

파일 단위 Cross-Torrent 기법을 이용한 BitTorrent 성능 향상에 관한 연구

조은상, 권태경, 최양희
서울대학교 컴퓨터공학부

escho@mmlab.snu.ac.kr, {tkkwon, yhchoi}@snu.ac.kr

Study on BitTorrent Performance Improvement Using Per-File Cross-Torrent Scheme

Eunsang Cho, Ted “Taekyoung” Kwon, Yanghee Choi
School of Computer Science and Engineering, Seoul National University
escho@mmlab.snu.ac.kr, {tkkwon, yhchoi}@snu.ac.kr

요 약

BitTorrent의 낮은 Content Availability는 이 시스템의 중요한 문제 중 하나이다. 이를 해결하기 위한 방법 중 하나로 단일 torrent가 아닌 복수 torrent 사이에서 공유를 촉진시키려는 cross-torrent 기법이 제안되어 왔다. 이 가운데 cross-torrent bartering은 서로 다른 복수 torrent에 동시에 참여하는 복수 peer들이 서로 다른 torrent의 데이터도 교환할 수 있도록 하는 것인데, 최근의 measurement 연구에 의하면 이는 측정된 전체 경우 중 약 0.1%에 해당하여 기대되는 성능 향상이 미미한 것으로 알려져 있다. 본 논문에서는 이러한 cross-torrent bartering 개념을 bundled torrent에 적용하여 파일 단위의 bartering을 수행할 수 있는 기술적 가능성을 분석해보고 measurement 연구를 통하여 이 기법의 BitTorrent 성능 향상 기여에 대해 예측한다. Bundled torrent란 하나의 torrent 파일이 내부적으로 여러 개의 파일로 구성된 경우를 이르는데, 이러한 여러 개의 파일들은 BitTorrent 클라이언트의 종류에 따라 약간 상이한 구성 방식을 가지고 있다. 이러한 차이를 기반으로 하여 bundled torrent에 대한 파일 단위 bartering이 가능한 경우를 살펴본다. 또한 가장 잘 알려진 BitTorrent Index Site 중 하나인 The Pirate Bay에 약 한 달 간 등록된 모든 torrent 파일에 대해 측정된 결과를 바탕으로, 파일 단위 bartering 기법의 성능 향상 기여에 대하여 예측한다.

1. 서론

BitTorrent[1]는 데이터 공유를 목적으로 최근 가장 널리 쓰이는 Peer-to-peer(이하 P2P) 시스템이다[2, 3]. 이 시스템은 데이터 공유를 그 목적으로 하므로 가용성(availability)이 매우 중요한 요소이며, 이에 대한 연구가 활발히 진행 중이다. 최근의 연구 결과에 따르면 torrent 파일이 생성된 지 약 한 달의 시간이 흐르면, 약 62%의 torrent만이 가용한 상태로 남게 된다.[4]

BitTorrent의 가용성 측면에서 볼 때 특히 데이터의 완

전한 사본을 가지고 있는 peer, 즉 seed의 존재가 매우 중요하다. Seed가 존재하는 경우 해당 torrent 파일의 데이터는 완전히 다운로드 받을 수 있기 때문이다. 그러나 seed가 존재하지 않더라도 torrent 파일의 데이터를 완전히 다운로드 받을 수 있는 경우가 존재하는데, 이는 두 가지로 나누어 생각해 볼 수 있다. 첫째는 여러 peer들이 데이터의 서로 다른 부분을 나눠 갖고 있어서 전체적으로 완전한 사본이 존재하는 경우이다. 다음으로는 seed 또는 가용하지 않던 데이터의 부분을 가진 peer가 간헐적으로 BitTorrent 시스템에 참여하는 경우이다. 이렇게 하여 완전히 다운로드

받은 peer가 발생하면, 이 peer가 곧 seed가 되므로 해당 torrent 파일은 다시 가용해진다.

이 두 가지 경우에서 첫 번째 경우는 해당 torrent 자체에 완전한 사본이 분산되어 있는 것으로, 외부의 도움이 없이 완전한 사본을 구성할 수 있는 가능성이 있다. 그러나 두 번째 경우는 외부의 도움, 즉 현재 해당 torrent에 참여하지 않는 peer들의 참여가 있어야 완전한 사본이 구성된다. 이 점에 착안하여 단일 torrent가 아닌 복수 torrent 사이에서 공유를 촉진시키려는 cross-torrent 기법이 제안되어 왔다.

