

머신 러닝 기법을 활용한 SDN 트래픽 엔지니어링 기법 개선 방안

최준목*, 한윤선**, 현종환*, 홍원기*

포항공과대학교 *컴퓨터공학과, **정보전자융합공학부

{juk909090, seon054, noraki, jwkhong}@postech.ac.kr

요 약

IT 기술의 발전으로 인해 데이터 센터 트래픽이 급증하고 있다. 따라서 네트워크 트래픽을 관리하는 트래픽 엔지니어링의 중요성이 커지고 있다. 트래픽 엔지니어링은 트래픽을 분석하고 조절하여 전체 네트워크의 성능을 최적화하는 기법이다. 그러나 기존의 제어 평면과 데이터 평면이 고정 되어 있는 네트워크에 기반하는 트래픽 엔지니어링 기법들은 효율적이지 못 하다. 그 이유는 다중 경로를 효율적으로 활용하지 못 하거나 네트워크의 전역 상황을 활용하지 못 하며, 트래픽 간의 조율이 불가능하기 때문에 네트워크를 효율적으로 활용하지 못 하기 때문이다. 따라서 최근 화제가 되고 있는 Software Defined Networking (SDN) 기술을 활용하여 이러한 문제를 해결하고자 하는 트래픽 엔지니어링 기법들이 등장하였다. 본 논문에서는 SDN 을 활용 하는 트래픽 엔지니어링 기법들을 소개하고 이러한 기법들의 성능을 저하 시킬 수 있는 문제 점을 분석한다. 그리고 분석한 문제점을 머신 러닝 기법 알고리즘을 활용하여 개선 시킬 수 있는 방향을 제시하고자 한다.

1. 서론

오늘날 SNS 서비스의 발달, 동영상 스트리밍 기술, 클라우드 컴퓨팅 기술 등과 같은 급진적인 IT 기술의 발전으로 인해 네트워크 트래픽이 급증하고 있다. 이로 인해 네트워크의 트래픽을 효율적으로 분산시키거나 적합한 경로를 할당하여 관리하는 트래픽 엔지니어링의 중요성이 커지고 있다. 트래픽 엔지니어링은 트래픽을 분석하고 조절하여 전체 네트워크의 성능을 최적화하는 기법이다 [1]. 기존의 트래픽 엔지니어링 기법들은 다중 경로를 효율적으로 활용하지 못 하거나 네트워크의 전역 상황을 활용하지 못 하며, 트래픽 간의 조율이 불가능하기 때문에 네트워크를 효율적으로 활용하지 못 한다는 문제점을 가지고 있다.

이러한 문제점을 극복하기 위해 최근 Software Defined Networking (SDN) 기술을 활용하는 트래픽 엔지니어링 기법들이 제시되었다 [1][2]. SDN 은 네트워크의 제어 평면과 데이터 평면을 분리하고 제어 평면에 대한 유연한 제어를 가능하게 하여 네트워크를 효율적으로 관리 할 수 있게끔 해주는 기술이다. 이러한 기술의 도입으로 인해 SDN 에 기반한 트래픽 엔지니어링 기법들은 몇 가지 이점을 활용할 수 있게 되었다. 먼저 트래픽 엔지니어링에 활용할 수 있는 네트워크의 전역 상태를 집중형 (centralized) 컨트롤러에서 다룰 수 있으며 이를 통

해 네트워크를 최적의 상태로 빠르게 수렴시킬 수 있으며 높은 활용도 실현이 가능하다. 또한 플로우 테이블을 통해 유연한 트래픽 엔지니어링 구현이 가능하다.

본 논문에서는 SDN 을 활용하는 트래픽 엔지니어링 기법들을 분석하고 이러한 기법들의 성능을 저하 시킬 수 있는 문제점을 분석한다. 그리고 분석한 문제점을 머신 러닝 기법 알고리즘을 활용하여 개선 시킬 수 있는 방향을 제시하고자 한다.

