 다음 단계로 웹 서버는 해당 서비스 채널인 방송서비스 제공자(Broadcasting Service Provider=소스루트)와 New 노드의 연결을 중재하여 준다. 소스루트는 New 노드에게 가장 빠르게 연결하여 최적의 서비스를 받을 수 있는 잠재적 부모노드(Potential Parent List) 테이블 리스트를 보내준다. New 노드는 전송 받은 테이블 리스트 정보를 가지고 각각의 잠재적 부모 노드에 대한 RTT(Round-Trip-Time) 체크루틴을 수행함으로써 최적의 부모노드를 선택하게 되며, 부모노드로부터 오버레이 멀티캐스트 기반의 방송서비스를 받게 된다.

제안 시스템에서 New 노드가 실제적으로 접속(access)하여 서비스를 받는 세션에 대한 구조를 자세히 살펴보면, 먼저 New 노드는 New_B_REQ(New Client Broadcast Request) 메시지를 BSP(Broadcasting Service Provider)에 전송한다. 그러면 mTBCP 기반 오버레이 멀티캐스트 트리 관리 모듈은 BSP에 속한 모듈로서 New_B_REQ를



(그림 1) 개인 방송 시스템 프레임워크

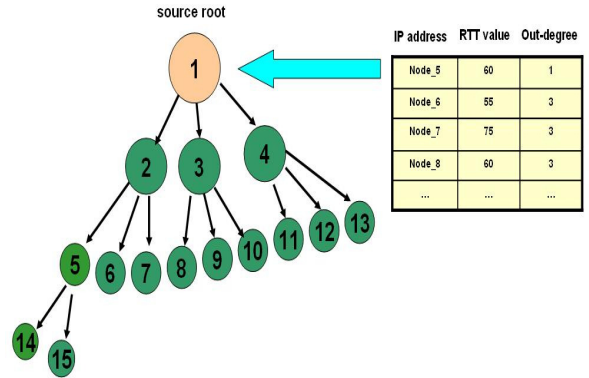
받은 BSP는 mTBCP 모듈을 이용하여 New 노드에게 서비스가 가능한 부모 노드 리스트인 PPL(Potential Parent List)을 응답 메시지와 함께 보내게 된다. PPL 정보를 받은 New 노드는 RTT 체크 루틴을 수행하여 최적의 부모노드를 선택함으로써 성공적으로 Join하게 된다. New 노드는 선택한 부모 노드에 성공적으로 Join 했다는 메시지를 BSP에 보내게 되며, 개인방송국의 서브 모듈인 mTBCP기반 오버레이 멀티캐스트 트리 관리 모듈은 PPL 정보를 업데이트 한다.

2.2 잠재적 부모 호스트(PPL) 구조

오버레이 멀티캐스트 트리에서 소스루트는 New 노드가 멀티캐스트 그룹에 가입하고자 할 때 일차적으로 접속하게 되는 랑데부 (rendezvous) 기능을 수행한다. 이때 소스

루트는 방송 서비스 그룹에 가입할 수 있는 정보를 주기 위해 잠재적 부모 호스트 목록인 PPL(Potential Parent List)를 New 노드에게 전송해 주는데 이 PPL 구조는 아래 그림과 같다.

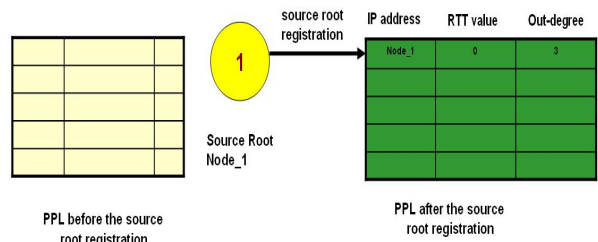
PPL은 방송 서비스를 제공해 줄 수 있는 잠재적 부모노드의 IP 주소, 상위 부모 노드를 거쳐 소스루트까지 가지는 RTT값과 여유 Out-degree로 구성된다.



(그림 2) 잠재적 부모호스트 구조

New 노드는 잠재적 부모 노드들 중 최적의 부모 노드를 선택하거나 기존의 노드가 방송 서비스 그룹에서 떠나는 경우 호스트들이 가지고 있는 정보가 변경될 때 마다 소스루트에 PPL Update 메시지를 보냄으로써 항상 최신의 트리 정보를 유지할 수 있게 한다.

