링크 안정성을 고려한 강화학습 기반 라우팅

염성웅, 홍 남 퀘엣, 김경백
전남대학교
yeomsw0421@gmail.com, quachhongnam1995@gmail.com, kyungbaekkim@jnu.ac.kr

Reinforcement learning-based Optimal Routing with Link Stability

Sungwoong Yeom, Hong-Nam Quach, Kyungbaek Kim
Dept. Artificial Intelligence Convergence, Chonnam National University.

요 약

점점 복잡해지는 네트워크에 따라 새로운 종단 간 서비스를 시작하는 것은 비용이 많이 필요할 수 있다. 기존의 Dijkstra 기반 라우팅 알고리즘은 복잡한 네트워크 토폴로지 전체를 탐색하여 종단 간의 서비스를 시작할 때 비용이 높을 수 있다. 이를 위해 시간 복잡성이 낮은 강화학습을 기반으로 라우팅을 한다면 서비스를 시작할 때 발생하는 비용을 줄일 수 있다. 이때, 네트워크상의 링크의 상태를 고려한다면 보다 최적의 라우팅이 가능하다. 본 논문에서는 링크 안정성을 고려하는 Q 라우팅 기반 최적의 라우팅 기법을 제안한다. 이 기법은 종단 간 서비스의 링크들로부터 대역폭, 지연 및 패킷 드랍율을 사용하여 링크 안정성을 계산하고 높은 링크 안정성을 보여주는 링크들을 Q 라우팅 알고리즘을 통해 최적의 라우팅을 한다.

Ⅰ. 서 론

최근 IoT 기기의 활성화에 따라 네트워크가 복잡해짐으로써 인터넷 애플리케이션의 서비스 QoS 요구 사항은 네트워크 트래픽 증가하고 품질이 다양하고 새롭게 생성된 패킷 흐름의 전송 연구가 활성화되고 있다 [1]. 하지만, 제한된 정보를 사용하여 라우팅 경로를 설정하기 때문에 동적 트래픽 변경에 대한 적응 속도가 느려지고 다양한 QoS 요구사항에 대한 지원이 제한될 수 있다.

최근 SDN (Software-defined networking)의 활성화는 네트워크 제어 및 패킷 포워딩, 프로그래밍 기능, 글로벌 뷰, 논리적 중앙 제어와 같은 SDN 기능을 활용하여 다양한 정보를 수집하여 라우팅 프로토콜의 성능을 향상 시킬 수 있다 [2]. BFS (Breadth First Search) 알고리즘 중 하나인 다익스트라 (Dijkstra) 기반의 라우팅 기법은 네트워크 토폴로지 전체에서 주어진 정점에 대해 하나의 소스 노드에서 하나의 대상 노드까지의 최단 경로를 찾을 수 있다 [3]. 하지만, 네트워크가 점점 복잡해진다면 새로운 서비스를 활성화할 때 비용이 많이 필요할 수 있다. 새로운 서비스를 활성화할 때 발생하는 비용을 줄이기 위해서 강화학습 (Reinforcement Learning) 중 하나인 Q 라우팅을 사용한다면 최적의 경로를 탐색할 수 있다 [4-6].

본 논문에서는 지능형 라우팅을 위해 SDN 환경에서 링크 안정성을 고려하여 Q 라우팅을 기반으로 최적의 패킷 경로 탐색 기법을 제안한다. 제안된 기법은 SDN을 활용하여 지식 평면을 추가하고 동적 트래픽 변경 중에도 지능형 라우팅에 대한 잠재적 경로를 탐색, 학습 및 활용하기 위해 이용 가능한 대역폭, 링크 지연, 패킷 손실과 같은 링크 상태 정보를 수집하고 링크 상태 정보를 사용하여 링크 안정성을 계산한다. 계산된 링크 안정성을 Q 라우팅에 적용한다면 최적의 경로를 탐색하는 것이 가능하다.

