

BitTorrent 시스템의 콘텐츠 공유 패턴 분석

조은상, 권태경, 최양희
서울대학교 전기컴퓨터공학부

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

Analysis on the Content Sharing Patterns of BitTorrent System

Eunsang Cho, Ted "Taekyoung" Kwon, Yanghee Choi

School of Computer Science and Engineering, Seoul National University

요 약

BitTorrent 는 가장 널리 사용되는 P2P 시스템으로, 수많은 콘텐츠의 유통이 이를 통해 발생하고 있다. 따라서 BitTorrent 의 성능 개선은 매우 중요한 이슈로 연구되어 왔으며, 특히 비디오나 소프트웨어의 배포와 같이 콘텐츠의 종류를 한정하여 효율을 높이는 시스템이 제안되고 있다. 그러나 이렇게 제안된 시스템들은 보통 용도제한적인 상황을 가정하고 있어, 범용적인 P2P 시스템에 있어서 성능 향상의 필요성이 있다. 범용적인 시스템에서 성능을 향상시키기 위해서는 현재의 시스템에 대한 이해가 절실히 필요한데, 본 논문에서는 현재 BitTorrent 시스템의 콘텐츠 공유 패턴을 수집, 분석하여 이 시스템에 대한 이해를 높이고자 하였다. 그 결과 어떤 종류의 콘텐츠 공유가 주류를 이루고 있으며, 각 콘텐츠 종류별 특징이 어떠한지를 분석하였다. 이를 바탕으로 향후 새로운 P2P 시스템을 개발하거나 기존 시스템을 개선함에 있어서 보다 효율적인 시스템을 구축하는데 기초를 제공할 수 있을 것으로 기대된다.

I. 서론

오늘날 인터넷 상의 콘텐츠 유통 방식은 BitTorrent[1]와 같은 P2P 시스템으로 인하여 혁신적으로 변화하는 중이다. 기존의 콘텐츠 유통방식은 공급자가 콘텐츠 서버의 하드웨어와 네트워크 대역폭을 마련하여 수요자가 직접 서버에서 내려 받는 방법을 통하였다. 이러한 방식의 대표적인 사례로는 CDN(Content Delivery Network)이 있다. P2P 시스템은 수요자가 자발적인 참여로서 콘텐츠를 내려 받기 위해 자신의 하드웨어와 네트워크 대역폭을 제공하므로, 공급자의 입장에서는 서버 하드웨어와 네트워크 대역폭이 필요하지 않다. 단지 콘텐츠를 처음 제공하는 역할만을 할 뿐이다.

이러한 콘텐츠 유통 구조의 변화로 인하여 콘텐츠 공급자들이 속속 자신들의 콘텐츠를 P2P 시스템을 통하여 배포하는 사례가 나타나고 있다. 잘 알려진 사례로는 World of Warcraft 라는 게임을 제공하는 미국의 Blizzard Entertainment Inc.는 게임 클라이언트의 업데이트를 P2P 시스템을 활용한 Blizzard Downloader 를 통해 배포한다.[2]

ipoque 에 따르면 P2P 는 전세계적으로 가장 많은 트래픽을 생산하고 있으며, P2P 중에서는 BitTorrent 가 가장 많은 비중을 차지하고 있다. BitTorrent 는 남미를 제외한 전지역에서 가장 많이 쓰이는 프로토콜이며, HTTP 가 BitTorrent 의 뒤를 따르고 있다.[3] 따라서 BitTorrent 의 성능을 개선하기 위한 연구가 계속되고 있다.

BitTorrent 의 성능 개선은 특히 비디오나 소프트웨어의 배포와 같이 콘텐츠의 종류를 한정하여 효율을 높이는 방법으로 진행되어 왔다. Phosphite[4]는 VoD(Video on Demand) 서비스를 함에 있어서 BitTorrent 를 기본으로 하여 성능을 향상시키는 기법을 제안한 것이다. Blizzard Downloader 의 경우에도 사용한 P2P 프로토콜이 BitTorrent 를 기반으로 한 것으로 알려져 있다.[5] 그러나 이러한 시

스템들은 용도제한적인 상황을 가정하고 있어, 범용적인 시스템에서 성능을 향상시킬 필요성이 있다.

범용 시스템에서 성능을 향상시키기 위해서는 현재 시스템이 어떻게 활용되고 있는지에 대한 이해가 절실히 필요하다. 따라서 본 논문에서는 현재 사용되고 있는 BitTorrent 시스템의 콘텐츠 공유 패턴을 수집, 분석하여 이 시스템에 대한 이해를 높이고자 하였다.

