

기회적 포그 컴퓨팅 노드를 활용한 CDN 오프로딩

경연웅*

CDN Offloading with Opportunistic Fog Computing Nodes

Yeunwoong Kyung*

요약

낮은 지연 시간 및 효율적인 네트워크 사용을 위해 포그 컴퓨팅 노드(FN; Fog Computing Node)를 활용한 콘텐츠 제공 서비스에 대한 연구가 진행되고 있다. 하지만 다수의 요청이 동일한 영역에서 짧은 시간 내에 발생하게 되면 FN의 과부하 현상이 발생할 수 있다. 본 논문에서는 이 문제를 해결하기 위해 기회적 포그 컴퓨팅 노드(OFN; Opportunistic FN)를 통해 부하를 분산시키고자 한다. 제안된 방법은 FN을 통해 콘텐츠 서비스를 진행하다가 OFN이 가용해지면 서비스 마이그레이션을 수행하여 OFN을 통해 서비스를 제공받도록 함으로써 서비스 연속성을 보장하고 FN의 부하를 분산시킬 수 있는 방법을 제안한다.

Key Words : CDN, offloading, opportunistic fog computing node, load distribution, migration

ABSTRACT

Content delivery service based on the fog computing node (FN) has been researched a lot to provide low service latency and efficient network resource utilization. However, overload on FN can occur when there are lots of service requests within the same area and short period. To solve the problem, this paper utilizes the opportunistic FN (OFN) to distribute the load of FN. The proposed scheme enables the service migration when OFN is available although the service starts with FN, which

can provide the service continuity as well as the load distribution of FN.

I. 서론

최근 네트워크 트래픽의 많은 양을 유튜브와 넷플릭스와 같은 비디오 콘텐츠들이 차지하고 있다. 또한 Internet of Things (IoT) 기기들을 통한 고화질 360도 비디오 및 augmented reality (AR), virtual reality (VR) 어플리케이션 사용이 증가하면서 이러한 현상은 심화되고 있다¹⁾. 이러한 상황 속에서 낮은 지연시간을 제공하고 네트워크 부하를 줄이기 위하여 클라우드 서버에서 제공되던 콘텐츠 서비스를 네트워크 엣지 영역으로 확장하여 포그 컴퓨팅 노드(FN; Fog Computing Node)에서 제공하는 연구들이 활발히 진행되었다²⁾. 하지만 서비스를 요청하는 기기들의 수가 증가하면서 다수의 요청이 동일한 영역에서 짧은 시간 내에 발생하게 되면 포그 컴퓨팅 노드의 부하 또한 증가할 수 있다. 이러한 문제를 해결하고자 스마트폰 및 차량 등과 같이 네트워크 내의 기회적으로 가용한 자원들인 기회적 포그 컴퓨팅 노드(OFN; Opportunistic Fog Computing Node)들을 통해 정적인 포그 컴퓨팅 노드의 부하를 분산시키고자 하는 노력들이 진행되었다³⁾. 하지만 OFN을 통한 서비스 제공은 OFN의 이동성으로 인해 서비스의 연속성을 보장하기 어렵다. 기존 연구들은 사용자의 이동 특성을 고려하여 최적의 콘텐츠 서버로 서비스를 마이그레이션 시키는 방법들을 제공하였지만⁴⁾ 콘텐츠 소스인 OFN의 이동성을 고려하진 못하였다.

그러므로 본 논문에서는 OFN을 활용하여 콘텐츠 서비스를 연속적으로 제공함으로써 정적인 FN의 부하를 줄일 수 있는 방법을 제공하고자 한다.

II. 시스템 모델

그림 1은 본 논문의 시스템 모델을 보여준다. MN(Mobile Node)은 BS(Base Station)에 연결된 FN을 통해 r_F 의 전송 속도로 콘텐츠 서비스를 제공받고 있다. MN은 τ_1 과 τ_2 사이에는 OFN과의 연결이 가능하고 FN에서 제공받던 콘텐츠가 p 의 확률로 OFN에 존재한다고 가정한다. 이 때, OFN을 통해 r_O 의 전송 속도로 콘텐츠 서비스를 제공받을 수 있다. 본 연구에

* 이 논문은 한신대학교 학술연구비 지원에 의하여 연구되었음.

