

다대다 응용 계층 멀티캐스트 프로토콜에서의 처리 지연 시간 영향 분석¹

김성훈⁰ 강경란 이동만
한국정보통신대학원대학교
(kimsh⁰, krkang, dlee)@icu.ac.kr

An Analysis of Impact of Processing Delay
in Many-to-many Application Layer Multicast Protocols

Sunghoon Kim⁰ Kyungran Kang Dongman Lee
Information and Communications University

요 약

IP 멀티캐스트가 현재 보편적 서비스로 사용되지 못하는 한계로 인하여 그 대안으로 응용 계층 멀티캐스트 기법들이 제안되고 있다. 기존의 응용 계층 멀티캐스트의 성능 분석 기준으로 네트워크 자원의 효율적 사용면을 고려하여 스트레치, 스트레치와 전송 트리 관리 비용 등이 사용되었다. 그러나, 응용 계층 멀티캐스트는 데이터의 전송이 참가자인 종단 시스템에 의해 이루어지고, 종단 시스템은 범용 장비로서 데이터 전달에 최적화된 성능을 갖지 못한다. 따라서, 데이터 전달을 위해 종단 시스템에서의 처리 지연이 발생하고 이로 인한 종단간 데이터 전송 지연에 영향을 끼치게 된다. 본 논문에서는 종단 시스템에서의 데이터 전달을 위한 전달 지연 시간을 모델링하고, 대표적 다대다 응용 계층 멀티캐스트 기법인 End-system Multicast와 NICE에 대하여 종단 시스템에서 발생하는 처리 지연 시간이 전체 참가자간의 데이터 전송 지연 시간에 미치는 영향을 분석한 결과를 기술한다.

1. 서론

네트워크의 보급 확대와 성능 향상으로 네트워크 게임, 오디오/비디오 컨퍼런싱, 멀티미디어 스트리밍, 데이터/컨텐츠 배포 등과 같이 네트워크 상에서 여러 사용자들 간에 데이터 교환이 이루어지는 다중 사용자 응용 프로그램들이 주목을 받고 있다. 다자간 데이터 교환을 지원하기 위한 대표적인 기술로 IP 멀티캐스트 [1]가 고려되어 왔으나, 실제 상용 인터넷 상에서 보편적인 서비스로서 사용되지 못하고 있다. 이러한 현실 상황을 고려하여, IP 멀티캐스트에 대한 대안으로 응용 계층 멀티캐스트 기법 [2]들이 제안되고 있다. 응용 계층 멀티캐스트 기법들에서는 각 참가자들 간에 IP 멀티캐스트 라우팅 트리와 유사한 데이터 전송 트리를 구성하고, 각 참가자들이 전송 트리 상의 지식 노드들에게 일대일 유니캐스트 연결을 이용하여 데이터를 전달(relay)함으로써 참가자들에게 데이터가 멀티캐스트된다.

응용 계층 멀티캐스트로는 하나의 소스로부터 다수의 수신자에게 데이터를 전달하기 위한 기법인 일대다 멀티캐스트 응용 계층 기법들 [3][4][5][6]과 참가자들 간의 상호 작용을 효율적으로 지원하기 위한 다대다 멀티캐스트 응용 계층 기법인 End system multicast [2][7]와 NICE [8]들이 개발되어왔다. 대표적 다대다 기법 중 하나인 End-system Multicast는 그룹의 규모가 작고 널리 분포된 참가자간에 메쉬(mesh) 구조의 가상 네트워크를 구성한 후, 메쉬 상에서 소스 기반의 데이터 전송 트리를 형성하며, 생성된 메쉬의 효율성은 주기적으로 검사되며 점차적으로 향상된다. NICE는 참가자들을 계층적으로 클러스터링 함으로써 참가자의 수에 대해 확장성 있는 전송 트리를 구성한다.

응용 계층 멀티캐스트 프로토콜의 성능을 평가하는데 있어, 전송 트리의 품질이라는 관점에서 다음과 같은 측정 기준들을 사용하여 왔다 [2].

