

## 네트워크 자원효율 및 QoE 향상을 위한 콘텐츠 인기도 기반 무선 캐싱 기술

김근욱 · 홍준표\*

### Wireless Caching Techniques Based on Content Popularity for Network Resource Efficiency and Quality of Experience Improvement

Geun-Uk Kim · Jun-Pyo Hong\*

Department of Information Communication Engineering, Pukyong National University, Busan 48513, Korea

#### 요 약

최근 발표에 따르면, 2020년까지 모바일 데이터 트래픽이 현재의 11배까지 증가 될 것으로 예상된다. 그 중 비디오 트래픽이 70%를 차지할 것으로 예상되는 만큼 방대해지는 모바일 비디오 트래픽의 문제를 해결하기 위해서는 비디오 트래픽의 특성을 이해할 필요가 있다. 최근, 인기 있는 유튜브 비디오와 같은 일부 인기 있는 콘텐츠의 반복적인 요청으로 인해 네트워크 트래픽 오버헤드가 많이 발생한다. 만약 콘텐츠 인기도를 알고 인기 있는 콘텐츠를 미리 캐싱 할 수 있는 네트워크 노드를 구성한다면 이용자의 요청에 대해 캐싱 된 콘텐츠를 이용함으로써 네트워크 오버헤드를 줄일 수 있다. 장치 대 장치 통신, 멀티캐스트, 헬퍼를 통해 비디오 처리량이 기존의 방법보다 약 1.5배에서 2배의 이득이 향상되었다. 또한, 프리픽스 캐싱을 통해 기존의 방법보다 약 0.2배에서 0.5배의 재생 지연이 감소되었다. 본 논문에서는 무선 네트워크 환경에서 콘텐츠 인기도를 기반한 캐싱 기술에 대한 최신 연구를 소개 한다.

#### ABSTRACT

According to recent report, global mobile data traffic is expected to increase by 11 times from 2016 to 2020. Moreover, this growth is expected to be driven mainly by mobile video traffic which is expected to account for about 70% of the total mobile data traffic. To cope with enormous mobile traffic, we need to understand video traffic's characteristic. Recently, the repetitive requests of some popular content such as popular YouTube videos cause an enormous network traffic overheads. If we constitute a network with the nodes capable of content caching based on the content popularity, we can reduce the network overheads by using the cached content for every request. Through device-to-device, multicast, and helpers, the video throughput can improve about 1.5~2 times and prefix caching reduces the playback delay by about 0.2~0.5 times than the conventional method. In this paper, we introduce some recent work on content popularity-based caching techniques in wireless networks.

**키워드** : 캐싱, 콘텐츠 인기도, 멀티캐스트, 헬퍼, 무선 콘텐츠 전송

**Key word** : Caching, Content popularity, Multicast, Helper, Wireless content delivery

Received 10 April 2017, Revised 11 April 2017, Accepted 09 August 2017

\* Corresponding Author Jun-Pyo Hong(jp\_hong@pknu.ac.kr, Tel:+82-51-629-6227)

Department of Information Communications Engineering, Pukyong National University, Busan 48513, Korea

Open Access <https://doi.org/10.6109/jkiice.2017.21.8.1498>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서 론

최근 스마트폰 및 태블릿의 보편화와 고용량 미디어 콘텐츠의 높은 수요로 인하여 모바일 비디오 데이터 트래픽 양이 급격하게 증가하고 있다. 채널 용량을 증가시키기 위한 전통적인 방법들은 방대해진 비디오 트래픽을 처리하기에는 한계가 있다[1].

모바일 비디오 데이터 트래픽의 급격한 증가 문제를 해결하기 위해 캐싱 기술이 각광을 받고 있다. 최근 가장 인기 있는 비디오 스트리밍 사이트인 유튜브(youTube)와 관련된 연구에 따르면[2], 방대해진 트래픽에 대처할 수 있는 비디오 트래픽의 흥미로운 특징을 보여주고 있다. 대부분의 비디오 데이터 트래픽이 몇 개의 인기 있는 비디오 클립이 차지하고 있다는 것이다. 즉, 몇 개의 인기 있는 파일들은 제한된 시간 동안 사용자로부터 계속해서 요청되어지는 반면에 다른 파일들은 거의 요청 되어 지지 않는다.

유선 통신에서 캐싱은 검색된 데이터를 클라이언트에 가까이 보관할 수 있도록 한다[3]. 웹 환경에서 캐싱은 웹 서비스의 응답 시간을 줄이고 네트워크 혼잡을 완화 할 수 있다. 웹 브라우저(클라이언트), 웹 서버 및 웹 프록시를 포함하여 웹 서버에서 데이터를 캐싱 할 수 있는 몇 가지 방법이 있다. 클라이언트 측에서의 캐싱은 대부분의 기존 웹 브라우저에서 구현되었다. 이 방법을 통해 클라이언트가 동일한 위치에서 반복적으로 트래픽을 생성하는 것을 방지 할 수 있다. 서버 측에서의 캐싱은 클라이언트의 요청이 있을 때마다 원격 서버를 통해 제공 받는 대신 웹 서버가 클라이언트의 요청을 제공하기 위해 미리 데이터를 가져 오는 것을 허용한다. 하지만 두 가지 방법 모두 상당한 네트워크 성능 향상을 제공하지 못한다. 클라이언트 측 캐싱은 한 명의 클라이언트가 동일한 데이터를 연속적으로 가져 오도록 한다. 그러나 동일한 영역의 여러 명의 클라이언트가 짧은 시간 간격 동안 동일한 웹 서버에서 동일한 웹 파일을 다운로드하는 것을 제한할 수 없다. 서버 측 캐싱은 요청을 더 이상 전달하지 않는 문제만 완화하고 클라이언트가 경험할 수 있는 잠재적인 액세스 지연이나 가능한 혼잡을 완화하거나 해결할 수 없다.

