

SDN에서 SVC를 이용한 강화 학습 기반 하이브리드 비디오 스트리밍 기법

안준범, 염상길, 추현승
 성균관대학교 소프트웨어대학
 e-mail:{joonbeom, sanggil2, choo}@skku.edu

Hybrid Video Streaming using SVC based on Reinforcement Learning in SDN

Joonbeom Ahn, sanggil Yeom, Hyunseung Choo
 College of Software, Sungkyunkwan University

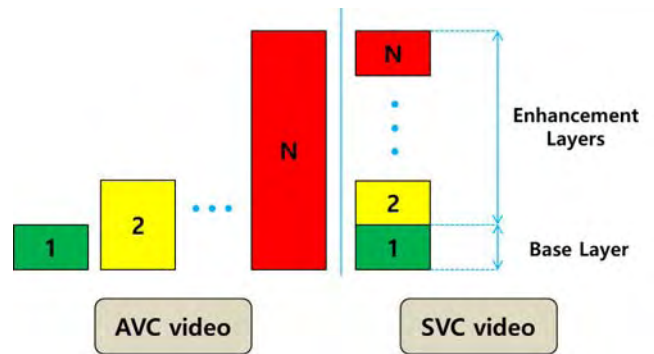
요 약

Scalable Video Coding(SVC)은 효율적인 스트리밍을 제공하기 위해 제안된 비디오 코딩 방식으로 비디오 파일을 계층적으로 구분을 하여 기초 계층에 항상 계층을 추가하여 비디오의 해상도, 프레임 재생율 그리고 화질을 개선시킬 수 있다. 현재 Software Defined Networking(SDN) 환경에서 SVC를 이용한 다양한 라우팅 기법 연구가 진행되고 있다. 본 논문은 SDN 환경에서 네트워크 상태에 따라 하나 또는 여러 개의 경로로 전송을 하는 하이브리드 비디오 스트리밍을 제안한다. 제안 기법에서 경로의 선택은 SVC의 계층별로 진행한다. 제안 기법을 통해 비디오 세그먼트의 손실율을 줄여 높은 QoE를 제공할 수 있고 비디오의 끊김 현상을 최소화 할 수 있다.

1. 서론

최근 멀티미디어 어플리케이션은 인터넷에서 가장 많은 네트워크 트래픽을 사용하는 어플리케이션이다. 따라서 멀티미디어 어플리케이션이 제공하는 비디오 스트리밍의 품질을 높이고자 하는 다양한 연구가 진행 중이다. 멀티미디어 어플리케이션에서 비디오 스트리밍을 위해 가장 일반적으로 사용되는 비디오 코딩 방식은 H.264/AVC이다. 비디오 스트리밍에서 H.264/AVC 방식은 네트워크 상태에 적용할 수 있도록 품질마다 다른 파일로 저장하여 보관한다. 하지만 H.264/AVC는 인코딩의 특성상 약간의 패킷 유실이 발생하더라도 영상 재생 중에 프레임 일부가 깨지는 현상이 발생 할 수 있다. 또한, 품질별로 여러 개의 개별적인 파일로 저장되기 때문에 저장 공간의 낭비가 발생한다. 이러한 문제점을 해결하기 위해 H.264/Scalable Video Coding (SVC)[1]이 등장하였다. 비디오 스트리밍에서 SVC는 하나의 비디오 세그먼트를 계층적으로 구성하는 비디오 코딩 방식이다. SVC의 계층들 중 기초 계층(base layer)은 최저 품질을 담당하며 항상 계층(enhancement layer)들은 기초 계층에 순서대로 추가되어 비디오의 품질을 향상시키는 역할을 한다. SVC는 계층적인 구조를 가지고 있기 때문에 파일의 중복성 문제를 해결할 수 있고 기초 계층의 전송을 보장함으로써 디코딩 오류를 개선시킬 수 있다.

Software Defined Networking(SDN)[2]은 스위치나 라우터와 같은 네트워크 하드웨어 장비에서 단순히 패킷 포워딩을 처리하는 부분인 데이터 플레인(data plane)과



(그림 1) SVC 컨셉

패킷의 경로 설정 또는 제어를 담당하는 컨트롤 플레인(control plane)을 분리하는 차세대 네트워킹 기술이다. SDN 환경에서 네트워크는 하나의 컨트롤러에 연결된 스위치들의 집합으로 나타낼 수 있으며 컨트롤러는 스위치들을 중앙 제어한다. 또한, SDN 컨트롤러는 프로그래밍이 가능하기 때문에 서로 다른 어플리케이션의 특성을 고려한 네트워크 환경 구현이 가능하며 최근 SDN 환경에서 SVC 비디오 스트리밍의 특성을 고려한 라우팅 기법의 연구가 활발히 진행 중이다.