Ⅱ. 배경

Q 라우팅 알고리즘은 환경 (Environment)과 상호작용을 하는 에이전트 (Agent)를 학습시킨다. 에이전트는 상태 (State)라고 부르는 네트워크 상황에 대한 행동 (Action)을 취한다. 먼저, 에이전트는 상태와 행동으로 구성된 Q 테이블 S,A를 최적의 상태로 전환하기 위해 일련의 스텝으로 구성된 에피소드 (Episode)를 정의한다. 이 에피소드에 정의된 스텝에 따라 행동 A 을 선택하고 상태 S_t을 변환하고 최적의 보상 (Reward) R_t 을 통해 최적의 Q 값 $Q_t(S_t,A_t)$을 근사화하는 최상의 정책 (Policy)을 설정한다. 따라서 Q 라우팅 알고리즘의 시간 복잡도는 네트워크 전체 노드를 탐색하는 다익스트라의 시간 복잡도에 비해 비교적 낮다. 이러한 Q 라우팅 알고리즘은 최적의 네트워크를 탐색하는데 사용될 수 있다. 이때, Q 라우팅 알고리즘의 에이전트가 여분의 대역폭은 많고 지연은 적으며 패킷 드랍이 적은 링크일수록 링크 안정성이 높은 링크들로 구성된 경로를 탐색할 필요가 있다. 본 논문에서는 링크 안정성을 고려하여 최적의 라우팅을 설정하는 기법을 제안한다.

Ⅲ. 링크 안정성을 고려한 강화학습 기반 최적의 경로 탐색 기법

본 논문에서는 링크 안정성을 고려하여 강화학습을 기반으로 최적의 경로를 탐색하는 기법을 제안한다. SDN(Software-Defined Networking)의 데이터 평면에 구성된 스위치 i는 상태 S_i로 나타내고 이러한 상태는 상태 공간 $S_i \in S$에서 관리한다. 이 상태 공간 S로 구성된 토폴로지는 데이터 평면의 스위치 토폴로지에 해당하는 그래프로서 에이전트에 전달된다. 상태 S_i의 인접 상태는 해당 스위치의 인접 스위치에 해당한다. 상태 S_i에서 상태 S_j로의 변환은 스위치 i와 스위치 j를 연결하는 링크에 해당한다. 에이전트는 각 상태 S_i에 대해 일련의 행동 중 하나 A_i를 수행하고 행동은 행동 공간 $A_i \in A$에서 관리한다. 행동 A_i은 현재 상태 S_i를 상태 S_j로 전환한다. 보상 R_t은 네트워크상에서 최상의 경로를 탐색을 위해 지식 평면으로부터 수집한 세 가지 특징인 이용 가능한 여분의 링크 대역폭 bws_t, 링크 지연 d_t 및 링크 패킷 손실 l_t을 사용하여 링크 안정성을 계산하고 이를 수식 (1)과 같이 나타낸다.

$$R = w_1 \quad bws + w_2 \quad (1-d) + w_3 \quad (1-\hat{l}) \qquad (1)$$

이때, , w_2 및 w_3 값은 보상 $_t$을 계산하기위한 행렬에 대한 가중치를 나타내는 매개변수이다. 가중치 w는 수식 (2)과 같이 나타낸다.

$$w_1+w_2+w_3=1, w_1,w_2,w_3\in[0,1] \quad (2)$$

수식 (1)에서 bws, d 및 $\hat{l}$은 각각 여분의 대역폭, 지연 및 패킷 손실 값을 정규화한 값이다. 에이전트의 학습 과정에서 각 링크 상태들은 각각 다른 단위로 구성되어 있기 때문에 각 특징들을 정규화를 시키기 위해 Min-Max 기법을 사용한다. 정규화에 대한 예시는 수식 (3)과 같이 나타낸다.

$$x_i=\frac{x_i-\min(X)}{\max(X)-\min(X)}, x_i\in X, 1 \quad i \quad m \quad (3)$$

에이전트는 학습률 α, 보상 R_t 및 새로운 상태 S_{t+1}를 사용하여 Q 값을 조정한다. Q 값에 대한 수식 (4)과 같이 나타낸다.