기존의 무선 통신 방법과는 달리 캐싱을 이용하는 무선 네트워크는 콘텐츠 재사용(content reuse)의 특징을

이용한다. 특히, 인기 있는 동영상 분포에 대한 정보를 기반으로 모바일은 향후 요청에 대비하여 인기 있는 비디오를 미리 캐싱한다. 캐싱 된 데이터는 원격 송신기로부터 수신 할 필요가 없기 때문에 캐싱은 트래픽 오버헤드와 인기 있는 콘텐츠의 반복된 요청과 관련된 신호 처리뿐만 아니라 불안정한 채널로부터 오는 성능 저하를 줄여준다. 즉, 캐싱은 네트워크 트래픽을 스펙트럼 대역폭에 비해 저렴한 네트워크 리소스인(모바일) 메모리로 오프로딩(offloading)하여 시스템 성능을 향상시킨다.

본 논문에서는 각 네트워크 노드가 비디오 클립을 캐싱 할 수 있는 저장 공간을 가지고 있는 무선 네트워크 환경에서의 콘텐츠 인기도를 기반한 캐싱 기술에 대한 최근 연구를 소개한다. 주파수 자원 소모와 재생 지연을 최소화 하는 것과 같이 시스템 목표에 따라 여러 가지 다른 캐싱 전략이 있다. 후속 섹션에서는 서로 다른 환경에서 여러 가지 캐싱 전략에 대해 논의하고 캐싱 전략을 비교한다.

## II. 네트워크 자원 효율 향상을 위한 헬퍼 노드를 활용한 캐싱

방대해져가는 모바일 비디오 트래픽을 해결하고 시스템 스펙트럼 효율을 달성하기 위한 방법은 소형 기지국을 배치하여 셀 크기를 축소하는 것이다. 그러나 셀 크기를 축소시킴으로써 늘어난 기지국의 요청에 대응하기 위해 백홀(backhaul)의 부담이 증가하게 된다. 백홀 용량의 부담과 비디오 트래픽의 지연을 줄이기 위해 매크로 셀과 피코-셀 그리고 펌토-셀이 병합 된 이종 네트워크(Heterogeneous cellular network, HCN)가 최근 많은 관심을 받고 있다.

백홀 용량을 소형 셀 액세스 포인트의 저장 용량으로 대체하는 것으로 구성 된 펌토 캐싱이 [1, 4, 5]에서 논의되었다. 간단하게, 오직 무선 연결만 가능하고 큰 스토리지 용량을 가지고 있는 특별한 노드를 셀에 배치하여 사용량이 적은 시간대에 서비스를 제공하는 셀룰러 기지국으로 대체 될 수 있다. 이러한 헬퍼(helper)라고 불리는 특별한 노드들이 조밀하게 배치되어 사용자가 하나 이상의 펌토 기지국과 통신 할 수 있는 펌토 셀 네트워크에서 지연을 최소화 하기위한 무선 분산 캐싱 인

프라를 구성한다. 그림 1은 하나의 셀로 구성된 시스템 구조를 보여준다.

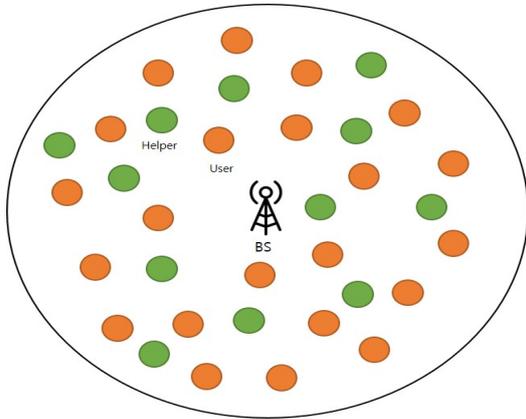


Fig. 1 An example of the single-cell system model.

셀룰러 기지국(BS)은 전체 비디오 파일 라이브러리를 가지고 있고 셀 내의 모든 사용자의 요청을 지원할 수 있다. 헬퍼는 전체 비디오 파일 라이브러리 중 일부의 파일을 가지고 있다. 무선 사용자들은 헬퍼로부터 콘텐츠를 다운로드하거나 헬퍼에 저장 되지 않은 파일은 기지국으로부터 직접 전송 받는다.

[1, 4]는 콘텐츠 배치 문제, 즉, 주어진 특정한 네트워크와 파일 인기도 분포에서 어떤 파일을 어떤 헬퍼에 캐싱 할 것인가 하는 문제를 다루었다. 특히, 콘텐츠 배치 문제에서 기대되는 총 파일 다운로드 지연을 최소화하기를 원한다. 다음과 같이 두 가지 방법의 콘텐츠 배치 문제를 고려하였다.

1) 코딩되지 않은 콘텐츠 배치 문제, 오직 완전한 파일(complete files)만이 전체 평균 지연을 최소화하기 위해 헬퍼 캐시에 저장되는 것이다. 최적의 파일 배치 방법은 NP-hard 문제로 덜 복잡한 탐욕 알고리즘(greedy Algorithm)을 제공하였는데 사용자와 연결된 최대 헬퍼의 수  $d$ 에 따라서 전체 평균 지연 시간은  $1 - (1 - 1/d)^d$  로 축약 되는 것을 보여주었다.

2) 코딩 콘텐츠 배치 문제에서 인트라-세션 코딩(Intra-session coding)을 고려하였다. 간단하게, 비디오 파일은 랩터 코드(Raptor codes, [6])에 의해 실현 가능한 이상적인 손실 없는 MDS 코드로 인코딩 되었다고 가정하였다. 이와 같은 경우, 전체 파일을 저장 할 필요

없이 비디오 파일의 일부분만을 헬퍼에 저장하므로 비디오 파일을 구성하는 비트의 개별적인 ID는 관련이 없으며 중요한 것은 주어진 파일의 패리티 심볼을 각 사용자가 헬퍼로부터 검색하는 것이다. 코딩 콘텐츠를 배치를 통해 헬퍼의 캐시를 더욱 효율적으로 사용할 수 있고 근처의 헬퍼로부터 더욱 다양한 비디오 파일을 다운로드 할 수 있다.