아래 그림은 방송하고자 하는 소스루트가 트리를 구성할 때 제일 먼저 PPL 모듈에 자신을 등록 시키는 과정을 보여주고 있다.



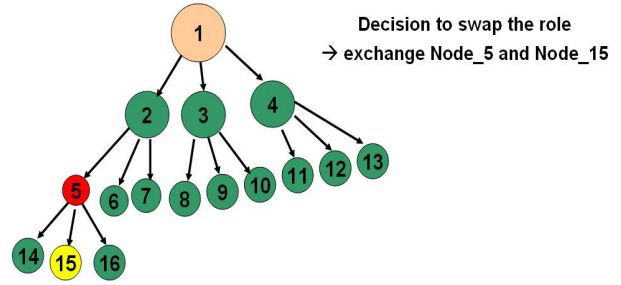
(그림 3) 트리 구성 초기화

2.3 성능최적화 계획

본 논문에서 제안하는 성능 최적화 기법인 로컬 변환 메커니즘은 피드백 되어오는 정보(RTCP-probing)에 의해 부모노드들의 성능/능력 등의 모니터링이 가능하며 부모노드로서의 능력이 현저히 떨어진다고 판단되어질 때 제안하는 메커니즘이 자동으로 수행되어진다.

최초로 구성된 트리는 가장 최적화 된 트리의 상태를 유지하고 있다. 그러나 오버레이 멀티캐스트의 특성을 반영하는 자유로운 노드의 임의 이탈은 시간이 지날수록 최적화 트리를 유지하지 못하게 되는 상황을 유발시킨다. 이는 주기적으로 보고되어지는 RTCP-probing 정보에 의해 발

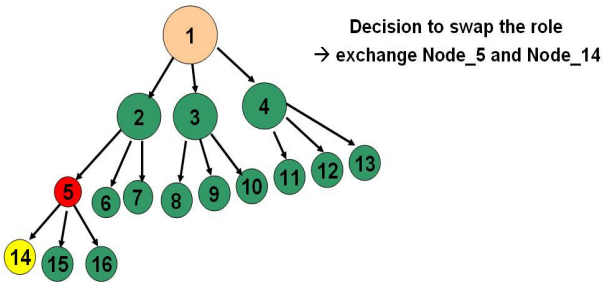
견할 수 있으며 이때, 제안하는 알고리즘을 적용하였을 때 전체적인 서비스 성능이 향상된 모습을 시뮬레이션을 통하여 결과를 확인하였다. 비효율적인 서비스 노드는 RTCP-probing에 의해 보고되는 피드백 정보에 의해 감지되며 방송 서비스를 받고 있는 자식노드들이 모두 부모 노드보다 성능이 더 좋다고 판단되어졌을 경우, 로컬 변환 모듈인 ChiPaRoS(Child-Parent-Role-Swapping) 알고리즘을 수행할 준비 상태로 진입한다. ChiPaRos 알고리즘은 아래의 2가지 상황에 대해서만 스왑 모듈을 가동한다.



(그림 5) ChiPaRoS 수행 (Step 2)

Step 1: Child-Parent-Role-Swapping 수행

RTCP-probing의 보고에 의해 효율적인 부모 노드로서의 서비스 역할을 하지 못한다고 판단이 되면 자식노드들의 상태(서비스 수행능력 및 여유 Out-degree 등)를 판단하게 된다. 자식노드들 중 최적의 성능을 가지고 있는 노드와 여유 Out-degree가 가장 많은 자식노드가 동일한 노드라면 곧바로 부모 노드와 자식 노드 사이의 스왑 모듈을 수행 한다.



(그림 4) ChiPaRoS 수행 (Step 1)

Step 2: Child-Parent-Role-Swapping 수행

자식노드들 중 최적의 성능을 가지고 있는 노드와 여유 Out-degree가 가장 많은 자식 노드가 동일한 노드가 아닐 때 여유 Out-degree가 가장 많은 자식노드와 부모 노드 사이에 스왑 모듈을 수행한다. ChiPaRoS 알고리즘 수행의 예를 (그림 3)과 (그림 4)에서 자세히 보여주고 있다.