- 스트레스(stress) - 물리적 링크 혹은 노드 상에 존재하게 되는 동일 패킷의 수
- 스트레치(stretch) - 유니캐스트에 의한 패킷이 지나가는 경로의 길이(홉 수)에 대한 응용 계층 멀티캐스트에 의한 경로의

길이의 비

- 전송 트리 관리 비용 - 전송 트리를 구성하고 유지하기 위해 참가자간에 교환되는 컨트롤 패킷의 양과 전송 트리 생성, 유지, 품질 향상 등을 위해 세션 참가자간에 교환하는 컨트롤 패킷의 양

높은 스트레스는 네트워크 대역폭을 많이 사용함을 뜻하고 네트워크에 혼잡을 초래할 수 있다. 스트레치는 높을수록 패킷이 크게 우회하여 수신자들에게 도달하게 됨을 의미한다. 일반적으로 스트레스가 높으면 스트레치가 낮아지게 되고, 반대로 스트레스가 낮으면 스트레치가 높아지게 되는 상충적인 효과를 갖는다.

실제로 응용 계층 멀티캐스트 세션의 참가자들은 범용 장비(종단 시스템, End-system)이므로 라우터와 같이 데이터 전달에 최적화되어 있지 않다. 그러므로, 종단 시스템에서 데이터를 전송 트리의 지식 노드에게 전달하는 과정에서 발생하는 처리 지연 시간이 전체 응용 계층 멀티캐스트 세션의 성능에 영향을 줄 수 있다. 특히, 성능이 크게 낮은 종단 시스템이 전송 트리의 중앙에 위치하게 되면, 해당 시스템에서 발생하는 처리 지연 때문에 전송 트리 상에서 해당 시스템의 자손에 해당하는 참가자들에게 데이터가 전달되기까지의 지연 시간이 크게 증가될 수 있다. 해당 시스템의 지식 노드의 수가 증가할수록 그리고 데이터의 전송률이 클수록 그 전달 지연 시간은 더 커진다. 그러므로, 비디오 컨퍼런싱이나 실시간 공유 문서 작업 등과 같은 실시간성이 중요한 응용에서는 전송 트리를 구성하는데 있어 이 같은 종단 시스템에서 발생하는 처리 지연 시간에 대한 고려가 포함되어야 한다.

본 논문에서는 기존의 대표적인 다대다 응용 계층 멀티캐스트 프로토콜인 End-system Multicast (ESM)와 NICE에 대해, 종단 시스템에서의 처리 지연 시간이라는 관점에서 데이터 전송 트리의 품질을 평가한다. 종단 시스템에서의 멀티캐스트 라우팅을 위한 처리 지연 시간은 대량의 입력을 받아서 처리하는 $M^X/M/1$ Queue 모델로서 설명이 가능하다 [9]. Queue 모델과 시뮬레이션을 통해서 종단 시스템에서의 데이터 전달을 위한 처리 지연을 관찰하였다. 실험 결과에 의하면, NICE가 스트레치나 전송 트리의 안정성이라는 관점에서 참

¹ 본 연구는 과학기술부가 주관하는 국가지정연구실사업 (NRL: M1-0104-00-0130)의 지원으로 수행되었음

가자의 수에 대해 확장성 있다고 평가되어 왔지만, 전체 클러스터의 중앙에 위치하는 참가자에게 데이터 전달 부담이 집중됨으로써, 해당 참가자로부터 발생하는 처리 지연 시간이 전체 세션의 전달 지연 시간에 중요한 영향을 끼친다는 것을 확인하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 ESM과 NICE를 좀 더 자세하게 설명하고, 3장에서 중단 시스템이 생성하는 처리 지연 시간 모델에 대해 기술한다. 4장에서는 ESM과 NICE에 대해 중단 시스템에서 발생하는 처리 지연 시간이 전체 참가자들의 데이터 전달 지연 시간에 미치는 영향에 대한 시험 결과를 기술한다. 마지막으로 5장에서 본 연구의 결론과 앞으로의 연구 방향을 제시한다.