여러 cross-torrent 기법 가운데 cross-torrent bartering은 서로 다른 복수 torrent에 동시에 참여하는 복수 peer들이 각각 다른 torrent의 데이터도 교환할 수 있도록 하는 것이다. 즉, 한 peer가 어느 torrent에 이미 참여하고 있는 경우, 다운로드 받고자 하는 다른 torrent에서 이미 참여 중인 torrent의 데이터를 업로드하는 조건으로 원하는 데이터를 다운로드 받을 수 있는 방식이다. 이러한 방식은 seed 또는 일부 데이터를 가진 peer가 해당 torrent에 보다 오래 남아 있도록 만들 수 있어 BitTorrent의 가용성을 향상시키는 효과가 있다.

그런데 Cross-torrent bartering은 서로 다른 peer들이 교환 조건에 맞는 torrent들에 동시에 참여하고 있어야 하는 가정에 기반하고 있다. [4]에 따르면 이러한 경우는 전체 측정치의 0.1%에 해당한다. 즉, 전체적인 BitTorrent 가용성에는 큰 영향을 주지 못할 것으로 예측할 수 있다.

본 논문에서는 이러한 cross-torrent bartering 개념을 bundled torrent에 적용하여 파일 단위의 bartering을 수행하는 기법을 제안하고 평가한다. 2장에서는 파일 단위 bartering 기법을 제안한다. 3장에서는 BitTorrent 시스템에 대한 대규모 측정 방법과 그 결과에 대하여 기술하고, 4장에서는 제안한 기법이 BitTorrent 가용성에 어떠한 영향을 줄 것인지에 대하여 전 장의 측정 결과 분석을 바탕으로 예측한다.

2. 파일 단위 Cross-Torrent Bartering

가. Bundled Torrent

Bundled torrent란 하나의 torrent 파일에 포함된 데이터가 실질적으로 여러 개의 파일 형태로 구성되어 있는 것을 의미한다. 즉, 이러한 torrent는 포함된 데이터 전체를

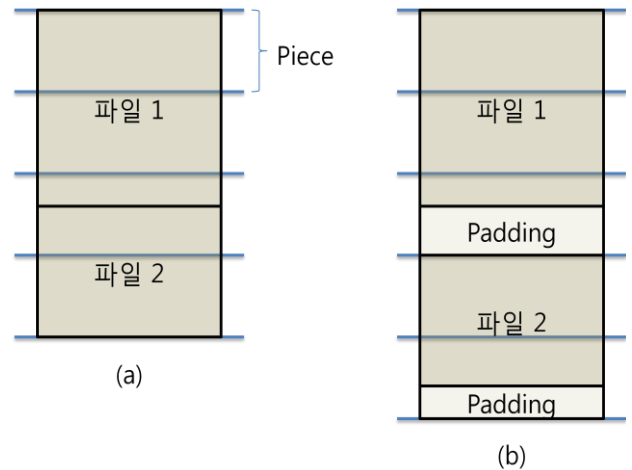


그림 1. Bundled torrent의 구조. (a) 파일을 연달아 배치한 경우. (b) 한 파일의 시작점이 항상 piece의 시작점과 일치하도록 배치하는 경우.

공유하는데 이용할 수도 있고, 일부 파일만을 원할 때 이용할 수도 있다. 예를 들면, 아서 코난 도일(Sir Arthur Ignatius Conan Doyle)의 추리 소설¹을 공유한다면 《A Study of Scarlet》 한 권을 torrent로 만들어 공유할 수도 있지만, 셜록 홈즈(Sherlock Holmes)가 등장하는 여러 소설들을 묶어 하나의 torrent로 만들 수도 있다. 이 중 후자의 경우를 bundled torrent라고 한다.

이러한 bundled torrent를 만드는 방법은 그림 1과 같이 크게 두 가지가 있다. 첫째는 파일을 연달아 배치하는 것이다. 이 경우에는 한 파일의 시작점이 어느 BitTorrent piece의 시작점과 일치하지 않는 경우가 발생할 수 있다. 다음으로는 한 파일의 시작점이 항상 piece의 시작점과 일치하도록 배치하는 것이다. 이 경우 어느 파일이 끝나는 지점이 piece의 끝나는 지점과 일치하지 않을 경우에는 padding을 두어 해결한다.

첫 번째 방법은 torrent에 포함된 데이터 크기가 파일 크기의 합과 같은 반면, 두 번째 방법은 padding이 포함되어 torrent의 데이터 크기가 더 커지는 단점이 있다. 그러나 두 번째 방법은 일부 파일만을 다운로드 하고 싶을 때 더 적은 piece만을 다운로드 하게 되는 장점이 있으며, piece 크기가 같은 여러 torrent에 동일한 파일이 존재할 경우 piece의 hash값이 같게 되는 특징이 있다.