$$Q_{t+1}(S_t,A_t)=Q_t(S_t,A_t)+\alpha\times[R_t+\max Q_t(S_{t+1},A_t)-Q_t(S_t,A_t)] \quad (4)$$

에이전트는 누적 보상을 최대로 만들기 위해 예상되는 최적의 행동을 선택하는 탐색(exploration) 및 미래에 더 큰 보상을 얻을 수 있기를 바라면서 다른 행동을 선택하는 이용(exploitation)을 반복한다. 하지만, 에이전트가 탐색하지 않는다면 보상 값을 최대로 만들지 못할 수 있고, 탐색만 한다면 기존의 경험을 사용하지 못해 비효율적이다. 이러한 탐색과 이용의 트레이드오프 딜레마를 해결하기위해 에이전트는 학습 프로세스 중 ε 확률로 이용을 진행하고 $1-\varepsilon$ 확률로 탐색을 진행하는 ε -greedy를 사용한다. ε -greedy를 사용하여 정의된 행동은 수식 (5)와 같다.

$$A=\begin{cases}\min Q_t(S_t,A), & \text{if } x<\epsilon\\ another\ action, else\end{cases} \quad (5)$$

에이전트는 모든 쌍의 행동-상태를 거쳐 최적의 Q 함수 $Q_{t+1}(S_t,A_t)$를 근사한다. 노드 쌍에 대해 근사화된 Q 값은 Q 테이블에 업데이트하고 저장 후 에피소드가 종료된다. 이 정책은 Q 값을 최소화하기 위해 사용 가능한 대역폭이 크고 링크 지연과 패킷 손실이 낮은 링크의 우선순위로 채택한다.

Ⅲ. 토론

제안된 알고리즘은 특정 상황의 네트워크 토폴로지 상에서 운용되고 있는 서비스가 있을 경우 서비스 운용 상태에 따라 보상 함수의 매트릭스가 규정이 된다. 하지만, 네트워크 서비스 상황은 지속적으로 변하기 때문에 보삼 함수가 고정이 되어있지 않을 수 있다. 따라서 고정되어있지 않은 보상 테이블을 사용하여 Q 테이블을 학습가능한지 확인할 필요가 있다. 만약 이러한 보상 테이블을 사용하여 Q 테이블을 학습 가능하다면 특정 종단 간 서비스를 위해 취적의 경로를 탐색할 수 있다. 하지만, 기존에 활성화된 종단 간 서비스의 경로는 최적이 아닐 수 있다. 따라서 네트워크상에 새로운 종단 간 서비스가 운용이 된다면 이전에 운용되던 종단 간 서비스의 경로 또한 탐색할 필요가 있다.

Ⅵ. 결론

본 논문에서는 링크 안정성을 고려하여 Q 라우팅을 기반으로 최적의 패킷 경로 탐색 기법을 제안한다. 제안된 기법은 기존의 다익스트라 기반 라우팅 기법에 비해 낮은 시간 복잡성 보여주기 때문에 새로운 서비스를 활성화할 때 발생하는 비용을 줄일 수 있다. 추후에 다수의 종단 간 서비스에 대해 최적의 경로를 탐색하는 기법을 연구하려한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음(IITP-2021-2016-0-00314*)

참 고 문 헌

[1] Marcus, J. Scott. "The economic impact of Internet traffic growth on network operators." Available at SSRN 2531782 (2014).

[2] Shu, Zhaogang, et al. "Traffic engineering in software-defined networking: Measurement and management." IEEE access 4 (2016): 3246-3256.

[3] J. Jiang, H. Huang, J. Liao and S. Chen, "Extending Dijkstra's shortest path algorithm for software defined networking," The 16th Asia-Pacific Network Operations and Management Symposium, 2014, pp. 1-4, doi: 10.1109/APNOMS.2014.6996609.

[4] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8.3-4 (1992): 279-292.

[5] Koenig, Sven, and Reid G. Simmons. "Complexity analysis of real-time reinforcement learning." AAAI. 1993.

[6] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.