콘텐츠 배치 문제는 주어진 파일 인기도 분포, 헬퍼 저장 용량 및 네트워크 토폴로지에 대해 평균 다운로드 지연을 최소화 하는 방법으로 파일을 어떻게 헬퍼 캐시에 배치하는 문제로 수식화 하였다. 사용자의 파일 인기도와 네트워크 연결 그래프가 알려져 있다고 가정하고 코딩되지 않은 배치와 코딩 된 배치의 수치적인 결과를 보여주었다.

- 헬퍼 수가 증가함에 따라 코딩되지 않은 경우와 코딩 된 경우의 성능 차이가 커진다. 제안된 시나리오의 배치 방식으로 약 1.5배에서 2배의 이득을 얻었다. 분명하게, 헬퍼의 수가 더 많을 때 더 많은 이득을 얻을 수 있다.

- 사용자 수가 증가함에 따라 평균 다운로드 속도가 느려진다. 헬퍼의 수가 고정되어 있는 상태에서 유저의 수가 증가함에 따라 헬퍼로부터 원하는 비디오 파일을 다운로드 하는 사용자 수는 적어진다.

최근에는 확률적 무선 캐싱 헬퍼 네트워크 모델이 [5]에서 논의되었다. 사용자는 콘텐츠 인기도 Zipf 분포에 따라 무작위로 하나의 콘텐츠를 요청한다. 요청된 콘텐츠를 캐싱한 헬퍼들 중, 순간 수신 신호 전력(instantaneous received signal power)이 가장 큰 헬퍼로부터 서비스를 받을 때, 평균 콘텐츠 전송 성공 확률과 콘텐츠  $i$ 의 저장 확률  $p_i$ 가 클수록 더욱 많은 헬퍼들이 콘텐츠  $i$ 를 저장한다. 즉, 사용자는 특정 콘텐츠  $i$ 에 대해 채널 선택 자유도가 커지게 된다. 반대로 다른 콘텐츠들에 대한 채널 선택 자유도는 감소한다. 따라서, 각 파일 인기도, 무선 채널 특성, 캐싱 헬퍼의 노드 수 등에 따라 적절하게 채널 선택 자유도를 제공해 주는 것이 필요한데 이는 확률 값을 적절하게 조정함으로써 가능하다. 최적 캐싱 확률 값은 [5]의 최적화 문제의 솔루션을 도출함으로써 찾을 수 있다. 제안된 확률 저장 기법이 기존의 다른 저장 기법들에 비해 채널 선택 다이버시티 최적화를 통해 더 우수한 성능을 가짐을 보여준다.

### III. 네트워크 자원 효율 향상을 위한 D2D 통신에서의 캐싱

각 사용자가 직접 데이터를 캐싱하고 또한 주변 사용자와 캐싱된 데이터를 공유할 수 있는 무선 장치 대 장치 (Device to Device, D2D) 네트워크에서 D2D 링크를 통한 로컬 통신은 기지국에서 대용량 트래픽을 사용자에게 오프로딩하여 스펙트럼 재사용을 향상시킨다. 주어진 파일 인기도 분포에서 D2D 통신을 활용한 캐싱 전략이 [1, 7, 8]에서 논의되었다.

사용자의 분산 캐싱 전략과 활성화된 D2D 링크의 평균 수(average number of active links)를 최대화하기 위한 D2D 통신 범위  $r$ 은 [1, 7]에서 논의되었다. 또한 최근에는 최적의 파일 배치 순서와 D2D 처리량과 아웃티지 확률(Outage Probability) 사이의 트레이드오프(tradeoff)가 [8]에서 논의되었다.

간단하게, 각 사용자는 단일 비디오 클립을 캐싱할 수 있고 셀 간 간섭은 없고 총  $K$  사용자가 셀에 균일하게 분포한다고 가정한다. 각 사용자는 한 번에 최대 하나의 비디오 클립을 요청하며 비디오 클립은 인기도 분포에 따라 요청된다. 특히, 사용자가  $i$  번째 인기 있는 비디오 클립을 요구하는 확률  $f_i$ 는 아래에 주어진 Zipf 분포에 의해 모델링된다고 가정한다.

$$f_i = \frac{i^{-\gamma_r}}{\sum_{j=1}^m j^{-\gamma_r}}, \quad 1 \leq i \leq m. \quad (1)$$

총  $m$  개의 후보 비디오 클립이 있고 Zipf 지수  $\gamma_r$ 는 인기도에 따른 요청 확률의 감소율을 결정한다. 즉,  $\gamma_r$ 이 증가할수록 몇 개의 인기 있는 비디오 클립에 더 많은 인기가 집중되는 것이다. D2D 통신을 이용하는 캐싱은 그림 2과 같이 동작한다.

- 상황 1 : 사용자가 이미 저장 공간에 캐싱된 비디오 클립을 원하면 비디오 클립을 즉시 사용할 수 있다. 이 같은 경우를 본인-요청(Self-Request)이라고 한다.

- 상황 2 : 사용자가 D2D 통신 범위 내에 있는 다른 사용자에게 캐싱된 비디오 클립을 원할 경우 해당 사용자와 D2D 통신 후 해당 비디오 클립을 사용할 수 있다.

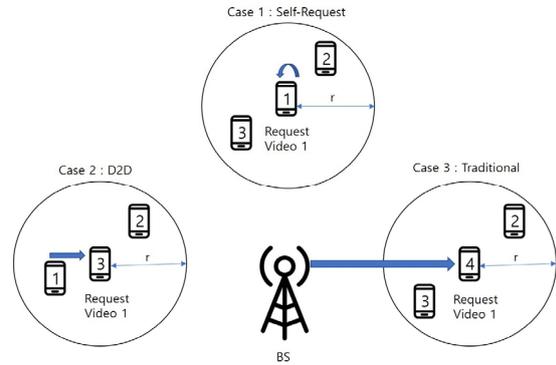


Fig. 2 D2D network with caching.

- 상황 3 : D2D 통신 범위 내에 요청된 비디오 클립을 캐싱한 사용자가 없으면 비디오 클립은 기지국과 통신한 후에 사용할 수 있다.