II. 자료 수집 방법

BitTorrent 시스템의 콘텐츠 공유 패턴을 알아보기 위하여 Torrent 메타 검색 사이트에서 제공하는 정보를 수집하였다. 널리 알려진 Mininova (<http://www.mininova.org/>)와 같이 torrent 파일들을 모아서 제공하는 웹사이트들을 인덱스 사이트라고 하는데, Torrent 메타 검색 사이트는 이러한 인덱스 사이트들의 정보를 한 데 모아서 제공한다. 따라서 Torrent 메타 검색 사이트를 통해 수집한 정보는 특정 인덱스 사이트만을 이용했을 때 보다 이용자의 성향에 의존적이지 않고 콘텐츠의 편중이 발생하는 것을 방지하는 측면에서 대표성을 가지는 자료이다.

자료 수집에 이용된 Torrent 메타 검색 사이트는 Torrentz (<http://www.torrentz.com/>)의 검증된 파일 목록을 이용하였다. 2009년 11월 16일부터 12월 14일까지 29일간 매일 22시경에 <http://www.torrentz.com/verified/> 웹 페이지에서 제공하는 일 최대 10만 건에 해당하는 torrent 파일 정보를 수집하였다. 수집한 정보는 총 1,599,572 건이며, 이는 총 120,407 개의 torrent 파일에 대한 정보이다. 수집한 정보의 종류는 torrent 파일의 제목, 생성일, 파일 크기, seed 와 leecher 의 수, 사용자가 분류한 콘텐츠의 종류 등이다.

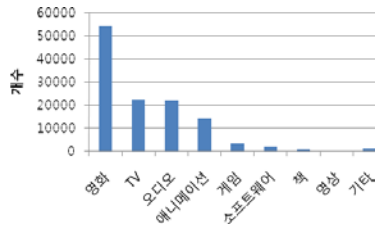


그림 1. 콘텐츠 분류별 torrent 파일 개수

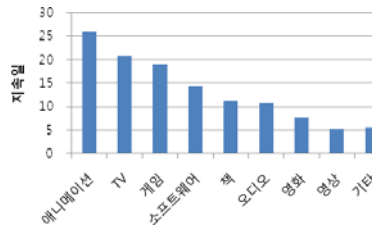


그림 2. 콘텐츠 분류별 공유 지속일

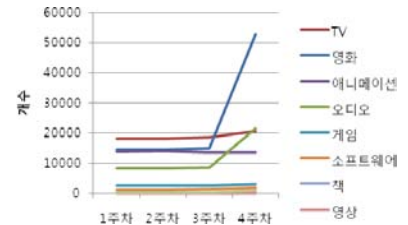


그림 3. 주차별/콘텐츠 분류별 torrent 파일 개수

분류	평균 Seed 수	평균 Leecher 수	평균 크기 (MB)
영화	52.335	42.334	1723.562
TV	81.418	33.973	573.608
오디오	28.197	6.795	641.702
애니메이션	8.213	5.178	486.480
게임	79.470	136.920	2848.930
소프트웨어	56.675	23.950	649.899
책	20.129	7.751	298.027
영상	1.425	1.525	101.333
기타	65.967	27.596	1269.347

표 1. 콘텐츠 분류별 평균 seed 수, leecher 수, 크기

완료하지 못한 leecher 상태로 측정되어 다른 종류의 콘텐츠와 반대의 경향을 나타내었을 가능성이 있다.

<그림 3>은 전체 29 일의 자료를 일주일 간격으로 나누어 분석한 것이다. 이를 통해 콘텐츠 종류별 순위가 TV, 영화, 애니메이션 순으로 꾸준히 유지됨을 확인할 수 있었다. 그러나 1~3 주차의 순위와 29 일간의 순위에서 상위에 속하는 콘텐츠가 조금 다르게 나타나는데, 이는 4 주차에 영화와 오디오 관련 콘텐츠가 상당히 큰 폭으로 증가하였기 때문에 발생한 결과이다. 이 것은 이 기간에 새로운 torrent 파일들이 새롭게 공유되기 시작하였음을 의미한다.

