JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 26 No. 1, pp. 111-118, January 2021
https://doi.org/10.9708/jksci.2021.26.01.111

Device Caching Strategy Maximizing Expected Content Quality

Minseok Choi*

*Assistant Professor, Dept. of Telecommunication Engineering, Jeju National University, Jeju, Korea

[Abstract]

This paper proposes a novel method of caching contents that can be encoded into multiple quality levels in device-to-device (D2D)-assisted caching networks. Different from the existing caching schemes, the author allows caching fractions of an individual file and considers the self cache hit event, which the user can find the desired content in its device. The author analyzes the tradeoff between the quality of cached contents and the cache hit rate, and proposes the device caching method maximizing the expected quality that the user can enjoy. Depending on the parameter of the relationship between the quality and the file size, the optimal caching method can be obtained by solving the convex optimization problem and the DC programming problem. If the file size increases faster than the quality, the cached fractions of the contents continuously increase as the popularity grows. Meanwhile, if the file size increases slower than the quality, some of the high-popularity files are entirely cached but others are not cached at all.

▸ **Key words**: device caching, wireless caching, content delivery network, convex optimization, DC programming

[요 약]

본 논문에서는 디바이스 캐싱 네트워크에서 다양한 퀄리티의 콘텐츠를 캐싱하는 기술을 제안한다. 하나의 파일을 온전히 캐싱하는 기존 기술들과 다르게, 저자는 콘텐츠의 일부 조각을 캐싱하는 것을 허용하였고, 사용자가 스스로 캐시 히트를 달성할 수 있는 경우를 고려하였다. 캐싱하는 콘텐츠의 퀄리티와 캐시 히트율 간의 트레이드오프를 분석하고, 사용자가 소비하는 콘텐츠의 기대 퀄리티를 최대화하는 디바이스 캐싱 기법을 제안한다. 퀄리티와 파일 크기의 관계 파라미터에 따라 볼록 최적화 문제와 DC programming 문제 두 가지 방식으로 나누어서 캐싱 문제를 풀어냈다. 퀄리티 증가 폭에 비해 파일 크기가 더 빠르게 증가하면, 인기도에 따라 캐싱할 콘텐츠의 부분 조각이 점차 증가하는 반면, 파일 크기가 더 느리게 증가하면, 일부 인기도가 높은 콘텐츠는 전체를 캐싱하고 그렇지 않은 콘텐츠는 아예 캐싱하지 않는 결과를 낸다.

▸ **주제어**: 디바이스 캐싱, 무선 캐싱, 콘텐츠 전송 네트워크, 볼록 최적화 문제, DC programming

• First Author: Minseok Choi, Corresponding Author: Minseok Choi
*Minseok Choi (ejaqmf@jejunu.ac.kr), Dept. of Telecommunication Engineering, Jeju National University
• Received: 2020. 12. 14, Revised: 2021. 01. 06, Accepted: 2021. 01. 06.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

시스코의 보고서에 따르면, 전 세계의 무선 데이터 트래픽의 많은 비율이 온디맨드 비디오 스트리밍과 같은 멀티미디어 서비스가 높은 비율을 차지한다 [1]. 멀티미디어 서비스 특성상 수많은 콘텐츠 중 극히 소수가 매우 높은 인기도를 지녀 다수의 사용자에게 반복적으로 요청된다. [2] 이 과정에서 사용자로부터 콘텐츠의 요청을 받은 기지국은 서버 혹은 클라우드에게 콘텐츠를 백홀 연결을 통해 전달받아 사용자에게 제공한다. 이때, 인기도가 높은 콘텐츠를 반복적으로 요청할 시 백홀 사용량이 비효율적으로 증가하는데, 이에 대응할 기술로 무선 캐싱이 등장하였다 [3-4].

무선 캐싱 네트워크에서는 독립적인 저장장치를 가진 캐싱 노드에 수요가 많은 콘텐츠 위주로 미리 캐싱해 놓고, 주변의 사용자가 해당 콘텐츠를 요청할 때 서버와의 통신 없이 바로 제공해줄 수 있도록 한 것이다. 캐싱 노드는 스몰 셀 기지국이 될 수도 있고 [5-6], 사용자 디바이스가 될 수도 있다 [7-8]. 무선 캐싱을 이용하면 백홀 사용량도 낮추고 콘텐츠 전송 지연 시간도 줄일 수 있어서, 무선 캐싱 네트워크는 온라인 영상 서비스를 지원하기에 매우 좋은 시스템으로 각광받고 있다 [9].