이 통신 전략을 사용할 때, 스펙트럼 재사용을 최대화하기 위한 두 가지 주요 논점이 있다. 첫 번째는 D2D 통신의 범위  $r$ 에 관한 것이다.  $r$ 이 커질수록 사용자가 D2D 통신을 통해 주변 사용자로부터 요청된 클립을 수신할 확률은 증가하지만 사용 가능한 D2D 링크의 최대 개수는 상호 간섭으로 인해 감소한다. 이 두 가지 측면의 균형을 맞추기 위해서는 적절한  $r$  값을 결정해야 한다.

두 번째는 캐싱 전략에 관한 것이다. 각 사용자는 가까운 사용자와 겹치지 않는 비디오 클립을 캐싱하는 것이 바람직하다. 사용자 위치에 기반하여 기지국에서의 중앙 집중식 비디오 클립 할당(centralized allocation)은 불필요한 중첩을 회피함으로써 최적의 성능을 제공할 수 있다. 그러나 중앙 집중식 캐싱 및 사용자 위치에 대한 신호 전송은 큰 오버헤드를 초래하고 또한 만약 사용자가 이동할 경우 재할당(re-allocation)이 필요하다. 이러한 이유로 모바일 사용자가 있는 네트워크 환경에서는 중앙 집중식 할당방식은 사용할 수 없고 적절한 분산 캐싱 전략이 고려된다. 분산 캐싱 전략에서 각 사용자는 지수  $\gamma_c$ 를 가지고 있는 Zipf 분포에 따라서 파일을 랜덤하고 독립적으로 캐싱한다. 이 전략에서  $\gamma_c$ 는  $\gamma_r$ 과 같을 필요가 없는 변수이기 때문에 D2D 링크의 수를 최대화 시키는 최적의  $r$  및  $\gamma_c$ 을 찾아야 한다. 또한, 사용자가 요청한 비디오 클립을 근처의 사용자로부터 찾는 확률은  $r$ 뿐만 아니라 캐싱 전략에도 의존하기 때

문에 최적의  $r$  및 캐싱 전략은 서로 관련(correlate)되어 있다.

[1, 7]은 각 사용자가 Zipf 분포 지수  $\gamma_c \geq \gamma_r$ 에 따라 비디오 클립을 독립적으로 캐싱 하는 랜덤 캐싱 전략을 제안 하였다. 이는 최적의  $r$  및 캐싱 전략을 찾는 문제를 최적의  $r$  및  $\gamma_c$ 을 찾는 문제로 단순화 시킨다. 이 캐싱 전략의 최적성은 도출되지 않았지만 적절한 값의  $\gamma_c$ 를 갖는 랜덤 캐싱 성능은 중앙 집중식 성능과 근사하고 기존 캐싱 방법에 비해 비디오 처리량이 약 1.5 ~ 2 배 향상 될 수 있음을 수치적으로 나타냈다.  $r$ 과  $\gamma_c$ 에 대한 수치 시뮬레이션 결과의 주요 관측 사항들은 다음과 같이 요약 될 수 있다.

- 인기 분포 지수  $\gamma_r$ 가 커질수록 D2D 통신의 최적 범위  $r$ 이 감소한다. 이는  $\gamma_r$ 이 높아질수록 인기도가 몇 개의 인기 있는 비디오 클립에 집중됨으로 D2D 통신을 필요로 하는 필요한 인접한 사용자 수는 적어진다.
- 최적의  $\gamma_c$  및  $r$  값에 따라 분산 캐싱 전략의 스펙트럼 재사용 성능(또는 활성화 된 D2D 링크의 평균 수)은 이상적인 중앙 집중식 캐싱 전략(전체 비디오 클립 개수  $m$ 과 사용자 수  $n$ 의 관점에서)의 성능 패턴과 비슷한 결과를 보인다. 특히 사용자 수  $n$ 이 크지 않은 경우 분산 전략과 중앙 집중식 전략의 성능은 비슷하지만  $n$ 이 커질수록 랜덤 캐싱 전략에서 캐싱 된 파일 간의 중복 가능성이 높아지므로 성능 향상은 중앙 집중식 캐싱 보다 적다.

[8]은 더욱 현실적인 시스템 모델을 가정하여 D2D 처리량과 아웃티지 확률(Outage Probability) 사이의 트레이드 오프(trade off)를 분석하였다. 기지국에서 모바일 이용자에게 전송하는 주파수 대역은 2.1 GHz long-term-evolution, LTE의 표준 주파수)의 주파수를 사용하고 D2D 통신의 주파수 대역은 2.45 GHz(ISM bands)와 38 GHz(Unlicensed bands)를 사용한다고 가정하였다. 여기서 파일은 가장 효율적인 전송 방식으로 이루어진다. 전송에 필요한 거리가 가깝다면 mm-wave 주파수 밴드를 이용하는 것이 가장 효율적이다. 그러나 mm-wave를 이용한 전송은 벽이나 사람의 몸과 같은 장애물의 영향을 크게 받으므로 전송이 쉽지 않다. mm-wave를 이용한 전송이 실패한다면 남은 방법 중의 최선책은 2.45 GHz 밴드를 이용한 D2D 전송이다. 만약 2.45 GHz 밴드를 이용한 전송이 실패하거나 D2D 통

신이 동작하는 범위 내에 이용자가 요청한 파일이 없다면 기지국을 통해서 파일이 전송된다. 수치 시뮬레이션을 평가할 때는 기존의 유니캐스트 방식 뿐만 아니라 멀티캐스트를 이용한 방식도 함께 평가하였다. 수치 시뮬레이션 결과를 통해 D2D 통신을 활용한 방식이 다른 방식들에 비해 현저히 높은 비디오 처리량을 보여주었다. 다른 방식의 경우 시스템 모델에서 다이버시티의 영향을 생각하지 않았기 때문에 채널 상태에 따라 아웃티지 확률이 영향을 크게 받는다.

