

캐싱 기반의 네트워크를 위한 콘텐츠 전송경로 선택 전략

한진영, 권태경, 최양희
서울대학교

iyhan@mmlab.snu.ac.kr, tkkwon@snu.ac.kr, yhchoi@snu.ac.kr

Strategy of Content Delivery Path Selection for Caching-based Network

Jinyoung Han, Taekyoung Kwon, Yanghee Choi
Seoul National University

요약

콘텐츠 전송은 콘텐츠의 위치를 찾는 단계 (Resolution)와 찾은 위치로부터 실제로 콘텐츠를 전송하는 단계 (Delivery)로 구성된다. 기존의 연구에서 Delivery 경로는 IP 레벨의 효율적인 라우팅 메커니즘을 통해 결정된다고 가정하고 있다. 하지만 캐싱 기능이 있는 네트워크에서는 단순히 자신에게 최적의 경로로 콘텐츠를 전송하는 것보다 캐싱 기능을 활용할 수 있는 방향으로 전송 경로를 선택하는 것이 필요하다. 본 논문에서는 캐싱이 있는 네트워크에서 콘텐츠 실제 전송 경로를 어떻게 정하는 것이 좋은지에 대해 논의한다. 또한 현실적인 실험 모델에 기반하여 Resolution 경로를 이용하는 콘텐츠 Delivery 경로 선택 전략이 전체 네트워크 성능을 향상시킨다는 것을 검증한다.

I. 서론

인터넷은 데이터나 서비스 등의 콘텐츠를 전송하는데 주로 사용된다. 예를 들어 P2P 어플리케이션이나 웹 서비스를 통해 사용자들은 필요한 콘텐츠를 주고 받는다.

콘텐츠 전송은 콘텐츠의 위치를 찾는 단계 (Resolution)와 찾은 위치로부터 실제로 콘텐츠를 전송하는 단계 (Delivery)로 구성된다. 각 단계를 위한 연구는 각각 많이 진행되어 왔다. 대표적인 Resolution 연구의 예로 현재 인터넷의 가장 중요한 시스템 중 하나인 DNS와 P2P 어플리케이션의 Distributed Hash Table (DHT) 구조 등이 있다. 그리고 최근 많이 사용되는 P2P 어플리케이션 중 하나인 BitTorrent의 전송 방식 등이 Delivery 연구의 예라고 할 수 있다. 현재까지 대부분의 Delivery 관련 연구는 콘텐츠를 갖고 있는 호스트로부터 콘텐츠를 요구한 호스트까지 IP 레벨의 효율적인 라우팅 메커니즘을 통해 Delivery 경로가 정해진다고 가정하고 있다. 하지만 콘텐츠 전송 능력을 향상시키기 위한 캐싱 기능이 추가되면서, 단순히 자신에게 최적의 경로로 콘텐츠를 전송하는 것보다 캐싱 기능을 잘 활용할 수 있는 경로를 선택하는 것이 전체 시스템 성능에 도움이 될 수 있다. 특별히 최근 많이 연구되고 있는 콘텐츠 중심의 네트워크[1]와 같은 경우에 캐싱 기능이 있는 콘텐츠 전송 아키텍처를 가정하기 때문에 Delivery 경로 선택을 어떻게 하는 것이 전체 성능에 좋은지 고찰하는 것이 필요하다.

본 논문에서는 콘텐츠 전송 성능의 지배적인 역할을 하는 Delivery를 하는 데 있어서 경로를 어떻게 선택할지에 대해 논의한다. 특별히 Resolution 구조가 대표적인 아키텍처 중 하나인 DHT로 구성되었을 때, Delivery 경로를 어떻게 정하는 것이 전체 성능 향상에 도움이 되는지 연구한다. 또한 본 논문에서는 현실적인 실험 모델에 기반하여 이러한 전략들의 성능을 검증한다.

II. 콘텐츠 Delivery 경로 선택 전략

대부분의 연구에서 콘텐츠 Delivery 경로는 Resolution으로 콘텐츠를 갖고 있는 호스트를 찾은 후에

IP 레벨의 라우팅 메커니즘을 통해 결정되는 것으로 가정하고 있다. 이렇게 선택된 경로가 콘텐츠를 받고자 하는 호스트 자체에게는 최적의 경로일 수 있지만, 캐싱 기능이 있는 네트워크에서는 전체 성능을 향상시킬 수 있는 콘텐츠 Delivery 경로 선택 전략이 필요하다. 다음은 본 논문에서 제시하는 Delivery 경로 선택 전략이다.