SVC 비디오 스트리밍의 QoE를 개선시키는 방법은 크게 두 가지로 구분할 수 있다. 체감 품질을 높이기 위하여 서버-클라이언트 전송 중 잃는 프레임의 수를 줄이는 방법과 경로의 변경 횟수를 줄여 영상의 끊김 현상을 줄이는 방법이다. 본 논문은 사용자의 QoE를 개선시키고 최고

품질의 영상을 제공하기 위해 하나 또는 여러 개의 경로를 이용하여 전송하는 비디오 스트리밍 기법을 제안한다.

2. 관련 연구

SDN 환경에서 비디오 스트리밍의 QoE를 효과적으로 개선하기 위한 다양한 연구들이 있다. 끊임 없는 영상을 사용자에게 제공하기 위하여 기초 계층과 향상 계층의 경로를 구분하여 전송하는 기법[3]이 있다. 이 기법에서 기초 계층은 제안한 라우팅 알고리즘을 통해 계산된 최적의 경로를 통해 전송된다. 품질을 높이는 향상 계층은 기초 계층을 전송하는 경로를 제외한 경로들 중 최적인 경로로 전송된다. 최적인 경로는 대역폭의 로드를 기준으로 계산되므로 기초 계층의 전송은 보장된다. 그렇기 때문에 서버는 연속적인 영상을 제공할 수 있다. 하지만 향상 계층의 수가 증가함에 따라 패킷 지연이 발생하게 되며 계층을 증가할수록 많은 양의 패킷 유실이 발생하기 때문에 QoE를 감소시키는 요인이 될 수 있다. SVC의 기초 계층과 향상 계층만을 구분하여 전송하는 것이 아닌 각각의 계층을 다른 경로로 전송하는 제안[4]이 있다. 이 제안은 링크의 패킷 유실율과 딜레이를 기준으로 경로를 계산한다. SVC는 향상 계층이 기초 계층에 의존적이기 때문에 기초 계층의 패킷 유실은 상위 계층에도 큰 영향을 미친다. 그렇기 때문에 하위계층의 패킷 유실율은 낮아야한다. 또한 상위 계층은 하위 계층과 비슷한 시간에 도착해야 디코딩 될 수 있기 때문에 딜레이는 상위 계층에 더 큰 영향을 미친다. 따라서 링크 딜레이와 패킷 유실율을 계층별 우선순위로 계산하여 계산된 경로를 각각의 계층에 할당한다. 하지만 이 제안은 링크의 잔여 대역폭을 고려하지 않았기 때문에 전체적인 네트워크 상황이 악화될 경우 많은 양의 패킷 유실을 발생시킬 수 있다.

비디오 스트리밍 기술의 Quality of Experience(QoE)를 개선하기 위한 연구로 강화 학습 알고리즘을 이용하여 비디오 품질 계층의 추가/삭제 또는 경로 변경을 결정하는 기법[5]이 있다. 이 기법은 전송 경로와 후보 경로의 비교를 통해 상태를 정의하고, 각 상태에서 수행 가능한 동작을 선택한다. 또한, 반복적인 학습을 통해 프레임 손실을 줄여 QoE를 높일 수 있는 최적의 결정을 할 수 있다. 하지만 갑작스럽게 네트워크 상태가 저하될 경우 프레임 손실이 극대화되며, 비디오 멈춤 현상을 발생시킬 수 있다.

3. 강화 학습

강화 학습[6]은 탐색 단계에서 환경 또는 시스템에 대한 정보를 수집하여 이 정보를 추가적인 실행 단계에 사용한다. 강화학습에서 의사 결정은 마코브 결정 프로세스로 모델링된다. 의사 결정은 모든 상태(state) S 에서 t 지점에 있는 상태를 s_t 로 정의한다. 상태 s_t 에서 선택할 수 있는 액션(action)들의 집합은 A 로 나타내며 상태 s_t 에 대한 액션 a_t 를 선택한다. 선택된 액션을 동작하며 다음상태 s_{t+1} 로

이동하며 보상(reward) r_{t+1} 을 획득하게 된다. 강화학습은 이러한 과정을 반복하여 최적의 결정을 찾는다. 이때 모든 행동에 대한 보상을 계산하기 위해 Q-learning 기술[7]을 사용한다. Q-learning 기술에서는 Q-table에 $Q(s, a)$ 의 형태로 상태 s 에 대한 동작 a 의 보상 값을 보관하며 업데이트 한다.