나. 파일 단위 Cross-Torrent Bartering

Cross-torrent bartering은 서로 다른 peer들이 교환

¹ 본문에 언급한 소설은 작가가 1930년에 사망하여, 2010년 현재 국내에서 원작 텍스트에 대한 법적인 저작권 보호 기간이 만료되었다.

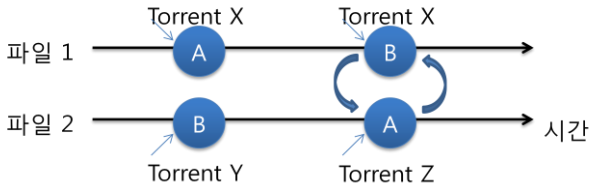


그림 2. 파일 단위 cross-torrent bartering

조건에 맞는 torrent들에 동시에 참여하고 있어야 하는 가정에 기반하고 있어, 이러한 조건을 만족시키기 어려운 단점이 있다. 이에 cross-torrent bartering을 bundled torrent에 적용하여 파일 단위로 확장하는 방식을 그림 2와 같이 제안한다.

각 peer는 bartering에 적용할 대상을 torrent 단위에서 bundled torrent에 포함되어 있는 파일 단위로 바꾸게 되며, 이를 통해 교환 조건을 만족시키는 가능성을 높일 수 있다. 이는 곧 BitTorrent 시스템의 가용성 향상에 영향을 주는 요인이 된다. 이 기법이 줄 영향을 분석하기 위하여 다음 장에 제시된 내용과 같이 BitTorrent 측정 시스템을 구축하고, 정보를 수집하였다.

3. BitTorrent 시스템 측정

BitTorrent 시스템을 측정하기 위하여 가장 잘 알려진 BitTorrent Index Site [5] 중 하나인 The Pirate Bay (<http://thepiratebay.org>)를 선정하고, 2010년 5월 1일부터 30일까지 약 한 달 여의 기간 동안 같은 기간 이 사이트에 등록된 모든 torrent 파일에 대해 측정하였다. 측정된 결과는 표 1과 같다.

표 1. BitTorrent 시스템 측정 결과

측정 기간	2010년 5월 1일 ~ 30일	
총 수집 torrent 수	50,247	100.0%
Bundled torrent 수	35,325	70.3%
Single torrent 수	14,922	29.7%
총 수집 파일 수	651,512	100.0%
Bundled torrent 포함	636,590	97.7%
Single torrent 포함	14,922	2.3%

측정 결과 해당 기간 총 수집된 torrent 수는 50,247건이었다. 이 가운데 bundled torrent는 35,325건으로 약

70.3%를 차지하였다. 특히 한 파일의 시작점이 항상 piece의 시작점과 일치하도록 배치하는 방식의 bundle을 구성한 torrent는 213건으로 이는 전체 bundled torrent 가운데 약 0.6%의 비중을 가지고 있었다. 한편, 같은 기간 bundled torrent에 포함된 파일 수를 살펴보면 총 636,590건으로, 하나의 bundled torrent 당 약 18개의 파일이 포함되어 있었다.

4. BitTorrent 가용성 영향 분석

Cross-torrent bartering이 가능한 경우는 어느 peer A와 B가 같은 시점에 두 개 이상의 torrent에 동시에 참여하고 있는 경우에만 가능하다. 그러나 파일 단위 cross-torrent bartering에서는 어느 peer A와 B가 같은 시점에 두 개 이상의 파일에 동시에 참여하고 있는 경우, 교환이 가능하다. 이 때 이 파일은 각각 서로 다른 torrent에 존재하는 것일 수 있다.

예를 들면, peer A가 torrent X, Y에 참여하고 있고, peer B는 torrent Y, Z에 참여하고 있을 때, 이 peer들은 cross-torrent bartering이 불가능한 경우이다. 그러나 파일 단위 cross-torrent bartering에서는 A와 B가 참여하고 있는 파일이 X와 Z에 동시에 존재하고 있는 경우, 서로 교환이 가능해진다. 이로부터 파일 단위 cross-torrent bartering이 기존 cross-torrent bartering에 비해 얻는 이득이 분명해진다.