무선 캐싱 네트워크에서의 주요 문제는 콘텐츠를 캐싱하는 방법 (캐싱 문제)과 캐싱된 콘텐츠를 전송하는 방법 (전송 문제)을 정하는 것이다. 캐싱 문제는 캐싱 노드의 분포를 이미 정확히 알고 있을 때, 어느 캐싱 노드의 어떤 콘텐츠를 저장할지 결정하는 문제와, 캐싱 노드의 확률 분포만 알고 있는 경우 확률적인 캐싱 기법을 제시하는 두 가지 방향이 있다. 전송 문제는 콘텐츠를 요청한 사용자가 본인의 주변에서 원하는 콘텐츠를 캐싱한 노드가 있는 경우, 즉 캐시 히트(cache hit)를 이루도록 하거나, 실제 물리 채널 환경에서 정확히 콘텐츠가 전송될 확률을 높이거나, 전송 지연 시간을 낮추어주는 등의 성능 지표를 활용하여 풀어낸다.

다양한 확률적 캐싱 기술들이 스토캐스틱 무선 네트워크에서 제안되었다. 대표적으로 캐시 히트율 [10] 또는 성공적인 전송률을 최대화하거나 [11], 네트워크 코스트 함수를 최소화하거나 [12], 사용자의 선호도에 따라 연속적인 사용자의 요구를 충족하는 방향으로 제안된 캐싱 기법들이 있다 [13].

대다수의 무선 캐싱에 대한 논문들은 모든 콘텐츠의 크기가 같다고 가정하였지만, 최근에 동일 콘텐츠가 여러 개의 퀄리티 레벨을 지니고 각 퀄리티 레벨 별로 파일의 크기가 다른 경우의 캐싱 문제를 본 연구들이 있다 [14-17]. 특히 영상 파일의 경우 다수의 버전으로 인코딩이 되어 최대 신호 대 잡음 비 또는 해상도 등을 다양하게 제공할 수 있으므로, 캐싱 노드는 어떤 버전의 콘텐츠를 캐싱하고 사용자에게 전송할지에 대한 결정도 내려야 한다. 콘텐츠 전송 관점에서, 채널 환경에 따라 전송 큐의 안정성을 추구하면서 제공하는 콘텐츠의 퀄리티를 최대화하는 전송 기법과 [14], 시변하는 영상의 화질에 대한 네트워크 유틸리티 함수를 최대화하는 기술이 제안된 바 있다 [15]. 시간에 따라 변하는 채널 상태에 따라서 전송하고자 하는 콘텐츠의 퀄리티를 비롯한 캐싱 노드 결정, 전송할 영상 조각의 양까지 한꺼번에 결정하는 기술은 [16]에서 제안되었다. [17]에서는 디바이스 캐싱이 가능할 때, 퀄리티가 다른 콘텐츠 전송을 고려한 디바이스 간 통신 스케쥴링 기법을 제안하였다.

콘텐츠 캐싱 관점에서도 동일 콘텐츠의 서로 다른 퀄리티 버전은 캐시 히트율과 사용자에게 제공할 수 있는 퀄리티 간의 트레이드오프를 발생시킨다. 캐싱 노드가 사용자에게 높은 화질의 영상을 제공하고자 캐싱한다면, 제한된 저장용량 때문에 다양한 콘텐츠를 캐싱할 수 없어 캐시 히트율이 낮아진다. 반대로, 캐시 히트율을 올리려고 다양한 콘텐츠를 저장하려면 필연적으로 콘텐츠의 화질을 낮추어야 한다. 이러한 트레이드오프를 분석하여 무선 캐싱 네트워크에서의 퀄리티별 확률적 캐싱 및 전송 기법을 [18]에서 처음 제안하였다. 영상 파일이 scalable video coding (SVC)로 인코딩되었을 때, 전송 지연 시간을 최소화하는 인코딩된 영상 파일의 레이어 개수를 결정하는 기술도 제안된 바 있다 [19, 20].

다양한 퀄리티 레벨로 인코딩될 수 있는 콘텐츠의 캐싱을 연구한 [18-20]의 논문에서는 콘텐츠를 요청하는 사용자와 캐싱 노드가 분리된 경우를 고려하였다. 본 논문에서는 [17]과 같이 사용자 디바이스가 본인의 저장공간을 활용하여 직접 캐싱 노드가 될 수 있는 디바이스 간 통신 기반의 무선 캐싱 네트워크를 고려한다. [17]에서는 콘텐츠 전송 관점에서의 콘텐츠의 퀄리티 레벨의 영향을 보았지만, 본 논문에서는 콘텐츠 캐싱 관점에서의 퀄리티 레벨의 영향을 살펴본다. 디바이스 캐싱이 가능한 환경의 특징은 디바이스가 직접 캐싱한 콘텐츠를 스스로 요청할 경우 다른 디바이스의 도움 없이 즉시 콘텐츠의 소비가 가능하다. 따라서 콘텐츠 요청 사용자와 콘텐츠를 제공하는 캐싱 노드가 분리된 네트워크와 다르게 최적화 문제가 정립되며, 확률적 캐싱 기법도 다르게 최적화될 것이다. 본 논문에서는 디바이스 캐싱이 가능한 환경에서 여러 퀄리티 레벨로 인코딩될 수 있는 영상 파일을 사용자들의 기대 퀄리티를 최대화하는 방향으로 캐싱하는 기술을 제안한다.