$$Q(s_t, a_t) = (1 - \eta) Q(s_t, a_t) + \eta(r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})) \quad (1)$$

Q 값은 (1)에 주어진 함수에 따라 업데이트된다. (1)의 공식에서, γ 와 η 는 각각 할인 계수(discount factor)와 학습 계수(learning factor)를 나타낸다. 할인 계수는 현재 및 미래의 보상의 중요성을 결정하며, 0과 1 사이의 값을 사용한다. 할인 계수가 1에 가깝다면 강화 학습 모델은 새로운 보상에 높은 가중치를 부여합니다. 할인 계수와 마찬가지로 학습 계수는 0과 1 사이의 값을 설정한다. 학습 계수가 1에 가깝다면 이것은 다음 상태 s_{t+1} 의 보상이 Q 값에 큰 영향을 미치고 모델이 빠르게 학습한다는 것을 의미한다.

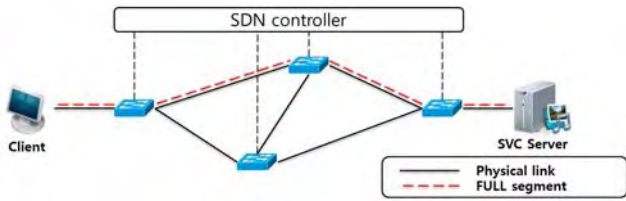
$$P(a|s) = \frac{\exp[Q(s, a)/T]}{\sum_{b \in A} \exp[Q(s, b)/T]} \quad (2)$$

Q 값은 처음에 0으로 초기화된다. 이는 모든 동작의 선택 확률이 동일하다고 설명한다. Q 값은 (1)에 주어진 함수에 따라 업데이트되지만, 각 액션의 확률을 계산하기 위해 (2)의 Softmax 함수를 사용하며 Softmax 함수의 변수인 T 파라미터는 Q 값이 업데이트 되면서 점차적으로 감소한다. T 값이 감소되면 확률의 차이를 더욱 크게 보이기 때문에 실행 단계로 자연스럽게 전환된다. 이는 탐색 단계에서 모든 작업의 선택 확률은 거의 동일하지만 실행 단계에서는 의사 결정자가 Q 값이 높은 작업을 선택하는 이유다. T 값과 마찬가지로 할인 계수와 학습 계수의 감소를 설정할 수 있으며 시간에 따라 Q-값에 의한 최적의 결정을 내릴 수 있다.

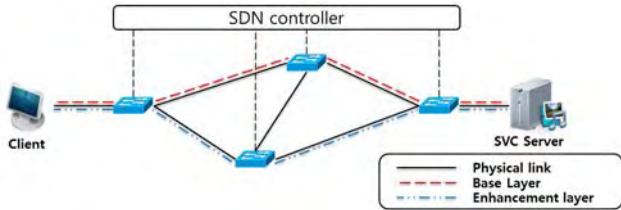
4. 강화 학습 기반 하이브리드 비디오 스트리밍

사용자에게 높은 품질의 영상을 제공하기 위해 가용 대역폭에 따라 두 가지 경우로 구분하여 전송한다. 첫 번째는 SVC 세그먼트의 최대 계층을 수용할 수 있는 잔여 대역폭을 가진 경로가 있는 경우이다. 해당 경우 경로를 나눠서 전송하는 것보다 하나의 경로를 이용하여 전송하는 것이 효율적이다. 두 번째는 여러 개의 경로를 이용하여 전송하는 경우이다. 여러 개의 경로로 전송을 하게 되면 SVC의 특성상 하위계층에 대한 의존성 때문에 하위계층의 패킷 loss가 가장 낮아야한다. 또한 품질을 향상시키는 상위 계층은 하위 계층과 같은 시간에 디코딩 될 수 있도록 비슷한 시간에 클라이언트에 도착해야만 한다. 컨

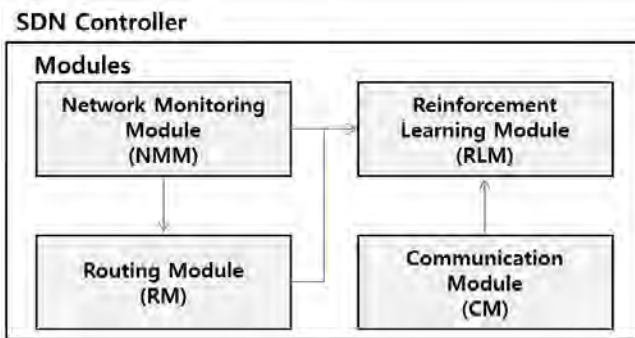
트roller는 비디오 스트리밍이 진행되는 동안 지속적으로 네트워크 상태를 모니터링하며 경로의 수를 결정한다.