만약 소스루트(노드 1)에서는 자식노드들인 노드 10, 노드 11, 노드 12보다 부모노드인 노드 4의 성능이 좋지 않음으로 인해서 전체적인 서비스가 안 좋은 상태로 진행되는 경우 : 최적의 성능을 가지고 있는 노드와 여유 Out-degree가 가장 많은 자식노드가 노드 10으로서 동일한 노드라면 Child-Parent-Role-Swapping 알고리즘을 수행한다.

만약 소스루트(노드 1)에서는 자식노드들인 노드 10, 노드 11, 노드 12보다 부모노드인 노드 4의 성능이 좋지 않음으로 인해서 전체적인 서비스가 안 좋은 상태로 진행되는 경우 최적의 성능을 가지고 있는 노드는 노드 10 이며 여유 Out-degree가 가장 많은 자식노드는 노드 11 으로서 동일한 노드가 아니라면 여유 Out-degree가 가장 많은 자식노드와 부모 노드 사이에 스왑이 발생하는 Child-Parent-Role-Swapping 알고리즘을 수행한다.

3. 결론 및 향후 연구과제

본 논문에서는 방송 시스템에서 방송을 서비스 하고자 하는 경우에 방송 서비스를 받고자 하는 자식 노드의 최적화 서비스 부분을 고려하기 위한 로컬 변환(Local Transformation) 모델을 제안하였다. 제안된 모델에서 부모노드가 부모노드로서의 역할을 하지 못할 경우 자식노드와의 스왑 과정을 통해 좀더 효율적인 서비스를 받게 하였다. 이러한 방법은 소스 루트의 부하를 약간 가중시키지만 서비스 측면에서는 최적의 서비스를 받을 수 있는 트리 상태를 유지하게 된다. 향후 본 논문의 제안 기법을 기존의 기법들과 비교하고 성능 최적화 작업에 적용하였을 경우 비효율적인 트리 상태에서 벗어나 효율적인 트리 상태에 접근하는 것을 확인하는 실험이 필요할 것이다.

참고문헌

[1] Y. Chu, S. G. Rao, H. Zhang, "A Case for End System Multicast," in Proceedings of ACM SIGMERTICS, June. 2000.
 [2] D. Pendarakis, S. Shi, D. Verma, M. Waldvogel, "ALMI: An Application Level Multicast Infrastructure", 3rd USENIX Symposium on Internet Technologies and Systems, Mar. 2001.
 [3] Y. Chawathe, S. McCanne, E. Brewer, "Scattercast: An Architecture for Internet Broadcast Distribution as an Infrastructure Service" PhD Thesis, University of California, Berkeley, 2000.
 [4] <http://www.icir.org/void/>
 [5] J. Jannotti, D. Gifford, K. Johnson, M. Kaashoek, J. O'Toole, "Overcast: Reliable Multicasting with an OVERlay Network", 4th Symposium on Operating Systems Design & Implementation, Oct. 2000.
 [6] H. Deshpande, M. Bawa, H. Garcia-Molina, "Streaming Live Media over Peers", Technical Report 2002-21, Stanford University, Mar. 2002.
 [7] S. Zhuang, B. Zhao, A. Joseph, R. Kartz, S. Shenker, "Bayeux: An Architecture for Scalable and Fault-Tolerant

Wide-Area Data Dissemination”, ACM NOSSDAV, June. 2001.

- [8] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, “Application-level Multicast using Content-Addressable Networks”, In Proceedings of NGC, 2001
- [9] D. Tran, K. Hua, T. Do, “ZIGZAG: An Efficient Peer-to-Peer Scheme for Media Streaming”, in proceedings of IEEE INFOCOM, Apr. 2003.
- [10] Mi-Young Kang, Omar F. Hamad, Choung-Ung Pom and Ji-Seung Nam, “FTTH-Enhanced Mini-System mTBCP-Based Overlay Construction and Evaluation”, in proceedings of SOFSEM Lecture Notes in Computer Science, Vol. 4362. Springer-Verlag, Berlin Heidelberg, pp612-623, Jan. 2007.