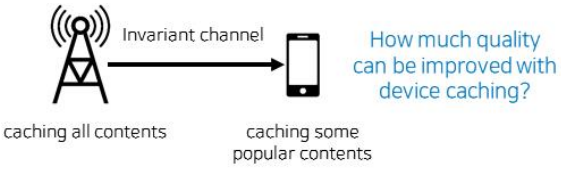

Fig. 1. Device caching-enabled network

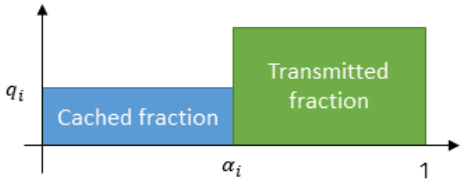

Fig. 2. Scenario of caching fractions of contents and receiving remaining via content delivery links

II. System Model

1. Contents caching model

본 논문에서는 사용자 디바이스가 직접 콘텐츠를 일부 캐싱할 수 있는 환경을 고려한다. 디바이스의 캐싱 용량은 M이라 하고, 총 F개의 콘텐츠가 $[q_{\min}, q_{\max}]$ 범위 내의 퀄리티 q를 가질 수 있다고 하자. 디바이스는 콘텐츠 i의 일부 조각 α_i만큼을 퀄리티 q_i로 캐싱한다. 즉, 캐싱 정책은 $\{\boldsymbol{\alpha}, \boldsymbol{q}\}$로 정의되고, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_F]$과 $\boldsymbol{q} = [q_1, q_2, \cdots, q_F]$ 이며, $0 \le \alpha_i \le 1$ $q_{\min} \le q_i \le q_{\max}$를 모든 $i \in \{1, 2, \cdots FRIGHT$ 에 대해 만족한다. 모든 디바이스가 동일 캐싱 정책을 이용한다고 하자. 일반적으로 퀄리티 q가 높아질수록 콘텐츠의 파일 크기 S는 증가하는데, 그 관계를 아래와 같다고 가정한다. 이때, A와 β는 시스템 파라미터 값이다.

$$q = A \cdot S^{\beta}$$

또한, $q_{\min}$과 $q_{\max}$ 퀄리티에 해당하는 파일 크기는 각각 $S_{\min}$과 $S_{\max}$라 하자.

2. Contents delivery model

사용자는 랜덤하게 F개의 콘텐츠 중 하나를 아래와 같은 M-Zipf 분포에 따라 요청하며, ν와 γ는 각각 인기도 분포 skewness와 plateau factor이다.

$$f_i = \frac{(i+\nu)^{-\gamma}}{\sum_{j=1}^{F}(j+\nu)^{-\gamma}}$$

사용자가 요청한 콘텐츠가 본인 디바이스 내 캐싱되어 있다면, 즉시 캐시 히트를 달성하고 콘텐츠 소비를 할 수 있다. 반면, 캐싱하지 않은 콘텐츠를 요청할 시에는 Fig. 1에 묘사된 것처럼 기지국 혹은 가까운 캐싱 헬퍼의 도움을 받아야 한다. 본 논문에서는 모든 디바이스가 같은 $\{\boldsymbol{\alpha}, \boldsymbol{q}\}$를 캐싱한다고 가정하여, 본인에게 해당 콘텐츠가 없다면 다른 디바이스에게도 도움을 받을 수가 없어서 기지국이나 캐싱 헬퍼에게서 콘텐츠를 받아야 한다.

사용자가 디바이스에 캐싱된 파일 i를 요청하였더라도, 파일 i의 일부 α_i만을 캐싱하였기에 온전한 콘텐츠를 소비하기 위해서는 캐싱하지 않은 $1-\alpha_i$ 부분의 조각을 캐싱 헬퍼에게 전달받아야 한다. 디바이스가 캐싱한 조각 α_i의 퀄리티 q_i와 캐싱 헬퍼가 전달해주는 $1-\alpha_i$ 조각의 퀄리티 q_h는 달라도 되며, 그 평균값인 $\alpha_i q_i + (1-\alpha_i) q_h$가 사용자가 소비하는 전체 콘텐츠의 평균 퀄리티 측정값이 된다. 이 관계는 Fig. 2에 나타나 있다.

캐싱 헬퍼와 사용자 사이의 채널은 고정적이라고 가정하고, 이때 채널 용량을 C라 한다. 이산 시간 간격 T마다 콘텐츠 전송이 이루어질 때, 총 전달할 수 있는 데이터의 양은 CT로 정의할 수 있다. 즉, 캐싱 헬퍼가 채널 용량이 지원하는 한 최대의 퀄리티의 파일을 전송해준다고 가정하면, 전송해주는 퀄리티 q_h는 아래와 같다.

$$q_h = \min\left\{A\left(\frac{CT}{1-\alpha_i}\right)^{\beta}, q_{\max}\right\}$$

위 식에서 최소 함수 내 왼쪽 값은 채널 용량이 지원하는 한 제공 가능한 최대 퀄리티 값인데, 이 값은 $q_{\max}$를 넘지 못하므로 위와 같이 정의한다.