2. 다대다 세션을 위한 응용 계층 멀티캐스트 프로토콜

End-system Multicast (ESM) [2]은 기본적으로 그룹의 규모가 작고 사용자가 널리 분포되어 있는 멀티캐스트 세션을 위해 설계되었다. 멀티캐스트 세션에 참여하는 중단 시스템들이 메쉬 구조의 가상 네트워크를 구성한 후, 메쉬 상에서 소스 기반의 데이터 전송 트리를 형성하여 멀티캐스팅을 수행한다. 각 중단 시스템은 각 소스에 따라 다음 전송 노드를 결정하는 라우팅 테이블을 유지하고, 상호 교환함으로써 소스 기반의 데이터 전송 트리를 형성할 수 있다. 그러나, 소스 기반 데이터 전송 트리의 구성을 위해서 교환되어야 하는 데이터의 양이 많고 각 소스별로 중단 시스템이 가져야 하는 라우팅을 위한 테이블의 크기 또한 커서 일반적으로 확장성이 없다고 평가된다.

NICE[8]는 멀티캐스트 세션에 참여하는 중단 시스템을 상호간의 거리를 이용하여 클러스터링하고, 각 클러스터에서 대표를 선출하여 이들 대표간에 또 새로운 클러스터 계층을 구성한다. 이러한 방식을 반복하여 다계층구조의 가상 네트워크를 형성한다. 가상 네트워크 상에서 각 중단 시스템은 자신이 속하는 클러스터 내의 중단 시스템만을 인지한다. 따라서, 데이터 전송 트리를 위한 가상 네트워크의 유지 관리에 있어서 확장성 있다고 볼 수 있다. 데이터 전송은 자신이 속하는 모든 계층에서 같은 클러스터에 속한 중단 시스템들에게 데이터를 전송한다. 따라서, 상단 계층으로 갈수록 데이터를 전송해야 하는 대상이 증가하게 되고, 트래픽이 집중되어 병목 현상이 발생할 수 있으며, 중단간 데이터 전달 지연 시간을 크게 증가시킬 수 있다.

3. 중단 시스템에서의 처리 지연 시간 모델

응용계층 멀티캐스트 세션에 참여하는 중단 시스템의 데이터 전달을 위한 큐 모델은 그림 1과 같다. 개별 중단 시스템은 수신된 데이터 패킷을 저장하기 위한 Incoming Queue와 전송 트리 상에서 자식 노드들에게 전달하기 위한 데이터 패킷을 저장하는 Outgoing queue를 갖는다. 각 중단 시스템은 λ 의 속도로 수신한 패킷을 저장하고, 응용 계층 멀티캐스트를 위한 데이터 전송 트리 상의 k 개의 자식 노드에게 전달하기 위해서 k 개의 Output Queue에 복사하여 저장한다. 데이터 패킷의 처리 속도를 μ 로 표시한다.

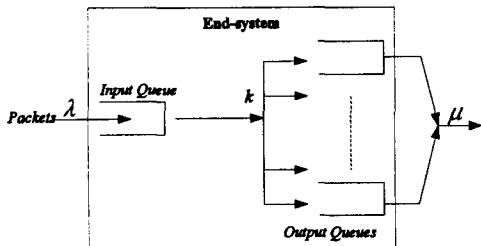


그림 1. 중단 시스템에서의 큐 모델

그림 1로부터 데이터 전송 트리 상에서 k 개의 자식 노드를 가지는 중단 시스템에서의 패킷 처리 상태를 나타내는 상태 변이 다이어그램(state transition diagram)을 그림 2와 같이 유도할 수 있다. 다이어그램 내의 상태는 큐에 남아 있는 패킷의 수를 의미한다.