III. 자료 분석

29 일 동안 수집한 콘텐츠를 사용자가 분류한 콘텐츠의 종류 정보를 바탕으로 하여 총 8 가지 종류로 분류하였다. 이 때 영화, TV, 애니메이션 등은 모두 동영상 항목으로 묶일 수 있다. 그러나 이 자료에서는 각 항목 자체가 큰 비중을 차지하고 있으며, 각각 서로 다른 특성을 보이고 있어 전체를 하나로 묶지 않고 별개의 항목으로 분류하였다. 또한 사용자가 분류한 콘텐츠의 종류 정보는 torrent 파일이 처음 등록되었을 때는 불분명한 경우가 있다. 이러한 경우는 따로 '기타'로 분류하였다. 그리고 사용자가 'misc' 등의 '기타'를 의미하는 방식으로 분류한 경우에도 '기타'로 분류하였다.

<그림 1>은 각 분류별 torrent 파일 개수를 나타낸다. 이를 통해 나타난 주요 콘텐츠는 영화, TV, 오디오, 애니메이션 등이며 동영상과 음악과 같은 멀티미디어 콘텐츠에 해당함을 알 수 있다.

같은 torrent 파일이 몇 일 간 나타났는가 하는 공유 지속일은 콘텐츠가 얼마나 지속적으로 공유되었는지를 보여준다. <그림 2>에서 보듯이 공유의 지속성은 애니메이션, TV, 게임 순으로 높았으며 애니메이션, TV 와 같이 콘텐츠 자체의 성격이 연속성을 가진 경우 다른 콘텐츠 보다 오랜 기간 공유되고 있었다.

한 torrent 파일에 대해서 얼마나 많은 수의 peer 가 참여하는지는 해당 콘텐츠가 얼마나 인기있는 콘텐츠인가를 나타낸다. 특히 seed 의 숫자가 많은 경우에는 해당 콘텐츠의 가용성이 높음을 나타내는 중요한 지표이다. <표 1>을 통해 볼 때 대부분의 종류에서 seed 의 숫자가 10 개 이상으로 충분히 가용한 상태에 있음을 알 수 있었다. 애니메이션의 경우 평균 seed 의 숫자가 8.213 개로 다른 콘텐츠 종류에 비해 상대적으로 작게 나타났는데 이는 애니메이션이 오랜 기간 공유되고 있지만 소수의 peer 들만이 참여하고 있다는 것을 보여준다.

한편, 많은 경우 seed 의 수가 leecher 의 수보다 큰 것으로 나타났는데, 게임의 경우에는 그와 반대의 경향을 보였다. 이것은 <표 1>에 나타난 콘텐츠의 평균 크기에서 연유하는 것으로 보인다. 즉, 게임의 경우는 약 2.8GB 정도의 평균 크기를 보이는데, 다른 콘텐츠에 비해서 그 크기가 월등히 크다. 따라서 많은 수의 peer 가 다운로드를

IV. 결론

29 일 동안 웹 크롤링을 통해 수집한 자료를 바탕으로 BitTorrent 시스템에서 어떤 콘텐츠가 공유되고 있는지에 대하여 분석하였다. 분석 결과 영화, TV, 애니메이션과 같은 동영상 및 음악 등의 멀티미디어 콘텐츠가 대부분을 차지하였다.

따라서 현재의 용도제한적 성능 개선 방법을 범용적인 시스템에 적용할 경우에도 어느 정도 성능 향상을 기대할 수 있을 것으로 보인다. 또한 이를 바탕으로 향후 콘텐츠 중심 네트워크 및 P2P 등을 연구함에 있어서 특히 동영상에 갖는 특성을 이용한다면 보다 효율적인 네트워크를 만들 수 있을 것이다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT 신성장동력핵심기술개발사업의 일환으로 수행하였음. [2007-F-038-02, 미래 인터넷 핵심기술 연구]

이 연구를 위해 연구장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사드립니다.

참고문헌

- [1] B. Cohen, "Incentives Build Robustness in BitTorrent", Workshop on Economics of Peer-to-Peer systems, June, 2003.
- [2] Blizzard Downloader F.A.Q. <http://www.worldofwarcraft.com/info/faq/blizzarddownloader.html>.
- [3] H. Schulze, et al., "ipoque :: Internet Study 2008/2009", http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009/.
- [4] M-L. Champel, et al., "Phosphite: Guaranteeing out-of-order download in P2P Video on Demand", IEEE P2P '09, September, 2009.
- [5] Blizzard Downloader, http://www.wowwiki.com/Blizzard_Downloader.