• First Author : (ORCID:0000-0002-5247-0296)School of Computer Engineering, Hanshin University, ywkyung@hs.ac.kr, 조교수, 정회원
논문번호 : 202104-090-B-LU, Received January 15, 2021; Revised March 23, 2021; Accepted March 23, 2021

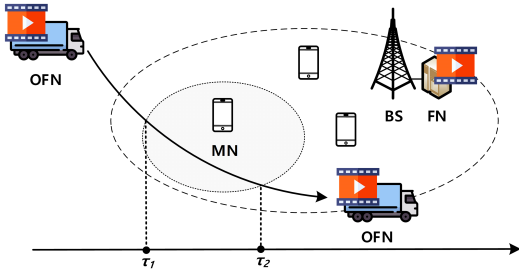


그림 1. 시스템 모델
Fig. 1. System model

서는 콘텐츠 서비스 세션 동안 OFN의 이동으로 인해 최대 한 번 연결이 가능해지는 것을 가정하였고, 다수의 OFN과 두 번 이상의 연결을 고려하는 상황은 향후 연구로 진행하고자 한다. MN이 FN과 연결이 가능한 시간을 FN 컨택 시간이라고 명명하고 FN 컨택 시간은 평균 $1/\mu_F$ 를 갖는 general 분포를 갖는다고 가정하였다⁴⁾. 또한 OFN의 이동으로 인해 MN과 연결이 가능한 시간을 OFN 컨택 시간이라고 명명하고 OFN 컨택 시간은 평균 $1/\mu_O$ 를 갖는 exponential 분포를 갖는다고 가정하였다⁴⁾. OFN 컨택 시간의 probability density function(PDF)와 cumulative density function(CDF)는 다음과 같다.

$$f_o(t) = \mu_o e^{-\mu_o t} \quad (1)$$

$$F_o(t) = 1 - e^{-\mu_o t} \quad (2)$$

그리고 사용자의 콘텐츠 요청의 수는 단위 시간 당 평균 λ 를 갖는 uniform 분포를 갖는다고 가정하였고⁵⁾, 사용자의 콘텐츠 서비스 시간은 평균 $1/\mu_d$ 를 갖는 exponential 분포를 갖는다고 가정하였다⁴⁾. 콘텐츠 서비스 시간의 PDF와 CDF는 다음과 같다.

$$f_d(t) = \mu_d e^{-\mu_d t} \quad (3)$$

$$F_d(t) = 1 - e^{-\mu_d t} \quad (4)$$

또한 FN 컨택 시간 내에서 서비스가 시작되고 나서 OFN 컨택 시간이 시작될 때까지의 시간을 t_e 라고 가정하면, t_e 는 residual life 정리⁶⁾에 의하여 다음과 같은 PDF를 갖는다.

$$f_e(t) = \mu_F(1 - F_F(t)) \quad (5)$$

III. 성능 분석

본 챕터에서는 제안하는 방법과 기존 방법의 성능 분석을 진행하고자 한다. 기존 방법들은 OFN으로의 서비스 마이그레이션을 고려하지 않았기 때문에 FN 으로부터 서비스를 제공받는 중 OFN의 활용이 가능 하더라도 지속적으로 FN으로부터 서비스를 제공받게 된다. 즉 기존 방법의 경우 OFN을 통해 서비스를 제공받는 경우는 그림 2(a)와 같이 OFN컨택 시간(t_o) 내에 서비스 세션 요청이 시작되고 종료되는 경우이다. 만약 콘텐츠 서비스가 t_o 내에서 종료되지 않으면 FN과의 세션 재설정을 통해 서비스를 재시작 해야 하기 때문에 OFN을 통한 오프로딩 효과가 없는 것으로 가정하여 해당 경우는 고려하지 않았다. 이에 더하여 제안하는 방법은 그림 2(b) 및 2(c)와 같이 FN을 통해 서비스를 제공받고 있더라도 OFN의 활용이 가능한 경우 서비스 마이그레이션을 통해 OFN의 활용이 가능하다. 이를 통해 FN의 부하 분배가 가능하다.

