

Afreeca 인터넷 방송 서비스에서의 사용자 행동 분석

박용덕⁰, 박건우, 권태경, 최양희
서울대학교 전기컴퓨터 공학부
{ydpark, kwpark, tk, yhchoi}@mmlab.snu.ac.kr

User behaviors in "Afreeca" Internet Broadcast Service.

Yongdeok Park⁰, Kunwoo Park, Taekyoung Kwon and Yanhee Choi
School of Computer Science and Engineering, Seoul National University

요 약

광대역 통신망의 발달과 더불어 정보통신 및 디지털 영상 기기의 대중화로 인해 텍스트 및 사진 위주의 정보 전달과 달리 소리와 영상을 기반으로 한 멀티미디어 정보 전달이 우리 생활 속의 일부로 자리 잡고 있다. 이러한 멀티미디어 정보 전달에 있어 트래픽 분산을 위한 P2P(Peer to Peer) 방식의 실시간 스트리밍 서비스에 대해 많은 연구가 있었다. 본 논문에서는 "사용자"를 중심에 두고 사용자의 서비스 이용에 대한 패턴을 파악하기 위해 사용한 "Afreeca" 인터넷 방송서비스와 사용자 패턴 수집 방법에 대해 설명한 다음, 수집된 결과를 이용하여 콘텐츠 특성별 사용자 패턴의 차이와 특징을 분석해 보았다.

1. 서 론

오늘날, 광대역 통신망의 발달로 인해 개인 사용자의 사용가능한 대역폭이 크게 늘어남과 더불어 멀티미디어 정보통신 기기의 광범위한 보급으로 인해 과거의 글 및 사진 위주의 정보 전달에 소리와 영상을 더할 수 있게 되었으며, 단순히 음악파일이나 동영상 파일을 찾아 저장 후 재생하는 단계의 eDonkey, Pruna[1,2]와 같은 서비스로부터 발전하여 사용자가 원하는 정보를 선택하면 실시간으로 재생할 수 있는 YouTube나 Daum TV팟[3,4] 등과 같은 실시간 멀티미디어 스트리밍 서비스(Realtime multimedia streaming service)가 우리 일상 생활의 일부로 자리잡고 있다.

하지만, 이러한 서비스를 전통적인 클라이언트/서버 방식으로 지원하기에는 여러 측면에서 단점이 있으므로 Peer-to-peer (P2P) 방식의 연구가 활발히 진행 중에 있다[5, 6, 7].

P2P방식으로 실시간 멀티미디어 스트리밍을 지원하려면 몇가지 넘어야 할 산이 있다[8]. 그 중에서도 우리는 콘텐츠별 사용자의 행동 패턴을 파악하는데 초점을 맞춰 연구를 하고자 한다. 콘텐츠별 사용자의 서비스 이용에 따른 행동 패턴의 차이와 특징을 파악함으로써 사용자로 인해 발생하는 부하를 콘텐츠 제공자가 예상할 수 있고

그 결과를 통해 스트리밍 서비스의 질을 개선하거나 효율적인 P2P 시스템을 디자인하는데 사용할 수 있을 것이다.

본 논문의 2장에서는 콘텐츠에 대한 사용자들의 행동 패턴을 파악하기 위해 이용한 인터넷 방송 서비스인 Afreeca에 대해 기술한다. 3장에서는 Afreeca에서 사용자들의 서비스 이용에 따른 사용 패턴을 어떻게 수집하였는가에 대한 방법에 대해 설명하고 마지막으로 수집한 콘텐츠별 사용자들의 Afreeca 서비스 이용 행동 패턴 결과와 특징을 확인한 결과를 제시하고 향후 연구에 대해 기술한다.

2. Afreeca 소개

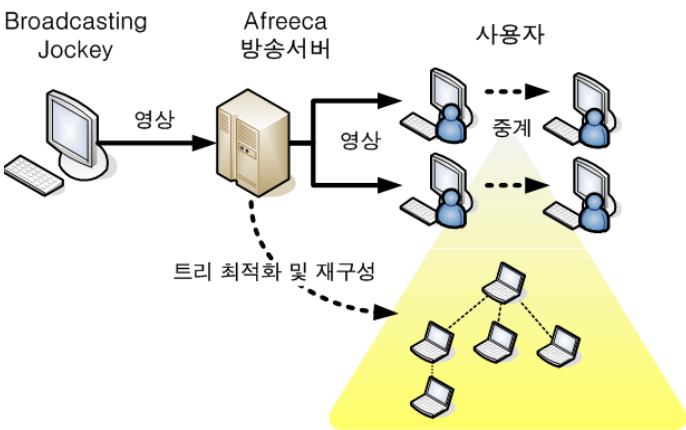
지금까지의 관련 연구들은 콘텐츠 제공자의 백본을 통해 직접 로그를 수집하거나[9], Bot이나 Web crawler등과 같은 것을 이용하여 간접적으로 사용자의 패턴에 대한 정보를 수집했다[10].

