

# 엣지-포그 캐싱 환경에서 재생 지연 감소를 위한

## Prefix 캐싱 기법

정준호 · 성은산 · 이현섭 · 윤주상\*

동의대학교

### Prefix Caching for Playback Delay Reduction in Edge-Fog Caching Environment

Junho Jeong · Eun San Seong · Hyounsop Lee · Joosang Youn\*

Dong-Eui University

E-mail : jeong@junho.dev / ses0137112@naver.com / lhskmj@deu.ac.kr / jsyoun@deu.ac.kr

#### 요 약

엣지 캐싱은 동영상 스트리밍 서비스에서 backhaul 네트워크의 트래픽을 감소시키고 응답 지연을 줄여 높은 QoE를 제공할 수 있다. 그러나 엣지 캐시의 용량은 제한적이기 때문에 많은 양의 콘텐츠를 캐싱할 수 없다. 본 논문에서는 동영상 콘텐츠의 prefix를 엣지에 캐싱하고 나머지 부분을 포그 캐시에 저장하여 재생 지연을 감소시키는 edge-fog prefix caching 기법을 제안한다.

#### ABSTRACT

Edge caching can provide high QoE by reducing traffic in the backhaul network and reducing latency in video streaming services. But due to the limited capacity of edge cache, large amounts of content cannot be cached. In this paper, we propose an edge-fog prefix caching that reduces playback delay by caching prefixes of video content on edges and storing the rest in fog cache.

#### 키워드

Edge Computing, Fog Computing, Caching, Quality of Experience(QoE)

#### I. 서 론

최근 동영상 스트리밍 서비스의 성장으로 무선 디바이스에서의 대용량 데이터 요청이 증가하고 있다. 동영상 스트리밍 서비스의 경우 인기 있는 일부 콘텐츠에 대한 요청이 반복적으로 이루어진다. 기존 클라우드 방식에서는 이러한 특성을 고려하지 않아 backhaul 네트워크에 많은 트래픽을 유발하고 응답 지연이 높아지는데, 이는 Quality-of-Experience (QoE)에 가장 큰 영향을 주는 요소 중 하나이다. 이를 해결하기 위해 엔드유저와 물리적으로 가까운 곳에 데이터를 저장하는 엣지 캐싱 기법이 제안되었다[1,2]. 또한 엣지 노드의 자원은

한정적이기 때문에 더 많은 데이터를 캐싱하기 위한 기법으로 포그 캐싱 또한 제안되었다[1]. 이를 통해 더 많은 콘텐츠를 낮은 응답 지연으로 사용자에게 제공할 수 있게 된다. 하지만 포그 노드는 엣지 노드보다 물리적으로 먼 곳에 위치하기 때문에 엣지에 캐시된 콘텐츠보다 응답 지연이 크다. 이를 해결하기 위해 본 논문에서는 엣지-포그 환경에서 콘텐츠의 캐싱을 보다 효율적으로 하기 위한 기법을 제안한다.

#### II. 관련연구

엣지 캐싱은 분산 컴퓨팅 중 하나로 클라우드와 엔드유저 사이에 위치한 엣지 노드에 서비스할 데

\* corresponding author

이터를 저장하는 기법이다[1,3]. 엣지 노드는 클라우드보다 물리적으로 가까운 곳에 위치하기 때문에 클라우드에서 콘텐츠를 제공하는 것보다 상대적으로 낮은 backhaul 비용과 적은 응답 지연으로 콘텐츠를 제공할 수 있다. 그러나 엣지 노드의 캐시의 용량은 한정적이기 때문에 모든 콘텐츠를 캐싱할 수는 없다. 이를 해결하기 위해 사용확률이 높은 콘텐츠를 캐싱하는 기법[4]과 사용자의 관심도에 따라 콘텐츠를 캐싱하는 기법[5], 동영상 콘텐츠의 초반 일부를 미리 캐싱하여 해당 부분이 발생하는 동안 남은 부분을 서버에서 받아와 응답 지연을 감소시키는 prefix caching 기법[6]과 엣지에 캐싱된 콘텐츠를 제거하거나 교체하는 정책들 또한 제안되었다[7,8,9,10].

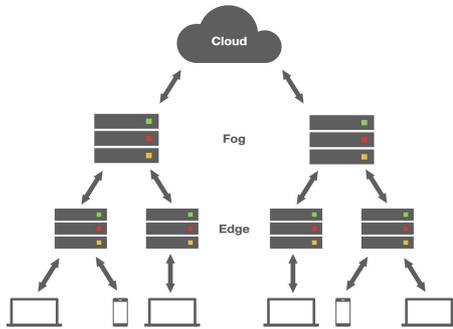


그림 1. 엣지-포그 캐싱 환경

### III. Edge-Fog Prefix Caching

엣지 노드의 캐시는 포그 노드에 비해 용량이 작아 많은 데이터를 캐싱할 수 없다. 이를 위해 여러 엣지 노드에 중복되어 캐싱되는 콘텐츠를 포그 노드에 캐싱하는 경우 다양한 콘텐츠를 캐싱할 수 있지만, 엣지 노드에 캐싱된 콘텐츠에 비해 사용자로부터 상대적으로 멀리 떨어져 있어 응답 지연이 증가한다. 이를 해결하기 위해 요청 확률이 높은 동영상 콘텐츠의 초반 일부만을 엣지 노드에 캐싱하고 나머지는 포그 노드에 캐싱하는 기법을 제안한다.