어떤 파일이 서로 다른 torrent에 동시에 존재하는 경우, 이를 BitTorrent 시스템에서 활용하려면 동시에 존재하는 파일을 찾아내는 방법이 필요하다. 이를 위하여 가능한 방법으로는 torrent에서 파일을 구성하는 각 piece의 hash 값을 바탕으로 비교하는 방법과 각 파일의 이름, 크기 등의 메타정보를 바탕으로 비교하는 방법이 가능하다.

첫 번째 방법은 동일한 두 파일을 찾아내는데 유용하게 쓰일 수 있다. 그러나 piece 크기가 동일한 경우 hash값 비교가 가능한, 한 파일의 시작점이 항상 piece의 시작점과 일치하도록 배치하는 방식의 경우 전 장의 측정 결과에서 보듯이 전체 bundled torrent 가운데 약 0.6%에 해당하므로 실제로 적용하기 매우 어렵다.

두 번째 방법은 각 파일의 이름, 크기 등의 메타정보를 활용하는 방식이므로 파일이 연속적으로 배치된 bundled torrent의 경우에도 적용이 가능하지만, 정확성이 떨어지는

단점이 있다. [6]에 따르면, torrent의 이름 유사성을 바탕으로 클러스터링할 경우, 유사성이 70% 이상일 때 클러스터링 결과가 90% 이상 정확하게 나타났다. 또한 유사성이 70% 이상인 경우, 실험에 사용한 약 5만 여 샘플로부터 약 3만 개의 클러스터가 생성되었다. 이는 각 샘플이 평균 0.7개의 유사한 샘플을 가지고 있다고 해석할 수 있다. 이러한 결과에서 볼 때 메타정보를 바탕으로 유사한 파일을 찾아내는 방식은, 파일 단위 cross-torrent bartering의 대상을 동일한 파일 뿐 아니라 유사한 파일로 까지 확대할 경우 가용성 증대에 큰 영향을 줄 수 있다.

5. 결론

BitTorrent는 최근 널리 쓰이는 데이터 공유 시스템으로, 시간이 지남에 따라 가용성이 저하되는 문제가 중요한 요소로 주목받고 있다. 본 논문에서는 BitTorrent 시스템의 가용성을 제고하기 위하여 파일 단위 cross-torrent bartering을 제안하고, BitTorrent 시스템의 실제 측정 결과를 바탕으로 제안한 기법이 가용성에 미치는 영향을 분석하였다.

분석 결과 cross-torrent bartering과 비교하여 파일 단위 cross-torrent bartering은 BitTorrent 시스템의 가용성을 보다 향상시킬 가능성이 있으며, 이를 달성하기 위하여 파일의 메타정보를 이용하여 찾아낼 수 있는 유사한 파일을 활용하는 것이 중요하였다.

이와 관련하여 향후 유사한 파일을 찾아내고 이를 활용할 수 있는 방안에 대한 연구가 이루어 질 필요가 있다. 또한 동시적인 peer 공유 관계 외에도 시간의 흐름을 고려한 분석이 필요한데, 이는 향후 연구를 계속 진행할 예정이다.

6. 감사의 글

본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업(정보통신)의 일환으로 수행하였음 [10035245: 미래인터넷에서의 이동환경 및 네트워크 다양성 지원구조 연구].

이 연구를 위해 연구장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사 드립니다.

7. 참고문헌

- [1] B. Cohen. "Incentives Build Robustness in BitTorrent," Workshop on Economics of P2P systems. Jun. 2003.
- [2] H. Schulze and K. Mochalski, "ipoque :: Internet Study 2007 Data about P2P, VoIP, Skype, File Hosters like RapidShare and Streaming Services like YouTube." <http://www.ipoque.com/resources/internet-studies/internet-study-2007/>.
- [3] H. Schulze and K. Mochalski, "ipoque :: Internet Study 2008/2009." http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009/.
- [4] S. Kaune, R. C. Rumín, G. Tyson, A. Mauthe, C. Guerrero, and R. Steinmetz, "Unraveling BitTorrent's File Unavailability: Measurements and Analysis," 2010 IEEE Tenth International Conference on Peer-to-Peer Computing (P2P). Aug. 2010.
- [5] C. Zhang, P. Dhungel, D. Wu, and K. W. Ross, "Unraveling the BitTorrent Ecosystem," to appear in IEEE Transactions on Parallel and Distributed Systems (IEEE TPDS), <http://cis.poly.edu/~ross/papers/PublicEcosystem.pdf>.
- [6] J. Han, T. Chung, H. Kim, T. Kwon, and Y. Choi, "Systematic Support for Content Bundling in BitTorrent Swarming" IEEE INFOCOM Student Workshop, Mar. 2010.