본 논문에서는 콘텐츠 제공자가 아닌 서비스 사용자로서 다른 사용자의 서비스 사용 패턴을 파악해 보기 위해 Afreeca라는 인터넷방송 서비스를 이용하였다.

Afreeca는 나우콤에서 개발한 DRD(Dynamic Relay Distribution ; 대용량 트래픽 분산 전송기술)을 이용한

애플리케이션 계층 멀티캐스트 서비스로서 [그림 1]과 같이 방송 서버를 서비스 제공자가 관리하고 콘텐츠 방송자(Broadcasting Jockey)가 동영상 파일이나 TV 수신카드로 수신중인 영상을 다른 시청자들에게 방송 서버를 통해 스트리밍 중계를 한다.

Afreeca에서 사용되는 DRD 기술은 방송 서버로부터 오는 스트리밍 콘텐츠 데이터를 여러 사용자가 받고 있을 때 서버에 부하가 집중되지 않도록 사용자들 간의 멀티캐스트 트리를 생성하여 서버로부터 직접 데이터를 받지 않고 이미 스트리밍 콘텐츠 데이터를 받은 사용자들 통해 중계 받을 수 있도록 한다[11, 12].



[그림 1] Afreeca 서비스 구성

3. 실험방법

기존 대부분의 연구들은 ‘컨텐츠’에 초점을 맞추고 사용자가 만들어서 올린 User Created Contents(UCC)의 랭킹 또는 평가와 같은 인기를 고려해서 어떻게 효율적으로 캐싱할 것인지 또, 인기는 어떤 특징을 가지고 있고 사용자는 몇 명이고 동영상 콘텐츠를 얼마나 몇 번 봤는가를 가정을 하였었다[13, 14].

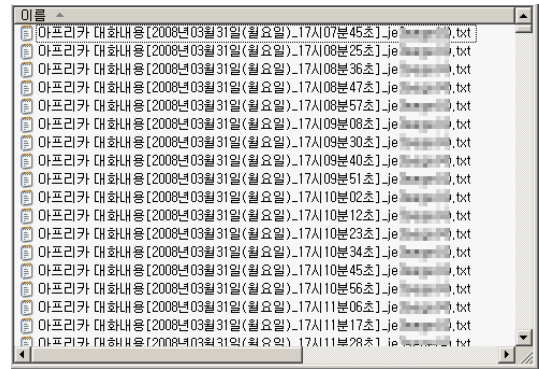
본 논문에서의 초점은 ‘사용자’이다. 사용자들의 서비스 사용 상황을 분석하기 위해선 사용자 출입에 관한 로그를 얻을 수 있어야 하는데 Winamp[15]나 곰플레이어[16]에서 사용자의 출입 패턴을 다른 사용자가 알기는 쉽지 않다.

그러나, [그림 2]에서와 같이 Afreeca에서는 채팅 기능을 제공하며 채팅창을 통해 표시되는 시청자 출입표시를 이용하여 사용자의 출입 상황을 확인할 수 있다.



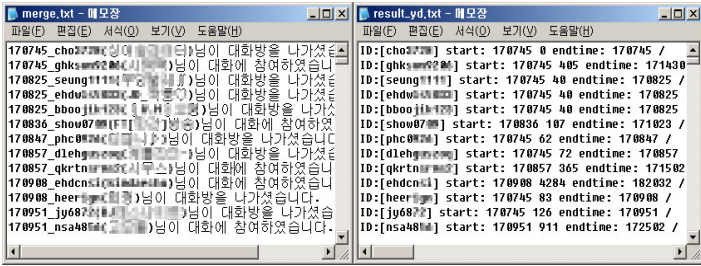
[그림 2] Afreeca 서비스 화면 및 채팅창

또, Afreeca에서 제공하는 기능에서 ‘저장’을 누르게 되면 저장한 시간이 생성되는 텍스트 파일명에 포함되고 채팅 내용이 텍스트 파일 내에 기록이 되게 된다. 이 파일을 계속 저장하여 [그림 3]과 같이 로그를 만든 후 기존의 파일 내용과 새로 저장된 파일 내용을 서로 비교하면 저장한 시점 간의 사용자들의 출입 상황을 알 수 있다.