위에서 언급한 그림 2(a)와 2(b) 두 경우를 고려하여 OFN을 통한 콘텐츠 서비스 전송량을 계산하고자 OFN을 통한 서비스 시간의 기대치에 대한 계산이 필요하다. 먼저 그림 2(a)의 경우 콘텐츠 서비스 요청 시점인 τ_s 가 t_o 내에 발생한 경우이고 t_d 동안 콘텐츠를 OFN으로부터 받을 수 있다. 그러므로 그림 2(a)의 OFN을 통한 서비스 시간의 기대치(Et_a)는 페이지의 하단에 있는 수식(6)과 같이 계산된다. 이와 더불어

$$Et_a = \int_{t_o=0}^{\infty} \int_{\tau_s=t_f}^{\tau_s+t_o} \int_{t_d=0}^{\tau_s+t_o-\tau_s} t_d f_d(t_d) f_s(\tau_s) f_o(t_o) dt_d d\tau_s dt_o \quad (6)$$

$$Et_b = \int_{t_o=0}^{\infty} \int_{\tau_s=0}^{\tau_s+t_f} \int_{t_e=0}^{\infty} \int_{t_d=t_e}^{\tau_s+t_o} (t_d - t_e - t_m) f_d(t_d) f_e(t_e) f_s(\tau_s) f_o(t_o) dt_d dt_e d\tau_s dt_o \quad (7)$$

$$Et_c = \int_{t_o=0}^{\infty} \int_{\tau_s=0}^{\tau_s+t_f} \int_{t_e=0}^{\infty} \int_{t_d=t_e+t_o}^{\tau_s+t_o} (t_o - t_m) f_d(t_d) f_e(t_e) f_s(\tau_s) f_o(t_o) dt_d dt_e d\tau_s dt_o \quad (8)$$

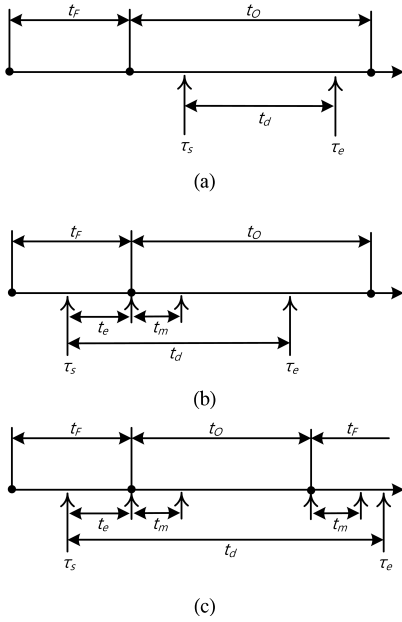


그림 2. 성능 분석을 위한 타이밍 다이어그램
Fig. 2. Timing diagrams for the performance

제안하는 방법은 그림 2(b)와 같이 t_s 가 t_o 내에 발생하더라도 OFN이 가용해지면 서비스 마이그레이션을 통해 마이그레이션 시간(t_m)을 제외한 시간인 $(t_f - t_e - t_m)$ 동안 OFN으로부터 콘텐츠 서비스를 제공받을 수 있다. 그러므로 그림 2(b)의 OFN을 통한 서비스 시간의 기대치(Et_b)는 수식(7)과 같이 계산된다. 또한 그림 2(c)와 같이 콘텐츠 서비스가 t_o 이후에 종료되는 경우에는 $(t_o - t_m)$ 동안 OFN으로부터 콘텐츠를 제공받을 수 있으며 서비스 시간의 기대치(Et_c)는 수식(8)과 같이 계산된다. FN에서 콘텐츠를 제공받다가

OFN으로 마이그레이션을 수행했을 때 콘텐츠 서비스 전송량은 $p \cdot r_o \cdot (Et_b + Et_c)$ 로 계산될 수 있고, OFN에서 콘텐츠 서비스를 시작하는 경우의 콘텐츠 서비스 전송량은 $r_o \cdot Et_a$ 로 계산될 수 있다.

한편, 전체 콘텐츠 전송량 대비 OFN을 통한 콘텐츠 전송량의 비율을 계산하기 위해서 FN을 통한 콘텐츠 전송량에 대한 계산이 필요하다. 앞서 OFN을 통한 서비스 시간의 기대치를 계산한 바와 같이 FN을 통한 서비스 시간의 기대치를 계산할 수 있다. 기존 방법에서 FN을 통한 서비스 시간의 기대치(Et_k)는 서비스 요청이 t_f 에서 발생하는 경우 및 t_o 에서 서비스 요청 발생 후 서비스 시간이 t_o 를 벗어나서 FN에게 콘텐츠 서비스 요청을 다시 하는 경우를 포함하여 본 페이지 하단의 수식 (9)와 같이 계산될 수 있다. 제안하는 방법에서는 t_f 에서 서비스 요청이 발생하는 경우 서비스가 t_f 내에서 종료되는 경우, t_o 에서 종료되는 경우, 그리고 t_o 를 지나 t_f 에서 종료되는 경우를 포함해야 한다. 또한 t_o 에서 서비스 요청이 발생하는 경우에는 서비스가 t_f 에서 종료되는 경우를 포함해야 한다. 이러한 경우들을 고려한 제안하는 방법의 FN을 통한 서비스 시간의 기대치(Et_{fp})는 수식 (10)과 같이 계산된다.