III. Proposed Caching Policy

1. Problem formulation

사용자가 소비하는 콘텐츠의 기대 퀄리티를 최적화 문제의 목적함수 $g(\boldsymbol{\alpha}, \boldsymbol{q})$로 정의하면 이는 다음과 같다.

$$g(\boldsymbol{\alpha}, \boldsymbol{q}) = \sum_{i=1}^{F} f_i \cdot \left[\alpha_i q_i + (1-\alpha_i) q_h\right]$$

이를 바탕으로 최적 캐싱 기법을 위한 문제를 정립하면 아래와 같다.

$$\{\boldsymbol{\alpha}^*,\boldsymbol{q}^*\}=\arg\max_{\boldsymbol{\alpha},\boldsymbol{q}} g(\boldsymbol{\alpha},\boldsymbol{q})$$
$$s.t.\ A^{-1/\beta}\sum_{i=1}^{F}\alpha_i q_i^{1/\beta} \le M$$
$$0 \le \alpha_i \le 1, \quad \forall\ i$$
$$q_{\min} \le q_i \le q_{\max}, \quad \forall\ i$$

위의 최적화 문제는 아래의 Lemmas 1, 2, Theorem 1, Proposition 1을 만족한다. 페이지가 부족하여 모든 증명은 생략한다.

Lemma 1. 최적의 캐싱 정책 $\{\boldsymbol{\alpha}^*\boldsymbol{q}^*\}$은 아래 등식을 만족한다.

$$\sum_{i=1}^{F}\alpha_i^*\left(\frac{q_i^*}{Aq_{\min}}\right)^{1/\beta}\cdot\ S_{\min}=M.$$

Lemma 2. 최적의 캐싱 정책 $\{\boldsymbol{\alpha}^*\boldsymbol{q}^*\}$은 아래 부등식을 만족한다.

$$A\left(\frac{CT}{1-\alpha_i^*}\right)^{\beta} \le q_{\max}$$

Proposition 1. 아래 부등식을 만족하면, 디바이스 캐싱은 필요하지 않다.

$$CT > \left(\frac{q_{\max}}{A}\right)^{1/\beta}$$

Theorem 1. 위의 캐싱 정책 최적화 문제는 $\beta \le 1$일 때 볼록 (convex) 최적화 문제이며, $\beta > 1$일 때는 볼록하지 않은 (nonconvex) 최적화 문제이다.

Lemmas 1과 2에 따르면, 최적화 문제는 다음과 같이 변환되며, 변환된 문제를 바탕으로 최적 캐싱 정책을 구한다.

$$\{\boldsymbol{\alpha}^*,\boldsymbol{q}^*\}=$$
$$\arg\max_{\boldsymbol{\alpha},\boldsymbol{q}}\sum_{i=1}^{F}f_i\cdot\left[\alpha_i q_i+(1-\alpha_i)\cdot A\left(\frac{CT}{1-\alpha_i}\right)^{\beta}\right]$$
$$s.t.\ A^{-1/\beta}\sum_{i=1}^{F}\alpha_i q_i^{1/\beta} \le M$$
$$0 \le \alpha_i \le 1-CT\left(\frac{A}{q_{\max}}\right)^{1/\beta}, \quad \forall\ i$$
$$q_{\min} \le q_i \le q_{\max}, \quad \forall\ i$$

2. Proposed caching when $\beta \le 1$

$\beta \le 1$일 때, 위 최적화 문제는 볼록 최적화 문제이므로, KKT 조건으로 문제의 해답을 얻을 수 있다. 위 문제의 KKT 조건들을 정리하면 아래와 같으며, 해당 KKT 조건들을 모두 만족하는 캐싱 정책 $\{\boldsymbol{\alpha}^*\boldsymbol{q}^*\}$ 값을 파라미터들의 범위에 따라 Theorem 2에 정리하였다.

$$f_i A(1-\beta)\left(\frac{CT}{1-\alpha_i}\right)^{\beta}-f_i q_i-\lambda\left(\frac{q_i}{A}\right)^{\frac{1}{\beta}}+\mu_i-\eta_i=0$$
$$-f_i\alpha_i-\lambda\frac{\alpha_i}{\beta}A^{-1/\beta}q_i^{1/\beta-1}-\rho_i+\delta_i=0$$
$$A^{-1/\beta}\sum_{i=1}^{F}\alpha_i q_i^{1/\beta}=M$$
$$0 \le \alpha_i \le 1-CT\left(\frac{A}{a_{\max}}\right)^{1/\beta}, \quad \forall\ i$$
$$q_{\min} \le q_i \le q_{\max}, \quad \forall\ i$$
$$\mu_i\left(\alpha_i-1+A\frac{CT}{q_{\max}}\right)=0, \quad \forall\ i$$
$$-\eta_i\alpha_i=0$$
$$\rho_i(q_{\min}-q_i)=0$$
$$\delta_i(q_i-q_{\max})=0$$
$$0 \le \mu_i,\eta_i,\rho_i,\delta_i, \quad \forall\ i$$

Theorem 2. 채널 용량이 C, 콘텐츠를 받는 이산 시간 T가 주어졌을 때, $\beta \le 1$인 경우 위 KKT 조건들을 모두 만족하는 최적의 $\{\boldsymbol{\alpha}^*,\boldsymbol{q}^*\}$는 아래의 조건들을 만족한다.