#### IV. 네트워크 자원 효율 향상을 위한 멀티캐스트를 활용한 캐싱

비디오 트래픽과 같이 데이터가 큰 동일한 콘텐츠를 멀티캐스트를 이용하여 한 번의 신호 전송만으로 다수의 사용자에게 제공 할 때 네트워크의 가용 대역폭을 효율적으로 활용할 수 있다. 직관적으로, 멀티 캐스트는 사용자가 동일한 콘텐츠 파일에 대한 요청을 동시에 생성할 때 효과적이다. 기존 유니캐스트 캐싱 방식에 비해 멀티캐스트 캐싱 방식의 우위는 (1) 동일한 콘텐츠에 대한 사용자 요구가 높고 (2) 콘텐츠에 대한 사용자 요청이 지연에 견딜 수 있는 경우에 매우 두드러진다. 이러한 시나리오는 스포츠 경기, 콘서트와 같은 수천 명의 참석자가 공동 콘텐츠에 대해 관심이 있을 때 더욱 일반적이다[9].

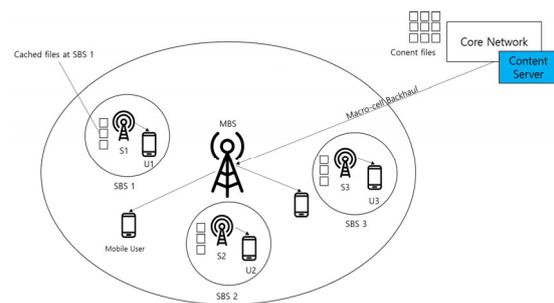
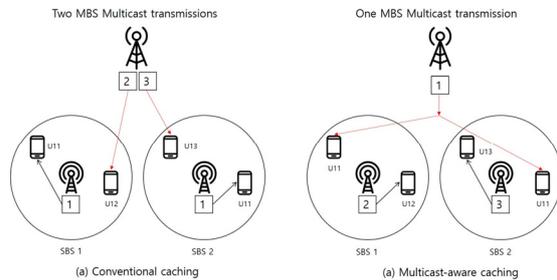


Fig. 3 Multicast-aware caching network with HCN.

이종 셀룰러 네트워크(Heterogeneous cellular network, HCN)에서 매크로 셀과 펌토 셀 기지국 간의 협력을 통해 멀티 캐스트 서비스를 지원할 수 있는 프로

토콜이 [10]에 논의되었다. 최근에는 캐싱과 멀티 캐스트를 결합한 캐싱 전략과 그에 따른 최소화 할 수 있는 에너지 비용과 이득이 [11]에서 논의되었다.

그림 3과 같이 모바일 사용자의 서비스에 대해 캐싱 및 멀티 캐스트를 지원하는 이중네트워크(HCN) 모델을 고려하였다. 짧은 시간의 윈도우(window) 동안 생성된 동일한 콘텐츠 파일에 대한 요청은 해당 윈도우가 만료 될 때 단일 멀티캐스트 전송을 통해 제공된다(batching multicast[12, 13]). 사용자 경험 지연을 줄려면 이 윈도우의 기간을 가능한 작게 해야 한다.



**Fig. 4** An example with two SBSs and three files when (a) conventional and (b) multicast-aware caching is applied.

그림 4는 멀티 캐스트가 캐싱 정책의 효율성에 미치는 영향을 보여준다. 그림 (a)에서 파일 1의 동일한 요청은 각 소형 셀 기지국(small-cell base stations, SBS)에서 사용자에게 제공된다. 파일 2와 파일 3의 요청에 대해서는 매크로 셀 기지국(macro-cell base stations, MBS)를 통해 각 유니캐스트(unicast) 송신을 통해 전송된다. 그림 (b)에서는 캐싱 전략을 다르게 하였다. 소형 셀 기지국 1에서 파일 2가 요청한 사용자에게 전송되고 소형 셀 기지국 2에서 파일 3을 요청한 사용자에게 전송된다. 그리고 파일 1에 대한 동일한 요청은 매크로 셀 기지국에서 하나의 멀티캐스트 송신을 통해 전송되었다. 즉, 파일 인기도 분포에 따라 파일을 저장하지 않아도 멀티 캐스트를 잘 활용한다면 더욱 효율적일 수가 있다.

[10]은 RS 파운데인 코드(Reed-Solomon Fountain code)를 이용한 협력 멀티캐스트 프로토콜을 소개한다. RS 파운데인 코드는(Reed-Solomon) 다음 이상적인 분수 코드의 두 가지 속성을 이용하였다.

- 유한 소스 데이터에서 무한대로 인코딩 된 데이터

를 생성 할 수 있다.

- 소스 데이터의 길이가  $k$ 라면, 수신기는 효율적인 디코딩 프로세스로  $k$ 개의 인코딩 된 데이터로부터 소스 데이터를 재구성 할 수 있다.

이론적으로 RS 코드는 분수 코드의 확장 버전으로 생각 할 수 있다. 소스 메시지의  $k$ 개의 심볼을 RS 코드로 인코딩하고 인코딩 된 데이터의  $n$ 개의 심볼을 얻는다면, 소스 메시지는  $n$ 개의 인코딩 된 심볼의  $k$ 개로부터 복구 될 수 있다. RS 파운데인 코드의 전체 인코딩 과정은 다음과 같이 두 단계로 구성된다. 먼저, 원래의 심볼은 RS 파운데인 코드로 인코딩된다. 그 다음, 다수의 파운데인 코드 워드(words)에 대응하는 심볼은 수신기가 정확하게 수신 된 심볼을 인식 할 수 있도록 보정 코드(correcting code)나 검출 코드(detecting code)로 인코딩된다. 수신 시에 패키지가 수신 된 후 에러 정정 또는 검출 코드의 디코딩이 먼저 수행된다. 다음 인코딩 된 패키지들이 정확하게 디코딩 된 후에, 원래 정보가 복구 될 수 있다. 수치 시뮬레이션 결과를 통해 제안 된 RS 파운데인 코드 협동 시스템은 파운데인 코드를 사용하지 않은 시스템보다 약 2배 가까운 채널 용량을 갖는 것을 보여주었다.