2. 관련 연구

2.1. SDN 기반 트래픽 엔지니어링 기법

SDN 에 기반하는 트래픽 엔지니어링 기법들은 트래픽에 효율적인 경로와 대역폭을 할당하여 네트워크의 활용도나 고장 허용 한계(Fault tolerance) 성능을 향상시키고자 한다. 이를 위해 별도의 모니터링 서버를 사용하거나 스위치로부터 트래픽 정보를 수집하여 알고리즘의 효율성을 위해 정보를 집계한다. 그리고 할당 알고리즘을 통해 트래픽마다 적절한 경로와 대역폭을 할당한다.

MicroTE [3]는 현재 트래픽의 값과 트래픽의 평균 값의 차이가 일정한 범위 내에 있는 지 확인하여 트래픽의 예측 가능성을 판단하여 이를 트래픽 엔지니어링에 활용한다. 할당 알고리즘은 트래픽 통계 정보를 사용하여 예측 가능한 트래픽에 대해 선형 계획법(Linear programming)이나 Bin-packing heuristic 을 적용하여 먼저 경로를 할당한다. 그리고 예측 불가능한 트래픽에 대해서 이미 할당된 대역폭을 가중치로 반영하여 가중치가 존재하는 ECMP 알고리

이 논문은 2015 년도 정부(미래창조과학부)의 재원으로 정보통신 기술진흥센터의 지원을 받아 수행된 연구임 (No. B0190-15-2012, 글로벌 SDN/NFV 공개소프트웨어 핵심 모듈/기능 개발)

증을 적용한다.

Hedera [4]는 트래픽의 크기 변화에 따라 능동적으로 대응 할 수 없는 Equal Cost Multipath(ECMP)의 단점을 해결하기 위해 트래픽의 크기에 따라 서로 다른 트래픽 엔지니어링 기법을 사용한다. 트래픽의 크기가 작을 때에는 ECMP 방식을 사용하며 트래픽이 커질 경우 SDN 컨트롤러에서 적절한 경로를 할당해주는 방식이다. 특정 트래픽을 선택적으로 스케줄링하는 방식을 통해 Hedera 는 알고리즘으로 인한 부하를 최소화하면서도 네트워크 활용도를 높일 수 있다.

2.2 머신 러닝 기법을 활용한 트래픽 관련 연구

머신 러닝 기법을 통해 컴퓨터에게 자료의 특징을 스스로 분석하고 예측 할 수 있는 능력을 부여할 수 있다. 이러한 머신 러닝 기법에는 K-means clustering, Regression 등 다양한 기법들이 있으며 이를 통해 자료를 분류하거나 예측 할 수 있다. 따라서 트래픽 패턴 특징에 따라 트래픽을 분류하거나 트래픽의 크기를 예측하여 이를 트래픽 엔지니어링에 활용 할 수 있다. 이러한 연구에는 TCP 처리량 예측 [5], 트래픽 분류 [6]가 있다.

3. 본론

MicroTE 는 트래픽의 predictability 와 같은 간단한 트래픽 패턴을 활용하고자 하였지만 이는 트래픽 크기 변화가 거의 없는 패턴에 대해서만 의미가 있다. 또한 트래픽을 할당 알고리즘에서 사용하기 위해 서버 랙(Rack) 단위로 집적함으로써 많은 정보를 잃어버리게 된다는 단점을 가지고 있다. Hedera 또한 일정한 한계에 따라 스케줄링할 트래픽을 선택하기 때문에 트래픽의 패턴에 따라 효율적인 트래픽 엔지니어링이 불가능하다는 단점을 가지고 있다. 즉 개별적인 트래픽에 대한 특징을 최대한 활용하면서도 확장성을 위해 어느 정도의 집적이 필요하다는 문제와 트래픽 패턴에 대한 예측 문제를 해결하는 것이 이러한 트래픽 엔지니어링을 좀 더 개선 할 수 있는 방안이 될 것이다.