(그림 2) 경로 하나를 이용한 전체 세그먼트 전송



(그림 3) 여러 경로를 이용하여 계층 단위로 전송



(그림 4) 하이브리드 비디오 스트리밍을 위한 SDN 컨트롤러 모듈

그림 2와 3은 제안하는 하이브리드 비디오 스트리밍의 개념을 나타낸다. 그림 2는 하나의 경로를 이용하여 전체 계층을 포함한 세그먼트의 전송을 나타낸다. 각 스위치들은 컨트롤러와 연결되어 있으며 컨트롤러가 설치한 포워딩 룰에 따라 패킷을 전송한다. 클라이언트는 수신한 세그먼트를 디코딩하여 재생할 수 있다. 그림 3은 빨간 점선과 파란 점선으로 경로를 나눠서 전송하는 모습을 나타낸다. 빨간 점선의 경로로 기초 계층을 전송하며 파란 점선으로 향상 계층의 전송을 담당한다. 각각의 경로는 서로 다른 소켓으로 전송하기 때문에 UDP 포트 번호를 이용하여 플로우를 구분한다. 컨트롤러는 하이브리드 비디오 스트리밍을 제공하기 위해 비디오 서버와 서로 통신하여 전송 가능한 비디오들에 관련된 정보를 주고받는다. 또한 컨트롤러는 각 경로에 할당된 계층의 정보를 비디오 서버에 전달한다.

컨트롤러에서 사용하는 모듈은 총 네 가지로 그림 4에 나타낸다. 컨트롤러는 Communication Module(CM)을 통해 비디오 서버와 통신하며 서버에 저장된 비디오 계층 등의 정보를 사전에 주고받는 역할을 한다. 또한

Reinforcement Learning Module(RLM)에 의해 할당된 경로별 포트 정보를 주고받는다. Network Monitoring Module(NMM)은 각 링크들의 잔여 대역폭과 패킷ロス 확률, 딜레이 등의 정보들을 수집하여 하나의 경로로 보낼지 여러 개의 경로로 보낼지 결정한다. Routing Module(RM)은 경로의 계산을 담당하는 모듈이다. RLM은 Q-Table을 저장하고 있으며 여러 개의 경로를 통해 전송할 경우 네트워크 상태에 따라 계층별로 적절한 경로를 할당한다.

4.1 라우팅 알고리즘

하이브리드 비디오 스트리밍 기법에서 라우팅 알고리즘은 경로가 하나인 경우와 여러 개인 경우로 구분한다. 잔여 대역폭은 경로의 수를 구분하기 위한 척도가 된다. 하나의 경로로 전송하는 경우 경로를 구하기 위해 다익스트라 알고리즘을 사용한다. 각 링크의 가중치는 잔여 대역폭 W_{band} 를 이용하여 식(3)을 통해 계산한다. P_{loss} 는 패킷 유실 확률을 나타낸다.

$$W = (1 - P_{loss}) \times W_{band} \quad (3)$$

라우팅 알고리즘으로 계산된 최적의 경로를 비디오 세그먼트의 최대 전송률과 비교하여 경로의 선정을 결정하며 최대 계층을 수용하지 못할 경우 여러 개의 경로를 이용하는 방법을 사용한다. 이때 경로는 (4)의 식을 가중치로 갖는 다익스트라 알고리즘으로 계산한다.

$$W = (1 - P_{loss}) \times W_{band} + \alpha \times (1 - 2^{-v_i}) \times D \quad (4)$$

식 (4)에서 v_i 는 계층별 우선순위를 나타내며 상위 계층일수록 높은 수를 갖는다. D 는 딜레이를 의미하며 상위 계층일수록 딜레이의 영향을 크게 받기 때문에 계층별 우선순위를 고려하여 가중치 값을 갖는다. 제안 알고리즘의 최종 목적은 사용자에게 최고 품질의 영상을 제공하는 것이다. 따라서 프레임 유실에 영향을 미치는 패킷 유실 확률과 모든 계층이 제시간에 도착하여 디코딩되기 위한 딜레이 그리고 계층의 전송을 수용하기 위한 잔여 대역폭을 이용한다.