조건 1. $A^{-1/\beta}\sum_{i=1}^{F}\alpha^{*}{}_i\left(q^{*}{}_i\right)^{1/\beta}=M$

조건 2.

1) $-f_i q_{\min}\left(\frac{A}{q_{\min}}\right)^{1/\beta} \le \lambda \le -f_i q_{\min}\beta\left(\frac{A}{q_{\min}}\right)^{1/\beta}$일 때, $q_i^*=q_{\min}$, $\alpha_i^*=\max\{0,\min\{\xi_i,1\}\}$이다.

2) $-f_i q_{\min}\beta\left(\frac{A}{q_{\min}}\right)^{1/\beta} \le \lambda \le -f_i q_{\max}\beta\left(\frac{A}{q_{\max}}\right)^{1/\beta}$일 때, $q_i^*=\left[\frac{-\lambda}{f_i\beta A^{1/\beta}}\right]^{\frac{\beta}{\beta-1}}$, $\alpha_i^*=\max\{0,\min\{\chi_i,1\}\}$

3) $-f_i q_{\max}\beta\left(\frac{A}{q_{\max}}\right)^{1/\beta} \le \lambda$일 때,

$q_i^*=q_{\max}$, $\alpha_i^*=\max\{0,\min\{\varphi_i,1\}\}$이다.

이때, ξ_i, χ_i, φ_i는 다음과 같다.

$$\xi_i=1-CT\left[\frac{1}{f_i(1-\beta)A}\cdot\left\{f_i q_{\min}+\lambda\left(\frac{q_{\min}}{A}\right)^{1/\beta}\right\}\right]^{-1/\beta}$$
$$\chi_i=1-CT\left(\frac{-\lambda}{f_i\beta A}\right)^{1/(1-\beta)}$$
$$\varphi_i=1-\left[\frac{1}{f_i(1-\beta)A(CT)^{\beta}}\cdot\left\{f_i q_{\max}+\lambda\left(\frac{q_{\max}}{A}\right)^{1/\beta}\right\}\right]^{-1/\beta}$$

3. Proposed caching when $\beta > 1$

$\beta > 1$일 때는, 최적화 문제의 목적함수를 아래와 같이 두 개의 볼록 함수의 차로 표현할 수 있다.

$$g(\boldsymbol{\alpha},\boldsymbol{q}) = u(\boldsymbol{\alpha},\boldsymbol{q}) - h(\boldsymbol{\alpha})$$
$$= \left(-\sum_{i=1}^{F} f_i \alpha_i q_i\right) - \left(\sum_{i=1}^{F} f_i \;\; A(CT)^{\beta}(1-\alpha_i)^{1-\beta}\right)$$

위에서 $u(\boldsymbol{\alpha},\boldsymbol{q})$와 $h(\boldsymbol{\alpha})$는 볼록 함수이며, 목적함수를 볼록 함수 두 개의 차로 바꾸었을 때, 위 최적화 문제는 DC programming 문제로 바뀐다. DC programming 문제는 전통적으로 아래와 같은 볼록 최적화 문제를 반복적으로 풀어 본래의 최적화 문제의 답을 얻을 수 있다.

$$\min_{\boldsymbol{\alpha},\boldsymbol{q}} u(\boldsymbol{\alpha},\boldsymbol{q}) - h(\boldsymbol{\alpha}^{(k)}) + \;\; (\boldsymbol{\alpha} - \boldsymbol{\alpha}^{(k)}) \cdot \;\; \nabla \; h(\boldsymbol{\alpha}^{(k)})$$

위 DC programming 문제의 목적함수를 다시 작성하면 아래와 같다.

$$\sum_{i=1}^{F} \alpha_i f_i \cdot \;\; \left[-q_i + (1-\beta)A(CT)^{\beta}(1-\alpha_i^{(k)})\right] -$$
$$\sum_{i=1}^{F} f_i \cdot \;\; A(CT)^{\beta}\left[(1-\alpha_i^{(k)})^{1-\beta} + \alpha_i^{(k)}(1-\beta)(1-\alpha_i^{(k)})\right]$$

또한, 임의의 작은 수 $0 < \epsilon < 1$에 대해서 아래의 부등식을 만족할 때, $\boldsymbol{\alpha}^{(k+1)}$과 $\boldsymbol{q}^{(k+1)}$은 위 DC programming 문제의 최적의 답이 된다.

$$\left| g(\boldsymbol{\alpha}^{(k+1)},\boldsymbol{q}^{(k+1)}) - g(\boldsymbol{\alpha}^{(k)},\boldsymbol{q}^{(k)}) \right| \le \;\; \epsilon$$