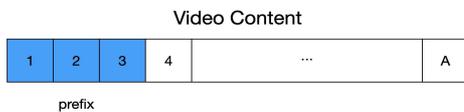


그림 2. 동영상 콘텐츠의 prefix

그림 3은 본 논문에서 제안하는 edge-fog prefix caching 기법을 도시하고 있다. 인기 동영상과 같은 요청횟수가 많은 콘텐츠나 사용자별 추천 동영상 등 요청 확률이 높은 콘텐츠들의 prefix를 엣지 노

드에 캐싱하고 나머지 부분은 포그 노드에 캐싱한다. 서비스의 전반적인 흐름은 그림 3과 같이 진행된다.

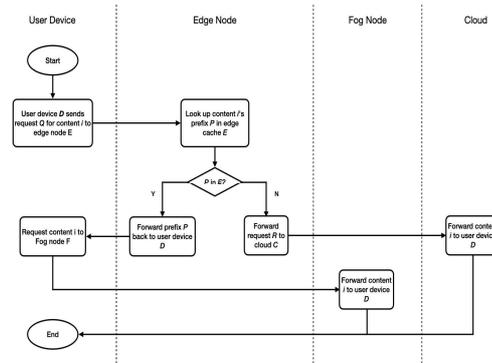


그림 3. Edge-Fog Prefix Caching을 통한 서비스

유저가 요청한 콘텐츠가 엣지 노드에 캐시되어 있다면 해당 콘텐츠의 prefix를 엣지 노드에 요청하여 동영상 스트리밍을 시작한다. 사용자의 디바이스에서 prefix의 로드가 완료되면 콘텐츠의 나머지 부분을 포그 노드에 요청하여 서비스를 진행한다. 엣지 노드에 요청한 콘텐츠가 없는 경우에는 기존 방식과 동일하게 클라우드에서 콘텐츠를 제공한다.

### IV. 결론

본 논문에서는 동영상 스트리밍 서비스의 QoS에 큰 영향을 미치는 재생 지연을 감소시키기 위한 방법으로 요청 확률이 높은 동영상 콘텐츠의 prefix를 엣지 노드에, 나머지 부분을 포그 노드에 캐싱하는 Edge-Fog Prefix Caching 기법을 제안하였다. 그러나 해당 기법은 요청 확률이 적은 콘텐츠를 요청하거나 사용자가 영상의 초반부에서 후반부로 스킵하는 경우에는 응답 지연 감소 효과를 기대하기 힘들다. 추후 테스트베드를 구축하여 해당 기법의 성능 향상률을 측정하기 위한 실험을 실시할 예정이다.

### Acknowledgement

본 논문은 IITP/NIST SW 컴퓨팅산업원천기술개발과제(20200001160012007)의 지원을 받아 수행된 결과임.

### References

[1] Safavat, S., Sapavath, N. and Rawat, D., "Recent advances in mobile edge computing and

- content caching”, *Digital Communications and Networks*, vol. 6, issue 2, pp.189-194, 2020.
- [2] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang and W. Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," in *IEEE Access*, vol. 5, pp. 6757-6779, 2017.
- [3] Y. Mao, C. You, J. Zhang, K. Huang and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322-2358, 2017.
- [4] R. J. Defouw, A. Sutton, R. W. Korngiebel, "Caching method for selecting data blocks for removal from cache based on recall probability and size," *US Patent* 6,742,084, 2004.
- [5] J. Iqbal and P. Giaccone, "Interest-based cooperative caching in multi-hop wireless networks," *2013 IEEE Globecom Workshops*, pp. 617-622, 2013.
- [6] J. Hong and W. Choi, "User Prefix Caching for Average Playback Delay Reduction in Wireless Video Streaming," in *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 377-388, Jan. 2016.
- [7] V. Martina, M. Garetto and E. Leonardi, "A unified approach to the performance analysis of caching systems," *IEEE Conference on Computer Communications*, pp. 2040-2048, 2014.
- [8] J. Liu, G. Wang, T. Huang, J. Chen and Y. Liu, "Modeling the sojourn time of items for in-network cache based on LRU policy," in *China Communications*, vol. 11, no. 10, pp. 88-95, Oct. 2014.
- [9] M. Bilal and S. Kang, "Time Aware Least Recent Used (TLRU) cache management policy in ICN," *16th International Conference on Advanced Communication Technology*, pp. 528-532, 2014.
- [10] H. Gomaa, G. G. Messier, C. Williamson and R. Davies, "Estimating Instantaneous Cache Hit Ratio Using Markov Chain Analysis," in *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1472-1483, Oct. 2013.