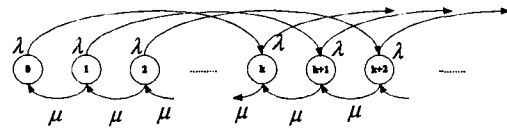


그림 2. 상태 변이 다이어그램

$M^k/M/1$ Queue 모델 [9]에 의해서 패킷이 시스템에서 기다리는 시간(W)은 다음과 같이 유도된다. 먼저 트래픽 밀도(ρ)는 다음과 같이 표현된다.

$$\rho = \lambda \cdot k / \mu \dots\dots\dots (1)$$

시스템에서 처리를 기다리는 패킷의 수(L)와 하나의 패킷이 시스템 내에서 처리되기까지 기다리는 시간(W)은 다음과 같다.

$$L = \frac{k+1}{2} \cdot \frac{\rho}{1-\rho} \dots\dots\dots (2)$$

$$W = \frac{L}{\lambda \cdot k} \dots\dots\dots (3)$$

패킷이 시스템에서 기다리는 시간(W)은 중단 시스템에서 데이터 전달을 위해 발생하는 처리 지연 시간을 의미하게 된다. 수식 (2)를 이용하여 수식 (3)은 다음과 같이 쓸 수 있다.

$$W = \frac{(k+1)}{2 \cdot (\mu - \lambda \cdot k)} \dots\dots\dots (4)$$

중단 시스템의 처리 지연은 해당 중단 시스템이 가지는 데이터 전송 트리 상에서의 자식 노드의 수(k)에 비례하여 증가하고, 또한, 데이터의 도착하는 속도(λ)에 비례하여 증가하게 되며, 중단 시스템의 처리 성능(μ)에 반비례한다. 다대다 응용 계층 멀티캐스트의 경우, 모든 참가자가 수신자인 동시에 데이터의 소스가 될 수 있으므로, 참가자의 수가 늘어날수록 각 중단 시스템에서의 데이터 도착률이 비례하여 증가하게 되고, 이로 인해 개별 중단 시스템에서의 전송 지연 시간이 비례하여 증가하게 된다. 특히, 신규로 추가되는 참가자가 전송 트리의 자식 노드가 되는 경우, 신규 참가자로 인한 처리 지연 시간 증가 효과의 비율이 더욱 커진다.

그림 3은 네트워크 시뮬레이터 NS2 [10]를 이용한 시뮬레이션을 통하여 측정된 중단 시스템의 처리 지연시간과 수식 (4)를 이용하여 계산한 처리 지연 시간을 보이고 있다. 그림 3에서 보는 바와 같이, 두 개의 결과가 거의 같음을 알 수 있다. 시뮬레이션에서 최초 데이터 전송 트리를 15개의 중단 시스템으로 구성하고, 측정하고자 하는 중단 시스템이 전송 트리에서 두 개의 자식 노드를 가지도록 하였다. 신규 참가자가 측정하고자 하는 중단 시스템의 자식 노드가 되도록 하고 전체 참가자의 수를 20까지 증가시켰다. 그리고, λ 는 1(패킷/초)이며, μ 는 200(패킷/초)이다.

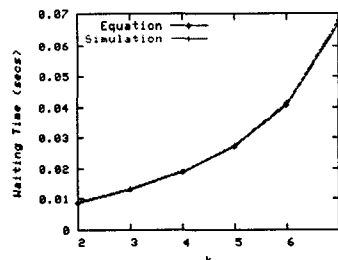


그림 3. 중단 시스템의 처리 지연 시간

4. 성능 분석

End-system Multicast (ESM)과 NICE에 대해, 중단 시스템에서의 처리 지연 시간이라는 관점에서 데이터 전송 트리의 품질을 평가한다.