위 DC programming 문제의 KKT 조건들을 정리하면 아래와 같으며, 해당 KKT 조건들을 모두 만족하는 캐싱 정책 $\{\boldsymbol{\alpha}^{*}\boldsymbol{q}^{*}\}$ 값을 파라미터들의 범위에 따라 Theorem 2에 정리하였다.

$$f_i\left[-q_i + (1-\beta)A(CT)^{\beta}(1-\alpha_i^{(k)})\right]$$
$$-\lambda\left(\frac{q_i}{A}\right)^{1/\beta} + \mu_i - \eta_i = 0$$
$$-\alpha_i f_i - \lambda \frac{q_i^{1/\beta - 1}}{\beta A^{1/\beta}}\alpha_i - \rho_i + \delta_i = 0$$
$$A^{-1/\beta}\sum_{i=1}^{F} \alpha_i q_i^{1/\beta} = M$$
$$0 \le \;\; \alpha_i \le \;\; 1 - CT\left(\frac{A}{q_{\max}}\right)^{1/\beta}, \;\; \forall \;\; i$$
$$q_{\min} \le \;\; q_i \le \;\; q_{\max}, \;\; \forall \;\; i$$
$$\mu_i\left(\alpha_i - 1 + A\frac{CT}{q_{\max}}\right) = 0, \;\; \forall \;\; i$$
$$-\eta_i \alpha_i = 0, \;\; \forall \;\; i$$
$$\rho_i(q_{\min} - q_i) = 0, \;\; \forall \;\; i$$
$$\delta_i(q_i - q_{\max}) = 0, \;\; \forall \;\; i$$
$$0 \le \;\; \mu_i, \eta_i, \rho_i, \delta_i, \;\; \forall \;\; i$$

Theorem 3. 채널 용량이 C, 콘텐츠를 받는 이산 시간 T가 주어졌을 때, $\beta > 1$인 경우 위 KKT 조건들을 모두 만족하는 최적의 $\{\boldsymbol{\alpha}^{*},\boldsymbol{q}^{*}\}$는 아래의 조건들을 만족한다.

조건 1. $A^{-1/\beta}\sum_{i=1}^{F} \alpha^{*i}(q^{*i})^{1/\beta} = M$

조건 2. $-f_i q_i \beta\left(\frac{A}{q_i}\right)^{1/\beta} \le \;\; \lambda$ 일 때,

$$\alpha_i = 1 - CT\left(\frac{A}{q_{\max}}\right)^{1/\beta}$$

조건 3.

1) $-f_i q_i \beta\left(\frac{A}{q_i}\right)^{1/\beta} \le \;\; \lambda \le \;\; -f_i q_{\max}\beta\left(\frac{A}{q_{\max}}\right)^{1/\beta}$일 때,

$$q_i = q_{\min}$$

2) $-f_i q_{\max}\beta\left(\frac{A}{q_{\max}}\right)^{1/\beta} \le \;\; \lambda \le \;\; -f_i q_{\min}\beta\left(\frac{A}{q_{\max}}\right)^{1/\beta}$일 때, $q_i = \left[\frac{-\lambda}{f_i \beta A^{1/\beta}}\right]^{\frac{\beta}{\beta-1}}$

3) $-f_i q_{\min}\beta\left(\frac{A}{q_{\max}}\right)^{1/\beta} \le \;\; \lambda$일 때, $q_i = q_{\max}$

Table 1. System Parameters

Item	Value
Number of contents (F)	20
Cache Size (M)	65 kB
Number of quaity levels (L)	4
minimum quality measure ($q_{\min}$)	34 dB
maximum quality measure ($q_{\max}$)	41.64 dB
minimum file size ($S_{\min}$)	2.6 kB
scaling parameter (A)	1

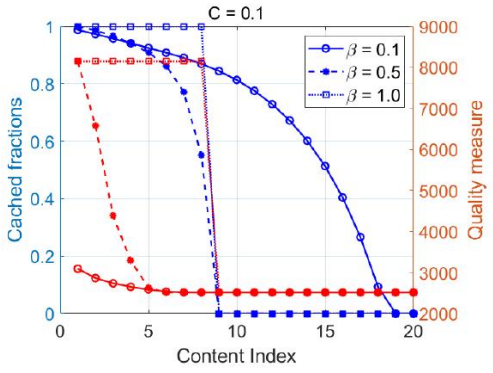

Fig. 3. caching policy when $\beta \le \;\; 1$ and $C = 0.1$

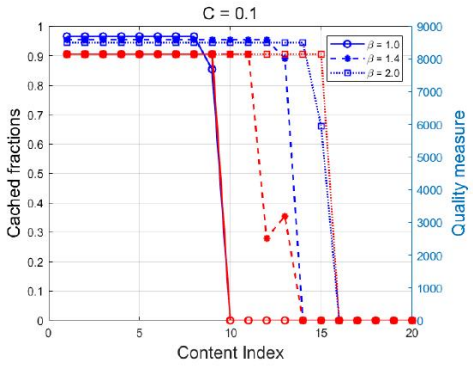

Fig. 4. caching policy when $\beta \geq 1$ and $C = 0.1$

IV. Simulation Results

1. Simulation Environments

본 시뮬레이션에서 사용한 파라미터 값은 Table 1에 정리되어 있다. 저자는 콘텐츠 퀄리티와 파일 크기의 관계 파라미터 β와 채널 용량 C를 다르게 하며 시뮬레이션을 진행하였다. 또한, 콘텐츠의 퀄리티 측정을 위해서는 최대 신호 잡음 비 값을 이용하였다.