3.1 집적 수준을 위한 머신 러닝 기법

K-means 클러스터링 알고리즘은 Unsupervised learning 기법의 대표적인 방식으로 주어진 데이터들을 거리를 이용하여 임의의 K 개의 클러스터로 묶는 알고리즘이다. 따라서 트래픽의 크기, 크기 변화와 같은 트래픽 패턴에 관한 정보와 포트 번호, 목적지 IP 주소와 같이 어플리케이션과 관련된 정보를 벡터로 정의하여 이들의 유클리디안 거리를 사용하는 K-means 클러스터링 알고리즘을 생각해 볼 수 있다. 이러한 클러스터링 알고리즘을 사용하게 된다면 유사한 특징을 가지는 트래픽끼리 집적 할 수 있게 될 것이며 이를 효율적인 트래픽 엔지니어링 구현에 사용 할 수 있을 것이다.

3.2 트래픽 패턴 예측을 위한 머신 러닝 기법

Unsupervised learning 기법과 달리 Supervised learning 기법은 데이터로부터 어떠한 관계식을 유추 해내기 위한 기법이다. 따라서 특정 시간대나 특정한 서버에서의 트래픽을 예측하는 데에 이러한 기법이 사용될 수 있다. 대표적인 기법으로는 회귀분석이 있으며 예측된 트래픽에 따라 다른 트래픽 경로를 미리 변경 해두는 등의 활용이 가능할 것이다.

4. 결론

기존의 트래픽 엔지니어링 기법들은 트래픽의 경로 할당을 위한 최적화 문제를 풀 때, ToR(Top of Rack) 단위와 같은 형태로 트래픽 정보를 집적하여 사용하였다. 이러한 집적은 알고리즘의 계산 복잡도를 줄이는 데에 도움이 되지만 특정 트래픽이 가지는 트래픽 패턴 같은 정보를 잃어버리게 되는 단점이 있다. 따라서 트래픽 패턴에 대한 정보를 활용할 수 있되 가급적 그 수를 줄일 수 있는 집적 알고리즘이 필요하다. 이에 K-means 알고리즘과 같은 클러스터링 알고리즘을 적용하는 집적 알고리즘을 고안하고자 한다. 또한 머신 러닝 기법을 활용하여 트래픽의 미시적인 특성뿐만 아니라 거시적인 관점에서 트래픽을 예측 할 수 있다면 트래픽 변화에 효율적으로 적용하는 알고리즘을 고안 할 수 있을 것으로 기대된다.

5. 참고 문헌

- [1] I.F. Akyildiz, A. Lee, P. Wang, M. Luo, and W. Chou, "A roadmap for traffic engineering in SDN-OpenFlow networks," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 71, pp. 1-30, Oct. 2014.
- [2] 서신석, 리건, 현종환, 유재형, 홍원기, 백성복, 황찬규, 이영우, "소프트웨어 정의 네트워킹 기술을 활용한 데이터 센터 네트워크 트래픽 엔지니어링," *KNOM Review*, Vol. 16, No. 2, pp. 12-25, Dec. 2013.
- [3] T. Benson, A. Anand, A. Akella and M. Zhang, "MicroTE: Fine Grained Traffic Engineering for Data Centers," in *Proc. 7th CONference on emerging Networking EXperiments and Technologies(CoNEXT '11)*, 2010.
- [4] M. Al-Fares, S. Radhakrishnan, B. Raghaven, N. Huang and A. Vahdat, "Hedera: dynamic flow scheduling for data center networks," in *Proc. 7th USENIX Conference on Networked Systems Design and Implementation(NSDI '10)*, 2010.
- [5] M. Mirza, J. Sommers, P. Barford and X. Zhu, "A machine learning approach to TCP throughput prediction," in *Proc. ACM SIGMETRICS(2007)*, Vol. 35(1), pp. 97-108, June. 2007.
- [6] T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys & Tutorials*, Vol. 10(4), pp. 56-76, 2008.