[11]은 전체 에너지를 최소화하는 멀티 캐스트-인식 캐싱 전략을 고안하기 위한 일반적인 멀티 캐스트-인식 캐싱 전략의 최적화 문제(Multicast-Aware caching problem, MACP)는 (1) 서로 다른 캐시 크기와 전송 비용 파라미터를 가지고 있는 기지국의 이질성, (2) 서로 다른 콘텐츠 파일을 요청함으로써 생기는 요청 패턴의 변화를 고려하였다. 멀티 캐스트-인식 캐싱 전략의 최적화 문제는 NP-hard 문제로 기존의 캐싱 방식 보다 성능이 크게 향상되는 멀티 캐스트-인식 캐싱 전략의 최적화 문제의 차선책인 간단한 발견적 알고리즘에 대해 설명한다. 제안된 알고리즘은 모든 캐시가 비어있다는 가정으로 시작한다. 각 반복마다, 지연을 최소화 할 수 있도록 캐시에 탐욕적으로(greedily) 파일을 배치하고 모든 캐시가 가득 차게 되면 종료한다.

기존 캐싱 전략에 비해 제안 된 멀티 캐스트 인식 캐싱 알고리즘으로 얻은 에너지 절감 효과를 수치적으로 평가하였다. 전반적으로, 기존의 캐싱 전략(인기도를 고려한 캐싱)에서 멀티 캐스트 인식으로 향상 된 캐싱 전략이 실제로 에너지를 줄일 수 있고 수요가 방대하고 사용자의 콘텐츠 요청이 지연에 견딜 수 있는

경우에 더 많은 이점이 생긴다는 것을 알 수 있다.

- 멀티 캐스트 기간의 영향: 직관적으로, 멀티 캐스트는 사용자 간의 콘텐츠를 액세스하는데 상당한 동시성이 있을 때, 즉, 동일한 파일에 대한 많은 요청이 멀티 캐스트 기간 내에 자주 나타나는 경우에 효과적이다. 제안된 멀티캐스트 인식 캐싱 기법은 지속적으로 기존의 멀티캐스트 전략보다 뛰어난 성능을 보였으며 멀티 캐스트 기간이 증가함에 따라 이득도 함께 증가(최대 31%)하였다. 비교적 작은 멀티 캐스트 기간 값에 대해서조차도 제안된 방식들은 기존의 멀티캐스트 전략에 비해 상당한 이득을 얻었다.

- 캐시 크기의 영향: 캐시 크기를 늘리면 매크로 셀 기지국의 참여 없이 더 많은 요청이 매크로 셀 기지국의 참여 없이 충족되므로 제안된 전략의 에너지 비용이 감소하였다. 멀티 캐스트 인식 전략은 많은 유니 캐스트 전송 대신 하나의 멀티 캐스트 신호를 통해 많은 요청을 처리하기 때문에 유니캐스트 방식은 나머지 방식에 비해 가장 큰 에너지 비용(최대 35% 차이)을 초래한다. 제안된 멀티 캐스트 인식 캐싱 기법은 지속적으로 인기도를 인식하는 기존의 캐싱 기법보다 성능이 뛰어나고 캐시 크기가 증가함에 따라 이득이 증가(최대 20%)한다.

experience)는 재생 지연(playback delay)에 가장 크게 좌우되기 때문에 비디오는 안정적이고 순차적으로 사용자에게 전달되어야한다. 즉, 느린 전송으로 인해 재생이 느려지지(retard) 않으면 비디오 전송에서 빠른 전송 속도는 중요하지 않다. 따라서, 이 섹션에서 캐싱의 목적은 전송에 약간의 지연이 있을 수 있음에도 불구하고 사용자는 비디오 재생을 할 때 어떤 지연(delay)도 경험하지 않도록 하는 것이다.

D2D 통신이 없는 다운링크 네트워크에서 사용자가 인지하는 비디오 재생 지연을 최소화하기 위해 사용자 프리픽스(prefix) 캐싱 전략이 [14]에서 논의되었다. 프리픽스는 비디오의 초기 부분을 의미하며 프리픽스의 크기는 시스템 환경에 따라 다를 수 있다. 재생 지연을 최소화하기 위해 사용자는 사용자의 개인적 선호도와 일치하는 범위에 맞춰서 일부 인기 있는 동영상의 초기 부분(즉, 프리픽스)을 캐싱한다. D2D 통신이 없기 때문에 각 사용자는 시청 이력 정보(viewing history) 및 개인적 선호도(personal preferences)에 따라 결정되는 개인적 인기 분포(personal popularity distribution)에 따라 비디오 클립을 캐싱한다. 즉, 근처 사용자의 캐시 메모리와 겹치는 것을 피하기 위해 사용되는 랜덤 캐싱(random caching)을 수행할 필요가 없다. 사용자가 상단  $i$  중에서 인기도가 높은 비디오를 요청할 확률은

$$p_i = 1 - e^{-\left(\frac{i}{\lambda}\right)^\kappa}, \quad (2)$$

### V. 비디오 스트리밍 환경에서 QOE 향상을 위한 캐싱

이전 섹션에서 네트워크 스펙트럼 재사용(network spectrum reuse)을 최대화하기 위한 캐싱 전략과는 달리 이번 섹션에서는 사용자가 받는 서비스 품질을 개선하기 위한 캐싱 전략에 대해 설명한다. QoE(quality of

$\lambda > 0$  and  $0 < \kappa \leq 1$  은 각각 크기(scale) 및 모양 매개변수(shape parameters)이다. 그런 다음, 캐싱 전략과 캐싱 통신은 다음과 같이 동작한다.

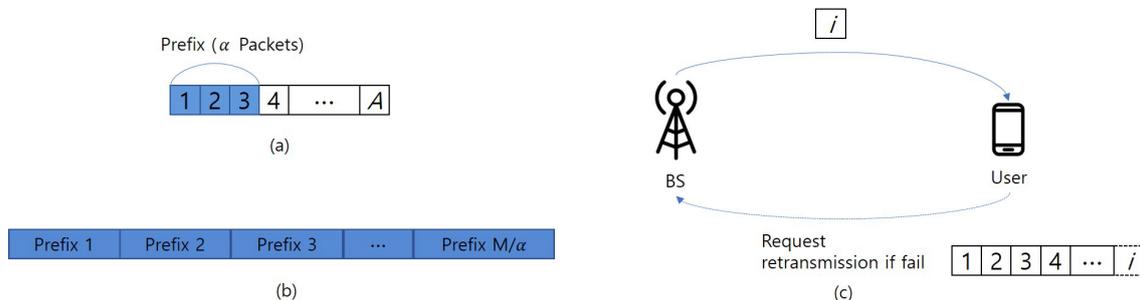


Fig. 5 System model. (a) Structure of video clip (b) User prefix caching (c) Wireless streaming.