네트워크 시뮬레이터인 NS2 [10]를 이용하여 성능 분석이 수행되었고, 사용된 컴퓨터는 펜티엄4 2Ghz의 CPU와 Redhat? Linux를 운영체제로 하고 있다. 사용된 네트워크 토폴로지는 GT-ITM [11]을 이용하여 생성되었고, 100개의 라우터와 각 라우터에 하나의 종단 시스템 (총 100개의 종단 시스템)이 연결되어 있다. 라우터간의 링크 지연은 5ms이고, 라우터와 종단 시스템 사이의 링크 지연은 2ms이다. 멀티캐스트에 참여하는 종단 시스템의 개수는 8, 16, 32, 64개로 실험하였으며, 토폴로지 상에서 세션에 참가하는 종단 시스템의 위치는 임의로 (randomly) 선택하였다. ESM의 경우 메쉬 상에서 노드가 가질 수 있는 최대 논리적 링크의 수를 8로 제한하였고 라우팅 정보의 교환을 5초마다 수행한다. NICE는 한 클러스터의 크기를 결정하는 인자를 3으로 설정하였다. 즉, 클러스터의 최대 크기를 8로 제한한다.

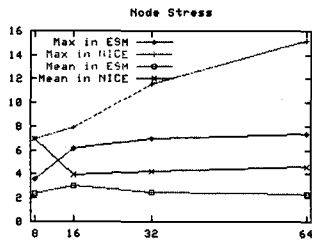


그림 4. 참가자의 스트레스

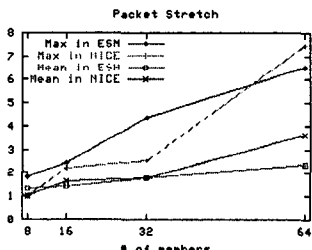


그림 5. 패킷 스트레치

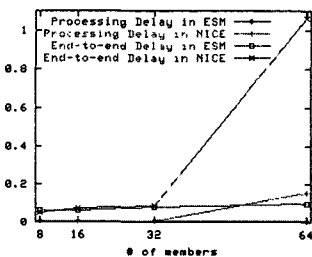


그림 6. 각 참가자에서의 처리 지연 및 종단간 전송 지연

그림 4, 그림 5, 그림 6은 시뮬레이션 결과를 나타낸다. 그림 4는 참가자가 갖는 평균과 최대 스트레스를 나타내고, 그림 5는 전송 트리의 평균과 최대 스트레치를 나타낸다. 그림 4에서 보듯이 End-system Multicast (ESM)의 경우 최대 스트레스가 제한값인 8보다 크지 않다. 반면, NICE는 한 세션의 최대 스트레스의 평균이 15에 가깝다. 왜냐하면 상위 계층의 종단 시스템은 하위 모든 계층에도 속하게 되기 때문에 전송 트리 상에서 자식 노드의 수가 많아지기 때문이다. 스트레스가 큰 NICE가 일반적으로 작은 스트레치를 갖지만, 그림 5에서 보이는 바와 같이 64개의 경우 그렇지 않은 결과를 보이고 있다. 이는 참가자 수가 많아짐에 따라 계층이 더 많이 만들어지고, 클러스터링의 규모에 제한이 있으므로 거리가 가까운 종단 시

템들이 서로 다른 클러스터에 속하게 되어 가까이 이웃하고 있음에도 불구하고 클러스터의 리더를 통해서 데이터가 전달되기 때문에 스트레치가 크게 나타날 수 있다.

그림 6은 각 종단 시스템에서 측정된, 데이터 전달을 위해서 소요되는 처리 지연 시간의 평균과 최대값, 그리고 소스로부터 각 수신자까지의 전송 지연 시간의 평균과 최대값을 나타낸다. 참가자 수가 32인 경우까지는 두 프로토콜이 거의 비슷한 성능을 나타내지만, 64가 되면 NICE의 경우 처리 지연 시간이 크게 증가한다. NICE는 스트레스에 상위 한계가 없고, 상위 계층에 속한 노드들에서 증가된 스트레스로 인한 처리 지연 시간 증가가 전체 전달 지연을 증가시키는 데 기여를 하게 된 것이다.