4.2 강화학습 모델링

하이브리드 비디오 스트리밍 기법에서 강화학습은 여러 경로로 비디오를 전송하기 위해 사용된다. 제안 기법에 강화학습을 적용시키기 위하여 상태, 동작 그리고 보상에 대한 정의가 필요하다. 상태는 계층의 번호와 할당된 경로로 결정한다. 각 상태에서 선택할 수 있는 액션은 해당 계층에 적절한 경로를 할당한다. 액션을 선택했을 때 얻을 수 있는 보상은 선택된 경로로 전송될 수 있는 전송량으로 계산한다.

(알고리즘 1) 제안 기법 처리과정

- 1 **Initialize** Q-table
- 2 Repeat
- 3 **Observe** state **s**
- 4 Calculate probability of actions in state **s** using softmax
- 5 **Choose** action **a** based on the probability
- 6 Observe reward **r** and new state **s'**
- 7 Calculate Q-value using Q-learning algorithm
- 8 **Update** Q-table
- 9 Until streaming ends

제안 기법의 처리과정은 알고리즘 1에 나타난다. 초기 시스템이 시작되면 Q-table을 초기화한다. 컨트롤러는 주기적으로 강화학습의 상태를 파악하며 파악된 상태에 따라 Softmax 함수에 의해 계산된 확률에 따라 최적의 액션을 선택한다. 선택된 액션을 동작시키고 보상을 획득하며 다음 상태를 파악한다. 컨트롤러는 획득한 보상을 이용하여 Q-learning 알고리즘으로 Q 값을 계산하여 Q-Table을 업데이트 한다. 스트리밍이 끝날 때 까지 위의 절차를 반복하며 각 계층에 대한 최적의 경로를 파악하여 여러 경로를 통해 최적의 품질을 클라이언트에 제공한다.

4. 결론

하이브리드 비디오 스트리밍 기술은 높은 QoE와 끊임 없는 영상을 제공하기 위한 목적을 가진 기법이다. 끊임 없는 영상을 제공하기 위해 패킷 유실 확률을 고려한 라우팅 알고리즘으로 비디오 프레임의 유실율을 줄일 수 있고, 높은 품질을 제공하기 위해 여러 개의 경로를 이용하여 네트워크 상태에 따라 최대 품질의 영상을 제공한다. 제안하는 기법을 구현하기 위해 floodlight와 mininet을 이용한 SDN 네트워크를 구현하여 각각의 모듈들의 개발을 진행하고 있다. 향후 개발을 완료하여 관련 연구의 기법들과 성능을 비교하여 개선점을 확인할 예정이다.

ACKNOWLEDGEMENT

본 논문은 기초연구사업 (NRF-2010-0020210)과 과학기술정보통신부 및 정보통신기술진흥센터의 Grand ICT연구센터지원사업 (IITP-2017-2015-0-00742), 방송통신인프라 원천기술개발사업 (B0101-17-1366, 자율 제어 네트워킹 및 자율 관리 핵심 기술 관리)의 연구결과로 수행되었음

참고문헌

- [1] H.Schwarz, D.Marpe, and T.Wiegand "Overview of the scalable video coding extension of the H.264/AVC standard" Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, no. 9, pp. 1103-1120, 2007
- [2] Open Networking Foundation(ONF), "Software

defined networking: the new norm for networks", White Paper, 2012

- [3] S. Laga, T. Van Cleemput, F. Van Raemdonck, F. Vanhoutte, N. Bouten, M. Claeys, and F. De Turck "Optimizing scalable video delivery through OpenFlow layer-based routing" IEEE Network Operations and Management Symposium (NOMS), pp. 1 - 4, 2014
- [4] Ankit Gangwal, Megha Gupta, Manoj Singh Gaur, Vijay Laxmi, and Mauro Conti "ELBA: Efficient Layer Based routing Algorithm in SDN" International Conference on Computer Communications and Networks(ICCCN), 2016
- [5] Tuba Uzakgider, Cihat Cetinkaya, and Muge Sayit "Learning-based approach for layered adaptive video streaming over SDN" Computer Networks, 92, pp. 357-368, 2015.
- [6] R. Sutton and A. Barto, "Reinforcement Learning: An Introduction", Cambridge, MA: The MIT Press, 1988.
- [7] D.P. Bertsekas, Dynamic Programming and Optimal Control, II, 3rd ed., Athena Scientific, 2007