2. Simulation Results

Figs. 3-6의 가로축은 모두 콘텐츠의 순번을 인기도 순서로 나열한 것이고, 왼쪽의 세로축은 디바이스에 캐싱된 부분 조각 α_i, 오른쪽의 세로축은 캐싱된 콘텐츠 조각의 퀄리티 측정 값, 즉 최대 신호 잡음 비 값이 나타나 있다. Fig. 3에서는 $\beta \leq 1$ 일 때의 최적의 캐싱 기법 $\{\boldsymbol{\alpha}^*, \boldsymbol{q}^*\}$을 보여준다. β가 작다는 것은 퀄리티가 일정 수준 증가한다면 파일 크기는 그보다 몇 배 이상으로 커진다는 의미이다. 따라서, β가 작을수록 높은 퀄리티의 파일을 캐싱할수록 저장공간을 많이 차지하여 굳이 높은 퀄리티의 파일을 캐싱하려 하지 않는다. 따라서 $\beta = 0.1$인 경우에는 모든 캐싱한 파일은 낮은 퀄리티이며, 대신 남는 저장공간에 비교적 낮은 인기도의 콘텐츠도 골고루 저장하는 것을 볼 수 있다. $\beta = 0.5$일 때는 인기도가 높은 일부 콘텐츠만이 높은 퀄리티로 저장되며, 대신 비교적 인기도가 높은 콘텐츠 위주로 캐싱한다. 반면, 퀄리티와 파일 크기가 비례할 때 ($\beta = 1$), 각 콘텐츠의 일부 조각을 캐싱하기보다는 파일 전체를 캐싱하거나, 아니면 아예 캐싱하지 않는 정책을 보인다.

Fig. 4는 $\beta > 1$인 경우의 캐싱 정책을 보여주는데, 이 경우는 퀄리티가 증가하는 폭 대비 파일 크기는 더 적게 증가한다. 따라서, 어떤 콘텐츠를 캐싱하기로 했으면, 될 수 있으면 높은 퀄리티를 캐싱하는 것이 더 좋으며, Fig. 4의 그래프들에서도 그 결과를 찾아볼 수 있다. 아무래도 $\beta \leq 1$일 때와 비교해서 높은 퀄리티의 파일 크기가 더 작으므로 비교적 더 많은 콘텐츠를 캐싱하는 것을 볼 수 있다. 또한, 대부분 캐싱하기로 한 콘텐츠는 거의 파일의 대부분을 디바이스에 저장하고, 극히 소량만을 기지국으로부터 전송받는 것을 알 수 있다.

Figs. 5-6에서는 채널 용량 C를 0.5까지 증가시켰을 때의 효과를 살펴보았다. C가 커지면 기지국이 전달해줄 수 있는 데이터 전송률이 증가하므로, 캐싱 공간에 여유분을 더 가져올 수 있다. 흥미롭게도, $\beta = 0.5$일 때는 C가 증가하면 캐싱 공간의 여유분을 더 높은 퀄리티의 콘텐츠를 저장하는 데 사용하고, 오히려 $C = 0.1$이ㄹ때는 캐싱했던 8번 콘텐츠의 캐싱을 포기해 버린다. 반면, Fig. 6에서 $\beta = 1.5$인 경우에는 캐싱 공간의 여유분을 인기도가 낮은 콘텐츠를 더 많이 저장하는 데 사용한다. 이처럼 퀄리티와 파일 크기의 관계 파라미터에 따라 캐싱하는 방법이 크게 영향받는 것을 관찰할 수 있다.