[그림 3] 시간별 로그파일

그러나, 사용자들의 출입 상황을 의미 있는 행동 패턴으로 변환하기 위해서 일정 간격으로 파일을 저장할 수 있어야 했는데 Macro 스크립트 프로그램을 이용하여 이를 해결하였고, 저장된 파일들에서 출입 상황을 분석하기 위해 Java 프로그래밍을 이용하여 [그림 4]와 같이 저장된 파일 로그로부터 시간별 사용자의 출입을 수집 후 이용자별 출입 시간을 분리하였다.



[그림 4] 사용자 출입시간별 분리 로그 파일

마지막으로 콘텐츠별 사용자의 행동 패턴을 분석하기 위해 Game, TV, 증권, 영화 방송방에 대해 짧게는 1시간 35분부터 길게는 3시간 동안의 BJ의 방송서비스 시간 동안 사용자들의 출입 상황을 수집하였다.

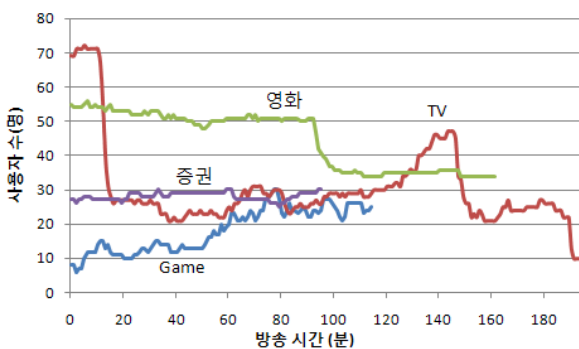
4. 결 과

4.1 콘텐츠별 사용자 패턴

B [표 1], [그림 5]는 앞 장에서 설명한 Afreeca의 로그 수집 방법을 통해 1~3시간 동안 서비스한 콘텐츠별 방송방에 대한 사용자 출입 상황 및 평균 서비스 이용 시간이다.

콘텐츠	방송시작 시간(시)	방송종료 시간(시)	평균 서비스 이용시간(초)
GAME	13:40	15:35	615
TV	10:24	13:38	1455
증권	10:53	12:48	1225
영화	16:39	19:07	1999

[표 1] 콘텐츠별 시작/종료시간 및 평균서비스 이용시간

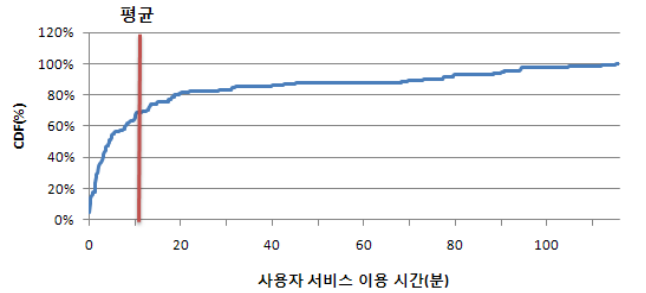


[그림 5] 콘텐츠별 사용자 출입상황

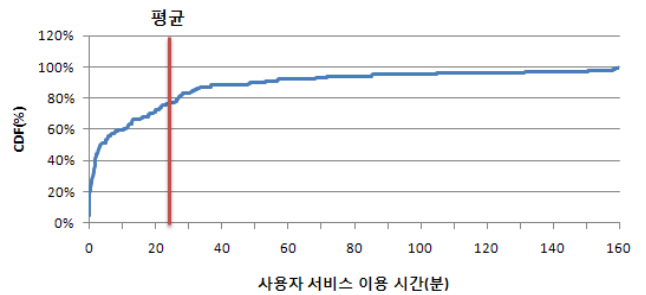
먼저 GAME 콘텐츠에서는 6~30명의 사용자가 약 2시간 동안 롤플레이팅 게임을 시청하였고 이들의 평균 서비스 이용시간은 약 10여분이었다. TV 콘텐츠에서는 약 3시간 동안 약 10~70명의 사용자가 서비스를 이용하였다. TV 콘텐츠에선 방송 프로그램 내용 변화에 따라 사용자

가 급격이 변화하는 모습을 볼 수 있었다. 증권방송에서는 사용자 26~30명 사이를 큰 변화 없이 유지되었고 영화에서는 34~55명의 사용자가 2시간 30분간 서비스를 이용하였으나 방 개설 95분 후인 영화 종료 시점에서 사용자가 급격히 줄어드는 모습이 관찰된다.