• 그림5-(a)에서, 각 비디오는 동일한 크기의 패킷으로 분할되고 패킷 단위의 순서대로 전송된다. 처음  $\alpha$  패킷은 프리픽스(그림 5에서  $\alpha = 3$ )로서 정의되고 일부 인기 비디오의 프리픽스는 사용자에게 캐싱 된다.

• 그림5-(b)에서, 각 사용자는 최대  $M$  개의 패킷을 캐싱 할 수 있으며 저장 공간은 단일 비디오 크기 ( $M \geq A$ )보다 큰 것으로 가정한다. 무한히 많은 캐싱 후보가 있지만 저장 공간은 제한적이므로 사용자는 저장 공간이 허용되는 만큼 비디오의 프리픽스를 선택적으로 캐싱한다.

• 그림5-(c)는 기지국에서 사용자로의 무선 패킷 전송을 보여준다. 전송 및 디코딩은 패킷 단위로 수행된다. 요청된 비디오의 패킷은 패킷 인덱스 순서대로 요청 사용자에게 전달되어야한다. 전송된 패킷  $i$ 가 사용자에서 성공적으로 디코딩된 경우에만 기지국은 패킷  $i + 1$ 을 전송할 수 있다. 사용자가 캐싱 되지 않은 비디오 중 하나를 요청하면 비디오의 패킷은 첫 번째 패킷부터 순서대로 전송된다. 반면, 사용자가 프리 픽스가 캐시된 비디오를 요청하면 요청된 비디오의 패킷은  $\alpha + 1$  인덱스가 있는 패킷부터 순차적으로 전송된다.

만약 사용자 쪽에서 패킷 디코딩 실패가 있는 경우 기지국은 사용자가 성공적으로 디코딩 될 때까지 패킷을 재전송한다. 디코딩 실패는 확률  $P_e$ 에 따라 독립적으로 발생한다. 패킷은 인덱스 순으로 재생되기 때문에 패킷 재전송은 재생 지연을 초래한다. 사용자가 어떤

프리픽스가 캐시된 비디오를 요청하면 기지국은  $\alpha + 1$  인덱스를 가진 패킷을 전송하면서 요청된 비디오를 재생하기 시작한다. 이런 프리픽스 재생 시간은 사용자가 재생 지연을 인지하지 않고 최대  $\alpha$  재전송으로부터 야기되는 지연을 보완하기 위한 버퍼로써 동작하면서 남아있는 패킷  $A - \alpha$ 를 수신하기 위한 여분의 시간 (extra-time)을 제공한다. 비디오 스트리밍에서 프리픽스 캐싱의 역할과 함께 중요한 문제는 평균 재생 지연을 최소화하는 프리픽스의 최적 크기  $\alpha$ 를 찾는 것이다.

제한된 저장 공간  $M$  [packets]을 사용하면 프리픽스 크기  $\alpha$ 와 캐싱된 비디오 수  $(1/\alpha)M$ 의 트레이드오프 (tradeoff)가 있다.  $\alpha$ 가 증가하면 캐싱된 비디오의 예상 지연은 줄어들지만 캐싱된 비디오의 수  $(1/\alpha)M$ 이 줄어들기 때문에 캐싱된 데이터를 사용할 수 없는 확률이 높아진다. 재생 지연을 최소화하려면 프리픽스 크기와 캐싱된 비디오 수 사이에 균형을 유지해야한다. 최적 프리픽스 크기는  $\alpha^* = AP_e$ 로 정의한다. 재생 지연이  $\lambda, \kappa, M, P_e$  및  $A$ 와 같은 많은 시스템 매개 변수에 따라 달라지는 것과는 달리 최적의 프리픽스 크기는  $A$  및  $P_e$ 의 영향만 받는다.

그림 6을 통해  $M = 4A$ 와  $M = 100A$ 의 두 가지 경우 모두  $\alpha = AP_e = 3400$ 에서 지연이 최소화되고  $M$ 에 관계없이 최적의 프리픽스 크기를 이용한 캐싱이 다른 값의 프리픽스 캐싱보다 우위에 있음을 알 수 있다. 또한, 기존의 방법보다 약 0.2배에서 0.5배의 재생 지연이 감

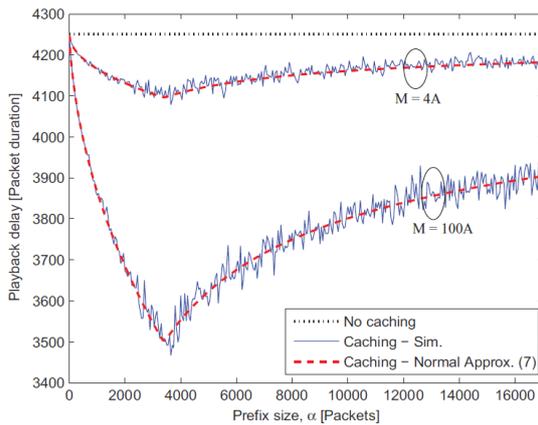


Fig. 6 Average playback delay versus the prefix size  $\alpha$  for some storage spaces ( $P_e = 0.2, \kappa = 0.516, \lambda = 11910$ ).