5. 결론

본 논문에서는 데이터 전달을 위한 처리 지연을 중심으로 다대다 응용 계층 멀티캐스트 프로토콜의 성능 분석을 수행하였다. 기존의 네트워크 자원의 사용에서의 효율성뿐 아니라 멀티캐스트 라우팅에 참여하는 종단 시스템의 특징을 고려하여 볼 때, 데이터 전달을 위한 처리 지연 또한 중요한 성능 분석 기준이 된다. 시뮬레이션을 통해서 종단 시스템에서의 처리 지연은 $M^X/M/1$ Queue 모델을 따르는 것을 확인하였다. 또한, 처리 지연은 종단 시스템이 수행하는 데이터 전달 (relay)의 양 (전송 트리의 자식 노드의 수)과 데이터의 수신률에 비례하기 때문에 다대다 응용 계층 멀티캐스트에서 더욱 더 고려하여야 할 요소이다.

End-system Multicast에 비해 확장성이 있는 가상네트워크 구조를 구성한다고 평가되는 NICE의 경우에도, 참가자의 수가 증가할수록 상위 계층의 종단 시스템의 스트레스가 급격하게 증가하게 되어 전체 세션에서 종단 시스템간의 전송 지연을 증가시키는 것을 볼 수 있다. 도리어, 평면 구조의 End-system Multicast이 각 종단 시스템에서의 최대 스트레스를 제한함으로써 NICE와 같은 급격한 처리 지연 시간의 증가를 유발하지 않았다. 즉, NICE에서 사용하는 클러스터링 구조에 최대 스트레스를 제한하는 기법을 결합함으로써, 확장성이 있는 가상 네트워크 구조뿐만 아니라 전송 지연 시간에서의 확장성을 동시에 이룰 수 있을 것이다.

앞으로 본 논문에서 관찰한 처리 지연 시간이 응용 계층 멀티캐스트 세션의 성능에 미치는 영향을 기반으로 하여, 스트레치, 스트레스 및 전송 지연 시간에 있어서도 확장성 있는 응용 계층 멀티캐스트 프로토콜의 개발을 진행할 것이다.

참고문헌

- [1] S. Deering, "Host Extensions for IP Multicasting," Internet RFC 1112, Aug. 1989.
- [2] Y. Chu, S. G. Rao, and H. Zhang, "A Case for End System Multicast," ACM SIGMETRICS'00, pp. 1-12, Jun. 2000.
- [3] L. Mathy, R. Canonico, and D. Hutchison, "An Overlay Tree Building Control Protocol," NGC 2001, Nov. 2001.
- [4] Y. Chawathe, S. McCanne, and Eric A. Brewer, "RMX: Reliable Multicast for Heterogeneous Networks," IEEE INFOCOM'00, pp. 795-804, Jun. 2000.
- [5] J. Jannotti, D. K. Gifford, K. L. Johnson, M. Frans Kaashoek, and J. W. O'Toole Jr., "Overcast: Reliable Multicasting with an Overlay Network," USENIX Symposium on Operating Systems Design and Implementation, pp. 197-212, Oct. 2000.
- [6] H. Deshpande, M. Bawa, and H. Garcia-Molina, "Transience of Peers and Streaming Media," HotNets-I (2002), Oct. 2002.
- [7] Y. Chu, S. G. Rao, S. Seshan and H. Zhang, "Enabling Conferencing Applications on the Internet using an Overlay Multicast Architecture," ACM SIGCOMM'01, pp. 55-67, Aug. 2001.
- [8] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, "Scalable Application Layer Multicast," ACM SIGCOMM'02, pp. 205-217, Aug. 2002.
- [9] D. Gross and C. M. Harris, "Fundamentals of Queueing Theory," 3rd edition, John Wiley & Sons, Inc., 1998, pp 116-122.
- [10] NS-2, Network Simulator 2, <http://www.isi.edu/nsnam/ns/>
- [11] GT-ITM, Georgia Tech Internetwork Topology Models, <http://www.cc.gatech.edu/projects/gtitm/>