- host-to-host: IP 레벨의 라우팅 메커니즘으로 결정된 최적의 경로이다. 줄여서 h2h라고 부른다.
- h2h (cache): Delivery 경로는 h2h로 하되 Resolution 아키텍처에 캐싱 기능이 있어서 콘텐츠를 제공하고 받는 호스트가 포함된 곳에서 캐싱된다.
- resolution (cache): Delivery 경로를 Resolution을 해운 경로를 따라 가도록 해서 Resolution 아키텍처에 캐싱이 되도록 한다. 이렇게 할 경우 자신의 Delivery 시간은 처음에는 길어질 수 있지만 다른 사용자들이 Resolution 아키텍처를 통해 콘텐츠를 전송 받고자 할 때 캐시 hit이 일어날 확률이 높아 시스템 전체적인 성능을 향상시킬 수 있다.
- h2h+resolution (cache): 실제 Delivery는 최적 경로인 h2h로 하고, Resolution 아키텍처에 캐싱을 하기 위한 목적으로 Resolution 경로로도 콘텐츠를 전송한다.

III. 실험 결과

A. 실험 설정

본 논문에서 네트워크 계층의 delay나 packet loss 등을 고려해서 보다 정확한 결과를 얻기 위해 Network Simulator v2 (ns-2)를 이용하여 실험하였다. 보다 현실적인 실험 환경을 사용하기 위해서 Heckmann이 제시한 GT-ITM 기반의 네트워크 토폴로지를 emulate 방법[2]을 사용하였다. 이 방법을 이용하여 312개의 노드로 이루어진 실제 AT&T의 토폴로지를 emulate 하였고, 그 중에 100개의 노드를

선택해서 오버레이로 DHT 기반의 Resolution 아키텍처를 형성하게 하였다. 그리고 네트워크 전체에 4000 개의 콘텐츠가 있고, 각 Resolution 아키텍처 노드들이 균등한 숫자의 콘텐츠를 관리하고 있다고 가정하였다. 이때, 각 콘텐츠의 크기는 대학 캠퍼스에서 수집한 웹 콘텐츠의 크기 분포를 반영해서 보다 실험의 현실성을 높였다. 한편, 호스트는 초당 0.5 번의 비율로 콘텐츠를 요청하는데, 이 때 개별 콘텐츠의 인기도가 Zipf 분포를 따른다는 것이 밝혀졌기 때문에 [3] 콘텐츠의 요청 분포를 Zipf 분포를 따르도록 하였다. 캐싱의 경우 총 2 기가의 용량까지 저장하도록 하였고, 용량이 모두 사용되면 기존의 효과적이라고 알려진 Least Recently/Frequently Used 교체전략 [4]을 사용하도록 하였다. 총 시뮬레이션은 안정상태에 이른 36000 초부터 120000 초까지 진행하였다.

B. Delivery 경로 선택 전략의 성능 비교

본 절에서는 앞장에서 설명한 콘텐츠 Delivery 경로 선택 전략에 따른 성능을 비교하였다. 캐싱 기능이 없는 경우는 있는 경우와 비교했을 때 성능 차이가 많이 나기 때문에 결과에서 제외하였다.