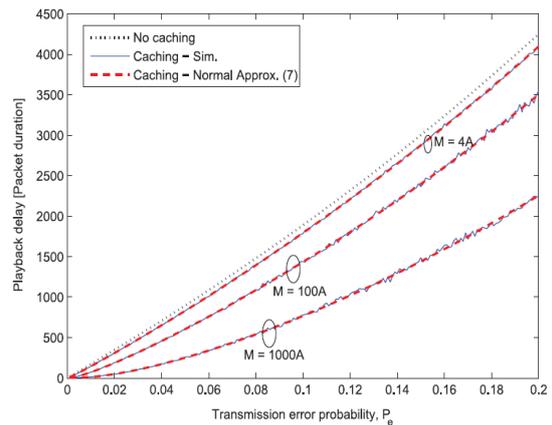


Fig. 7 Average playback delay versus transmission error probability  $P_e$  for some storage spaces ( $\kappa = 0.516, \lambda = 11910$ ).

소되었다.

그림 7은 평균 재생 지연이  $P_c$ 의 단조 증가 함수임을 명확하게 보여준다. 또한 저장 공간  $M$ 이 증가함에 따라 캐싱으로 인한 지연 감소 이득이 증가하는 것으로 나타났다.

## VI. 결 론

캐싱은 획기적인 통신 성능 향상을 제공하는 유망한 기술 중 하나로서 많은 주목을 받고 있다. 본문에서 소개된 헬퍼 노드와 무선 사용자로 구성된 웹토 셀 네트워크의 파일 저장 방법에 따라 코드화 된 콘텐츠 배치와 코드화 되지 않은 콘텐츠 배치 방식으로 약 1.5배-2배의 데이터 전송 속도 증가를 얻었고 헬퍼의 수가 더 많을 때 더 많은 이득을 얻을 수 있는 것을 보여주었다.

또한, D2D 통신을 사용하여 중앙집중식 및 랜덤 캐싱(deterministic & random caching) 전략에 대해 분석하고 기존 캐싱 방법에 비해 비디오 처리량이 약 1.5 ~ 2배 향상 될 수 있음을 보여주었다.

기존의 캐싱 방법에 멀티 캐스트를 추가적으로 사용하여 사용자에게 인기 있는 콘텐츠를 제공하는 캐싱 전략의 발견적 알고리즘(Heuristic Algorithm)을 분석하고 기존 캐싱 방식보다 약 19%의 데이터 전송 속도가 향상되었음을 보여주었다.

비디오 스트리밍 서비스의 평균 재생 지연을 최소화하기 위한 프리픽스 캐싱 전략을 제안하고 최적의 프리픽스 크기는 캐싱을 위한 저장 공간 및 인기도 분포의 파라미터와는 독립적이고 오직 비디오 클립 내의 패킷 수와 전송 실패 확률의 곱으로 나타났다. 프리픽스 캐싱 전략을 통해 약 0.2배에서 0.5배의 재생 지연이 감소되었다.

캐싱에 대한 작업은 초기 단계이므로 아직 해결해야 할 많은 문제가 있다. 많은 논문에서 분석상의 편의를 위해 동일한 크기를 가진 비디오 클립과 전송 전력, 모든 이용자에게 대해 동일한 파일 인기도 분포 등 여러 가지 가정을 두고 연구를 진행하였다. 향후 작업을 위한 한 가지 중요한 방향은 다양한 무선 전송 기술에 캐싱을 적용하고 보다 실질적인 무선 채널을 고려하는 것이 캐싱 연구의 다음 단계가 될 것이다.

## ACKNOWLEDGMENTS

This work was supported by a Research Grant of Pukyong National University (2016 year)

## REFERENCES

- [ 1 ] N. Golrezaei, A. F. Molisch, A. G. Dimakis, G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 142-149, Apr. 2013.
- [ 2 ] X. Cheng, J. Liu, and C. Dale, "Understanding the characteristics of internet short video sharing: A YouTube-based measurement study," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1184-1194, Aug. 2013.
- [ 3 ] B. Li, M. J. Golin, G. F. Italiano, X. Deng, K. Sohraby, "On the optimal placement of Web proxies in the Internet," in *Proc. 8th IEEE International Conference on Computer and Communications*, pp. 1282-1290, Mar. 1999.
- [ 4 ] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402-8413, Dec. 2013.
- [ 5 ] S. H. Chae and W. Choi, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6626-6637, Oct 2016.
- [ 6 ] A. Shokrollahi, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551-2567, Jun. 2006.
- [ 7 ] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Basestation assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no.7, pp.3665-3676, Jul. 2014.
- [ 8 ] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE Journal of Selected Areas in Communications*, vol. PP, no. 99, pp. 1, 2015.
- [ 9 ] J. Erman and K. K. Ramakrishnan, "Understanding the

- super-sized traffic of the super bowl,” in Proc. *ACM Internet Measurement Conference*, Nov. 2013, pp. 353-360.
- [10] X. Xie, B. Rong, T. Zhang, W. Lei, “Improving physical layer multicast by cooperative communications in heterogeneous networks,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 58-63, Jun. 2011.
- [11] K. Poularakis, G. Iosifidis, V. Sourlas and L. Tassiulas “Exploiting caching and multicast for 5G wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2995-3007, Apr. 2016.
- [12] M. A. Maddah-Ali, U. Niesen, “Fundamental limits of caching,” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856-2867, May 2014.
- [13] V. Tokekar, A. K. Ramani, and S. Tokekar, “Analysis of caching policy in view of user renegeing in VoD system,” in Proc. *IEEE India conference*, Dec. 2005, pp. 399-403.
- [14] J.-P. Hong and W. Choi, “User prefix caching for average playback delay reduction in wireless video streaming,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 377-388, Aug. 2016.



**김근욱(Geun-Uk Kim)**

2016년 8월 : 부경대학교 정보통신공학과 (공학사)  
2016년 9월 ~ 현재 : 부경대학교 대학원 정보통신공학과  
※관심분야 : 무선통신, 무선캐싱



**홍준표(Jun-Pyo Hong)**

2014년 KAIST 전기및전자공학과 박사  
2014년~2015년 KAIST 전자정보연구소 연수연구원  
2015년 ETRI 통신인터넷연구소 연구원  
2015년~현재 국립부경대학교 정보통신공학과 조교수  
※관심분야 : 정보통신 보안, 차세대 이동통신 시스템, wireless caching networks 등