그러므로 제안하는 방법의 전체 콘텐츠 전송량 대비 OFN을 통한 콘텐츠 전송량의 비율(R_o)은 $(p \cdot r_o \cdot (Et_b + Et_c) + r_o \cdot Et_a) / (p \cdot r_o \cdot (Et_b + Et_c) + r_o \cdot Et_a + r_f \cdot Et_{fp})$ 로 계산될 수 있으며 기존 방법의 R_o 는 $(r_o \cdot Et_a) / (r_o \cdot Et_a + r_f \cdot Et_k)$ 로 계산될 수 있다.

IV. 성능 결과

이 장에서는 MATLAB R2018a를 기반으로

$$Et_{fc} = \int_{t_o=0}^{\infty} \int_{\tau_s=0}^{t_f} \int_{t_d=0}^{\infty} t_d f_d(t_d) f_s(\tau_s) f_o(t_o) dt_d d\tau_s dt_o + \int_{t_o=0}^{\infty} \int_{\tau_s=t_f}^{t_f+t_o} \int_{t_d=t_f+t_o-\tau_s}^{\infty} t_d f_d(t_d) f_s(\tau_s) f_o(t_o) dt_d d\tau_s dt_o \tag{9}$$

$$Et_{fp} = \int_{t_o=0}^{\infty} \int_{\tau_s=0}^{t_f} \int_{t_d=0}^{t_f-\tau_s} t_d f_d(t_d) f_s(\tau_s) f_o(t_o) dt_d d\tau_s dt_o + \int_{t_o=0}^{\infty} \int_{\tau_s=0}^{t_f} \int_{t_e=0}^{\infty} \int_{t_d=t_e}^{t_f+t_o} t_e f_d(t_d) f_e(t_e) f_s(\tau_s) f_o(t_o) dt_d dt_e d\tau_s dt_o + \int_{t_o=0}^{\infty} \int_{\tau_s=0}^{t_f} \int_{t_e=0}^{\infty} \int_{t_d=t_e+t_o}^{2t_f+t_o} (t_d - t_o) f_d(t_d) f_e(t_e) f_s(\tau_s) f_o(t_o) dt_d dt_e d\tau_s dt_o + \int_{t_o=0}^{\infty} \int_{\tau_s=t_f}^{t_f+t_o} \int_{t_d=t_f+t_o-\tau_s}^{\infty} (\tau_s + t_d - t_f + t_o) f_d(t_d) f_s(\tau_s) f_o(t_o) dt_d d\tau_s dt_o \tag{10}$$

event-driven 시뮬레이터를 개발하여 extensive 시뮬레이션을 수행하였다. 각 분포를 고려하여 t_o , t_f , t_d 에 대해 약 3만 번의 랜덤 값 생성을 수행하여 t_d 와 t_b 의 평균값을 구했고, 이를 통해 콘텐츠 서비스 전송량을 계산하였다. μ_o , μ_b , p , r_f , r_o , λ 의 디폴트 값은 각각 10, 5, 0.5, 100Mbps, 100Mbps, 10으로 가정하였다.

그림 3은 $1/\mu_F$ 가 증가할 때, 제안하는 방법과 기존 방법의 OFN을 통한 콘텐츠 전송량을 나타낸다. 그림을 통해 시뮬레이션 결과(S)와 수식 분석 결과(A)가 동일함을 확인할 수 있었으며, $1/\mu_F$ 가 증가하면서 OFN 컨택 시간이 줄어들기 때문에 전체적으로 콘텐츠 전송량이 줄어드는 것을 볼 수 있다. 하지만 제안하는 방법은 서비스 마이그레이션을 통해 FN 컨택 시간 내에서 시작된 서비스들도 OFN을 활용할 수 있기 때문에 개선된 성능을 보인다. 또한 제안하는 방법은 마이그레이션 시간이 증가할수록 OFN 컨택 시간 활용도가 감소하기 때문에 성능저하가 발생하는 것을 볼 수 있다. 그러므로 사이즈, 통신환경 등에 영향을 받는 마이그레이션 시간³⁾ 예측을 통해 마이그레이션 수행 여부를 결정하는 효율적인 운영이 필요하다.