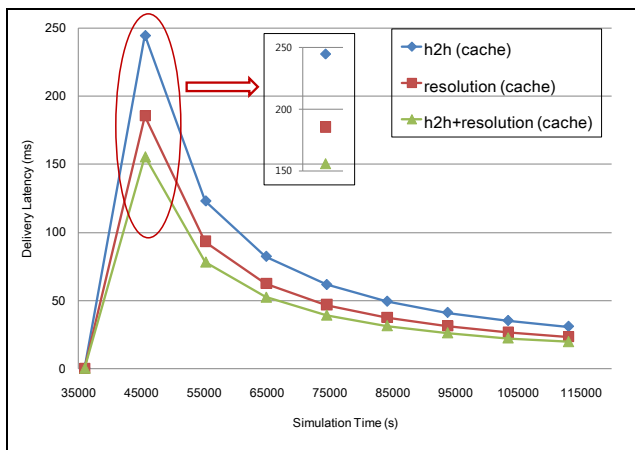


그림 1 콘텐츠 Delivery 경로 선택 전략에 따른 전송시간

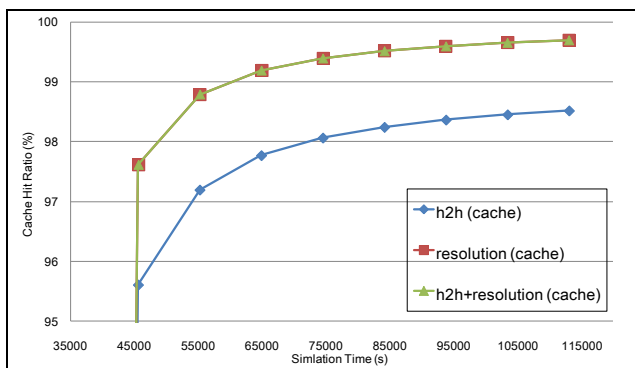


그림 2 콘텐츠 Delivery 경로 선택 전략 별 캐시 hit 비율

그림 1 은 콘텐츠 Delivery 선택 경로 전략에 따른 전송시간의 평균 결과이다. 전체적으로 Simulation 이 진행 될수록 캐시 hit 되는 경우가 많아져서 전송 시간이 줄어드는 것을 알 수 있었다. 기존의 보편적인 경로 선택 방식에 캐싱 기능이 추가된 h2h (cache)에 비해서 DHT Resolution 아키텍처를 따라 콘텐츠를 전송한 경우, 비록 h2h 보다 해당 콘텐츠의 실제 전송 시간은 더 길어지지만, 전체적으로 볼 때 캐시 hit 되는 경우가 더 많아서 전체 전송시간은 오히려 줄어드는 결과를 나타내었다. 즉 전송에 참여하는 커뮤니티의 경우

전체적으로 이득을 본 것이라고 할 수 있다. 전송은 h2h 방향으로 하고 콘텐츠의 캐싱을 위해 Resolution 방향으로 콘텐츠를 한번 더 보낸 h2h+resolution (cache)의 경우 캐시 hit 되는 효과와 실제 콘텐츠 전송을 최적으로 보내는 효과가 상승작용을 나타내어 가장 좋은 성능을 나타내었다. 그림 2 에서 각 전송 경로 선택 전략들의 캐시 hit 비율 중에 Resolution 방향으로 콘텐츠를 보낸 경우의 hit 비율이 더 높은 것을 볼 수 있고 이에 따라 성능 차이가 나게 된 것을 알 수 있다. 특별히 시뮬레이션 초반에 성능 차이가 두드러지게 많이 나는데, 그 이유는 h2h (cache)에 비해 Resolution 아키텍처로 콘텐츠를 전송한 경우에 전체 네트워크에 콘텐츠가 빨리 dissemination 되기 때문이다. 일반적으로 콘텐츠 전송을 빠르게 하기 위해서 네트워크에 해당 콘텐츠의 복제나 캐시를 잘 dissemination 하는 것이 필요한데, 콘텐츠 Delivery 시에 경로를 Resolution 경로로 보내게 되면 효과적으로 콘텐츠를 dissemination 하게 되어 전체 네트워크의 성능을 향상시킬 수 있게 됨을 알 수 있다.

IV. 결론

본 논문에서는 캐싱 기능이 있는 네트워크에서 콘텐츠의 실제 전송을 어떤 경로를 하는 것이 좋은지에 대해 논의하였다. 기존의 최적 경로로 전송하는 경우 (h2h (cache))보다 Resolution 경로 (resolution (cache), h2h+resolution (cache))를 따라 전송할 때 캐싱 기능을 효과적으로 이용할 수 있어서 전체 네트워크의 성능이 향상됨을 보였다. 본 논문에서 실험적으로 검증된 내용을 보다 철저히 검증하기 위해 수학적인 모델을 바탕으로 후속 연구를 진행할 계획이다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT 산업원천기술개발사업의 일환으로 수행하였음. [2007-F-038-03, 미래 인터넷 핵심기술 연구] 또한, 본 연구는 기초기술연구회의 NAP 과제 지원으로 수행되었음. 이 연구를 위해 연구장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사 드립니다.

참고 문헌

- [1] T. Koponen, M. Chawla, B.-G. Chun, A. Armolinskiy, K. H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," In ACM SIGCOMM 2007.
- [2] O. Heckmann, M. Piring, J. Schmitt, and R. Steinmetz, "On realistic network topologies for simulation," In ACM SIGCOMM workshop on Models, Methods and Tools for Reproducible Network Research, pages 28- 32, 2003.
- [3] G. K., "Zipf. Selected Studies of the Principle of relative Frequency in Language," Harvard University Press, 1932.
- [4] D. Lee, J. Choi, J. Kim, S. Noh, S. Min, Y. Cho, and C. Kim, "Lrfu: A spectrum of policies that subsumes the least recently used and least frequently used policies," IEEE Transactions on Computers, 50(12):1352- 1361, 2001.