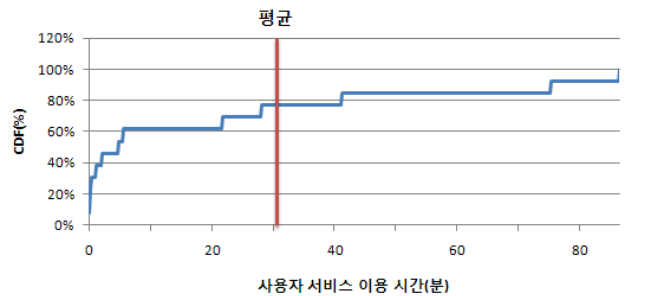
[그림 6]은 Afreeca의 로그 파일을 이용하여 만든 콘텐츠별 사용자 이용시간에 대한 누적 분포 그래프이다.



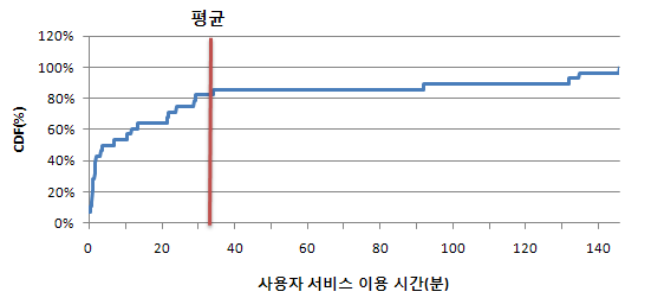
(a) Game



(b) TV



(c) 증권



(d) 영화

[그림 6] 콘텐츠별 사용자 누적분포

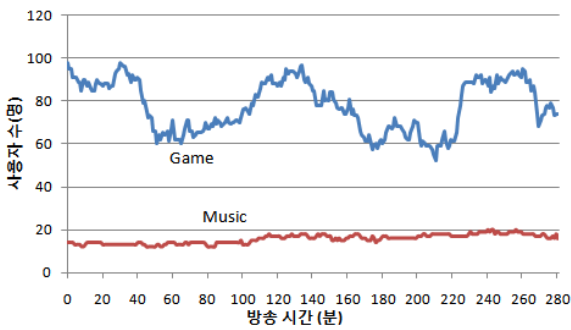
[그림 6(a)] Game방송의 사용자 누적 분포그래프에서는 총 115분의 서비스 시간 대비 평균 서비스 이용시간 비율이 8.6%로 4개 콘텐츠 중 가장 짧은 것을 확인 할 수 있다. [그림 6(b)] TV방송의 누적 분포그래프에서는 총 194분의 서비스 시간 대비 12.4%의 평균 서비스 이용시간을 보여준다.

반면 [그림 6(c)] 증권방송 누적 분포 그래프는 총 115분의 서비스 시간 대비 평균 서비스 시간 비율은 17.3%, [그림 6(d)] 영화방송 누적 분포그래프는 총 148분의 서비스 시간 대비 평균 서비스 시간 비율이 22.3%임을 알 수 있다. 따라서, 특정한 특징을 지닌 영화나 증권 및 누구나 즐겨볼 수 있는 TV, 음악 콘텐츠 사용자 패턴의 차이점을 [그림 6]을 통해 실제로 확인해 볼 수 있었다.

4.2 사용자 패턴의 특징

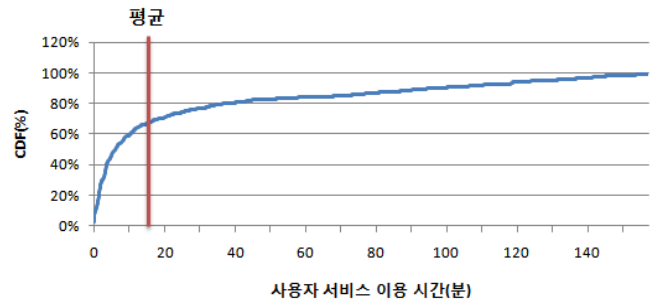
[그림 6]의 그래프가 가지는 특징으로는 각 콘텐츠 누적 접속자 수의 50%가 서비스를 10분 이하로 사용했다는 점이다. 이는 주로 사용자들이 어떤 방송을 하고 있는지 아니면 관심 있는 방송을 하고 있는지 확인 후 시청 여부를 바로 결정하기 때문이라 볼 수 있다. 또한 콘텐츠별 평균 서비스 이용 시간에 누적 사용자 비율이 70~90%에 도달함을 살펴 볼 수 있었다.

우리는 이러한 특징을 다시 한 번 확인하기 위해 [그림 7]과 같이 사용자수 변화가 급격한 Game방 및 사용자수의 변화가 크게 변하지 않은 음악방의 로그를 동일하게 4시간 40분간 측정을 해 보았다.