그림 4는 $1/\mu_F$ 가 증가할 때, 제안하는 방법과 기존 방법의 전체 콘텐츠 전송량 대비 OFN을 통한 콘텐츠 전송량의 비율을 보여준다. 그림에서 알 수 있듯이 제안하는 방법은 마이그레이션을 고려하기 때문에 기존 방법보다 많은 양의 콘텐츠를 OFN을 통해 서비스 받을 수 있는 것을 확인할 수 있다. 예를 들어 $1/\mu_F$ 이 $1/\mu_o$ 보다 5배 큰 경우에 기존 방법은 약 16.7%의 콘텐츠가 OFN을 통해 수신될 수 있지만 제안하는 방법의 경우 마이그레이션 시간이 서비스 시간의 0.5를 차지하는 경우에도 약 36.2%의 콘텐츠를 OFN을 통해 수

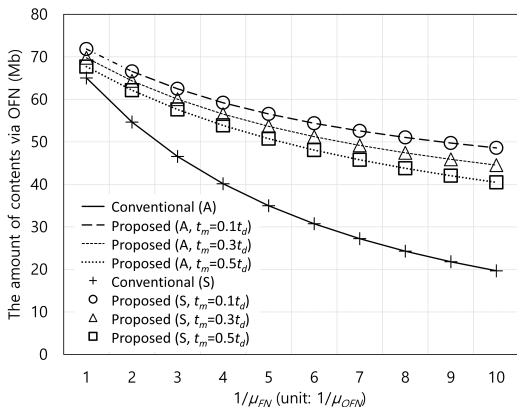


그림 3. OFN을 통한 콘텐츠 전송량
Fig. 3. The amount of contents via OFN

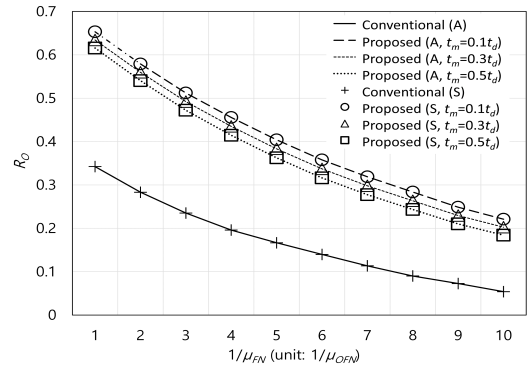


그림 4. 전체 콘텐츠 전송량 대비 OFN을 통한 전송량
Fig. 4. The ratio of the amount of contents through OFN to that of total contents

신할 수 있다.

V. 결론

본 논문에서는 FN에서 제공받던 서비스를 OFN이 가용할 때 서비스 마이그레이션을 통하여 서비스 연속성을 제공하면서 부하를 분산시킴으로써 OFN의 활용도를 높이고자 하였다. 성능 분석을 통해 제안된 방법이 기존 방법보다 OFN을 통한 콘텐츠 전송량이 개선됨을 확인하였고 향후 마이그레이션 시간 및 콘텐츠 존재 확률 등을 고려한 최적화 연구를 진행할 예정이다.

References

- [1] Y. Liu, M. Peng, G. Shou, Y. Chen, and S. Chen, "Toward edge intelligence: Multiaccess edge computing for 5G and internet of things," *IEEE Internet of Things J.*, vol. 7, no. 8, pp. 6722-6747, Aug. 2020.
- [2] H. Wu, J. Zhang, Z. Cai, F. Liu, Y. Li, and A. Liu, "Toward energy-aware caching for intelligent connected vehicles," *IEEE Internet of Things J.*, vol. 7, no. 9, pp. 8157-8166, Sep. 2020.
- [3] W. Bao, D. Yuan, Z. Yang, S. Wang, W. Li, B. B. Zhou, and A. Y. Zomaya, "Follow me fog: Toward seamless handover timing schemes in a fog computing environment," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 72-78, Nov. 2017.

-
- [4] D. Suh, H. Ko, and S. Pack, "Efficiency analysis of wifi offloading techniques," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3813-3817, May 2016.
 - [5] J. Dai, Z. Hu, B. Li, J. Liu, and B. Li, "Collaborative hierarchical caching with dynamic request for massive content distribution," in *Proc. IEEE INFOCOM*, vol. 8, no. 6, pp. 1550-1554, Dec. 2012.
 - [6] S. M. Ross, *Stochastic Processes*, 2nd Ed., NY: John Wiley & Sons, 1996.