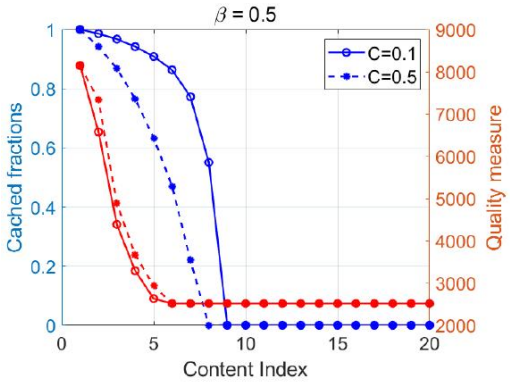

Fig. 5. caching policy when $\beta = 0.5$

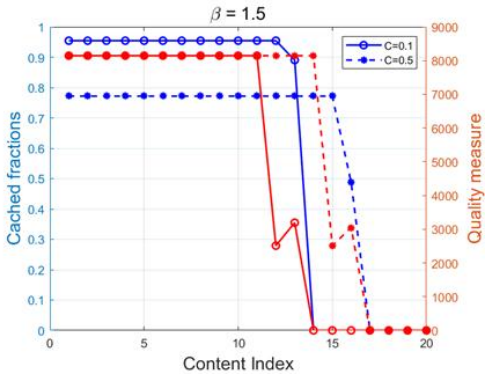

Fig. 6. caching policy when $\beta = 1.5$

V. Conclusions

본 논문에서는 디바이스 캐싱이 가능한 네트워크에서, 캐싱한 콘텐츠의 퀄리티와 캐시 히트율 간의 트레이드오프를 분석하고, 사용자의 기대 퀄리티 값을 최대화하는 캐싱 기법을 제안하였다. 또한, 콘텐츠의 퀄리티 대비 파일 크기가 증가하는 비율에 따라서 캐싱하는 기법이 달라짐을 시뮬레이션으로 보였다. 이는 어플리케이션마다 캐싱하는 기법이 달라질 수 있음을 의미한다.

ACKNOWLEDGEMENT

This work was supported by the 2020 education, research and student guidance grant funded by Jeju National University.

REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2018--2023 White Paper. Accessed: Mar. 3, 2020, Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[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, August 2013.

[3] E. Bastug, M. Bennis, and M. Debbah, ``Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82−89, August 2014.

[4] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley and G. Caire, "The Role of Caching in Future Communication Systems and Networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1111-1125, June 2018.

[5] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, ``FemtoCaching: Wireless Video Content Delivery through Distributed Caching Helpers," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, 2012.

[6] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, ``Cache in the Air: Exploiting Content Caching and Delivery Techniques for 5G Systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 131--139, February 2014.

[7] M. Ji, G. Caire, and A. F. Molisch, ``Fundamental Limits of Caching in Wireless D2D Networks," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 849--869, February 2016.

[8] M. Ji, G. Caire, and A. F. Molisch, "Wireless Device-to-Device Caching Networks: Basic Principles and System Performance", *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 176--189, Jan. 2016.

[9] Y. He, I. Lee, and L. Guan, "Distributed Throughput Maximization in P2P VoD Applications," *IEEE Transactions on Multimedia*, vol. 11, no. 3, pp. 509--522, April 2009.

[10] B. Blaszczyszyn and A. Giovanidis, "Optimal Geographic Caching in Cellular Networks," *Proc. IEEE Int'l Conf. on Communications (ICC)*, London, UK, 2015.

[11] 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, October 2016.

[12] M. Gregori, J. G´omez-Vilardeb´o, J. Matamoros, and D. G¨und¨uz, "Wireless Content Caching for Small Cell and D2D Networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1222−1234, May 2016.

[13] M. Choi, D. Kim, D. Han, J. Kim and J. Moon, "Probabilistic Caching Policy for Categorized Contents and Consecutive User Demands," *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, 2019, pp. 1-6.

[14] J. Kim, G. Caire, and A. F. Molisch, "Quality-Aware Streaming and Scheduling for Device-to-Device Video Delivery," *IEEE/ACM Transactions on Networking*, 24(4):2319−2331, August 2016.

[15] D. Bethanabhotla, G. Caire, and M. J. Neely, "Adaptive Video Streaming for Wireless Networks With Multiple Users and Helpers," *IEEE Transactions on Communications*, 63(1):268285, Jan. 2015.

[16] M. Choi, A. No, M. Ji and J. Kim, "Markov Decision Policies for Dynamic Video Delivery in Wireless Caching Networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5705-5718, Dec. 2019.

[17] M. Choi, A. F. Molisch and J. Kim, "Joint Distributed Link Scheduling and Power Allocation for Content Delivery in Wireless Caching Networks," *IEEE Transactions on Wireless Communications*, Early Access, Aug. 2020.

[18] M. Choi, J. Kim and J. Moon, "Wireless Video Caching and Dynamic Streaming Under Differentiated Quality Requirements," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1245-1257, June 2018.

[19] K. Poularakis, G. Iosifidis, A. Argyriou, I. Koutsopoulos and L. Tassiulas, ``Caching and operator cooperation policies for layered video content delivery," *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, San Francisco, CA, 2016, pp. 1--9.

[20] J. Meng, H. Lu and J. Liu, ``Joint Quality Selection and Caching for SVC Video Services in Heterogeneous Networks," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), May 2020.

Authors

Minseok Choi received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011, 2013, and 2018, respectively. He was a Visiting Postdoctoral Researcher in electrical and computer engineering with the University of Southern California (USC), Los Angeles, CA, USA, and a Research Professor in electrical engineering with Korea University, Seoul, South Korea. He has been an Assistant Professor with Jeju National University, Jeju, South Korea, since 2020. His research interests include wireless caching networks, stochastic network optimization, non-orthogonal multiple access, and 5G networks.