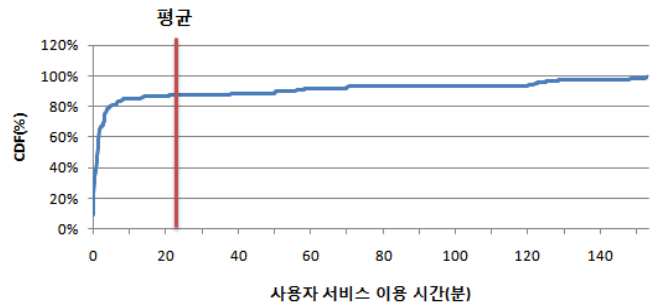


[그림 7] Game/Music 사용자 출입상황

4시간 40분동안 52~98명의 사용자가 Game 콘텐츠, 10~20명의 사용자가 Music 콘텐츠를 시청하였으며 [그림 8]은 콘텐츠별 사용자 누적그래프이다.



(a) Game



(b) 음악

[그림 8] Game/음악 콘텐츠 사용자 누적 그래프

[그림 8]에서도 공통적으로 서비스 시작 10분 동안 Game, 음악 콘텐츠 각각 60%, 85%의 누적 접속자 수에 도달함을 볼 수 있었고 각 콘텐츠의 사용자 평균 서비스 이용시간인 17.5분, 23분에는 각각 70%, 87%에 도달하였음을 보여주고 있다.

5. 결론

우리는 앞서 Afreeca에서 저장 기능을 이용하여 사용자의 출입 로그를 수집하는 방법과 이를 통해 얻은 로그를 이용하여 특징을 가지는 콘텐츠가 어떠한 사용자 행동 패턴을 보이고 있는지, 또 전체적인 사용자 패턴의 특징은 무엇인지 살펴보았다. 이러한 결과를 이용하여 사용자로 인해 발생하는 부하를 콘텐츠 제공자가 예상할 수 있고 그 결과를 통해 스트리밍 서비스의 질을 개선하거나 효율적인 P2P 시스템을 디자인하는데 사용할 수 있을 것으로 기대한다.

앞으로 좀 더 다양한 콘텐츠에 대한 특징을 이용, 세부적인 항목으로 구분하여 관심도에 따른 콘텐츠별 사용자 패턴의 특징에 대해 연구하고자 한다.

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 컴퓨터 연구소에 감사드립니다.

[15] Winamp. <http://www.winamp.com/>

[16] 그레텍 콤플레이어. <http://gom.gomtv.com/>

관련연구

[1] eDonkey. <http://www.edonkey2000.com/>

[2] Pruna. <http://www.pruna.com/>

[3] YouTube. <http://www.youtube.com/>

[4] Daum TV팟. <http://tvpot.daum.net>

[5] L. Guo, S. Chen, Z. Xiao, and X. Zhang. Analysis of

multimedia workloads with implications for Internet streaming. In Proc. of WWW, May 2005.

[6] Y. Tang, L. Sun, J. Luo, and Y. Zhong. Characterizing User Behavior to Improve Quality of Streaming Service over P2P Networks. PCM 2006, LNCS 4261, pp. 175-184

[7] L. Guo, E. Tan, S. Chen, Z. Xiao, O. Spatscheck, and X. Zhang. Delving into Internet Streaming Media Delivery: A Quality and Resource Utilization Perspective. IMC'06, October 25 - -27, 2006, Rio de Janeiro, Brazil.

[8] N. Daswani, H. Garcia-Molina, and B. Yang. Open Problems in Data-Sharing Peer-to-Peer Systems. Lecture Notes In Computer Science; Vol. 2572 Proceedings of the 9th International Conference on Database Theory table of contents, page 1-15

[9] H. Yu, D. Zheng, B.Y. Zhao, and W. Zhen. Understanding User Behavior in Large-Scale Video-on-Demand Systems. EuroSys'06, April 18 - -21, 2006, Leuven, Belgium.

[10] M. Kobayashoi, and K. Takeda, Information Retrieval on the Web, ACM Computing Surveys, Vol. 32, No. 2, June 2000

[11] 이일우, 박호진. P2P 서비스 응용 및 과금 기술 동향. ETRI 전자통신동향분석 제22권 제5호 2007년 10월

[12] Nowcom. <http://www.nowcom.co.kr>

[13] H. Yu, D. Zheng, B.Y. Zhao, and W. Zhen. Understanding User Behavior in Large-Scale Video-on-Demand Systems. EuroSys'06, April 18 - -21, 2006, Leuven, Belgium.

[14] M. Cha, H. Kwak, P. Rodrigues, Y. Ahn, and S. Moon. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. IMC'07, October 24-26, 2007, San Diego, California, USA.