

VNF의 자원 수요 예측 모델을 활용한 NFV 스케일링 시스템

최준묵*, 유재형†, 홍원기*

*포항공과대학교 컴퓨터공학과

† 포항공과대학교 정보통신대학원

{juk909090, styoo, jwkhong}@postech.ac.kr

NFV Scaling System using resource demand prediction model for VNF

Choi June Muk*, Jae Hyoung Yoo†, James Won-Ki Hong*

*Department of Computer Science and Engineering, POSTECH

† Graduate school of Information Technology, POSTECH

요약

본 논문에서는 사용자가 NFV 스케일링 과정에서 겪을 수 있는 서비스 품질 저하를 최소화 하기 위해 VNF의 자원 수요 변화에 따라 선제적으로 스케일링 할 수 있는 시스템을 제안한다. 본 시스템은 VNF의 자원 수요를 예측하기 위해 VNF의 자원 사용을 모니터링하는 모니터링 모듈, 이를 딥 러닝 알고리즘을 통해 VNF의 작업 특성과 다른 VNF와의 상관 관계를 분석하고 자원 수요를 예측하는 분석 모듈, 분석 모듈의 예측 결과를 통해 NFV의 스케일링을 수행하는 스케일링 모듈로 이루어져 있다. 이를 통해 NFV 스케일링 과정에서 사용자가 겪을 수 있는 서비스 품질 저하를 최소화하며 자원 할당을 최적화 하여 시스템의 자원을 효율적으로 활용할 수 있다.

I. 서론

최근 범용 하드웨어 성능이 발전함에 따라 하드웨어를 가상화 하여 제공하는 클라우드 컴퓨팅 환경이 보편화 되었다. 이러한 추세에 맞춰 네트워크 기능 또한 소프트웨어 형태로 가상화 하여 제공하고자 하는 수요가 발생하였다. 네트워크 기능 가상화(NFV: Network Function Virtualization)[1]은 전문 하드웨어에서 소프트웨어 기능과 하드웨어를 분리하여 제공하는 기술이다. 네트워크 관리자는 NFV를 통해 네트워크 투자 비용과 운용 비용을 절감 할 수 있으며 변화하는 네트워크 환경에 쉽게 대응 할 수 있다는 장점이 있다. IRTF(Internet Research Task Force)의 NFVRG(Network Function Virtualization Research Group)은 NFV 실시간 감시, 분석 기술, 스케일링 이슈 등 NFV에 대한 몇 가지 연구 이슈를 제시하고 있다. 그 중 스케일링 이슈는 VNF(Virtual Network Function)에 대한 서비스 수요를 충족시키기 위해 VNF에 할당된 자원이나 VNF의 수를 동적으로 변경하는 문제를 다룬다. 만약 VNF에 대한 서비스 수요가 VNF가 할당 받은 자원으로 충족시키기 어려운 경우 서비스 품질 보장을 위해 자원을 더 할당하거나 VNF 수를 늘리는 스케일 과정이 필요하게 된다.

본 논문에서는 VNF가 사용하고 있는 자원 상황을 모니터링하고 VNF가 필요로 할 자원을 예측 할 수 있는 알고리즘을 제안하며, 예측한 자원 수요를 충족시키기 위해 스케일링 작업을 미리 진행하여 사용자가 겪을 수 있는 서비스 품질 저하를 방지 할 수 있는 NFV 스케일링 시스템 구조를 제안하고자 한다.

II. 관련 연구

1. VNF 자원 수요 예측

VNF가 수행하는 작업은 CPU, 메모리, 디스크 등 다양한 자원을 요구한다. 이러한 자원의 상관 관계를 분석하여 VNF의 자원 수요를 예측하고자 하는 선행 연구 [2]가 있다. 이 연구에서는 CPU 사용률, 메모리 사용량, 네트워크 대역폭, 지연 시간 등 다양한 성능 지표와 함께 서비스 체이닝을 구성하고 있는 다른 VNF들의 성능 지표를 딥 러닝 알고리즘인 FNN(Feedforward Neural Network)와 GNN(Graph Neural Network)를 통해 분석하여 예측한다.

2. NFV Auto scaling

SRSA[3]는 scale-in 이나 scale-out 할 적절한 시간과 새로 추가하거나 삭제해야 할 인스턴스의 수를 결정하여 NFV 스케일링 이슈를 해결하였다. SRSA는 MDP(Markov Decision Process) 알고리즘을 사용하여 최적화 문제를 해결하고자 하였는데, 이를 위해 MDP의 상태(state)는 작업 부하를 나타내는 인덱스와 서비스 부하를 나타내는 인덱스 중 최대값, VM의 수, 시스템의 성능을 수치화 하여 정의된다. MDP의 행동(Action)은 추가하거나 삭제해야 할 인스턴스의 수를 사용하였다. 또한 MDP에서 가장 중요한 보상(Reward)는 VNF의 성능 보장과 전력 소모 최소화를 추구 할 수 있도록 두 지표를 수치화하여 합산한 값을 사용한다.

III. 본론

1. 개요

VNF에 할당된 자원이 VNF에 대한 서비스 수요를 충족시키지 못하면 사용자는 서비스 품질 저하를 경험하게 된다. 그러나 성능 문제가 발생하고 나서 스케일링 작업을 수행하게 되면 지연 시간으로 인해 서비스 품질을 효과적으로 보장 할 수 없게 된다.

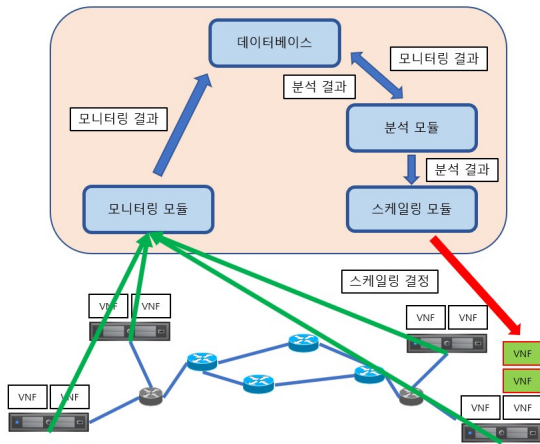


Figure 1. 구조도

일반적으로 컴퓨터에서 수행되는 작업은 CPU 의존 작업이나 디스크 I/O 의존 작업 등과 같이 작업 특성에 따라 사용하는 자원이 다르다. 이러한 점에 착안하여 본 논문에서는 VNF가 사용하는 자원을 모니터링하고 각 자원들의 변화와 상관 관계를 분석하여 VNF의 자원 수요를 예측하여 이를 통해 VNF를 선제적으로 스케일링하여 서비스 품질을 보장하고자 한다.

2. 구조

VNF의 자원 수요를 예측하기 위해서 먼저 VNF의 작업 특성을 파악해야 한다. 이를 위해 VNF가 배치되어 있는 VM의 자원을 모니터링 해야 하며 서비스 체이닝을 구성하고 있는 여러 VNF들 사이의 네트워크 자원도 모니터링 해야 한다. 이는 모니터링 모듈을 개발하여 각 VNF가 배치되어 있는 물리 머신 상의 OpenStack 모니터링을 통해 이루어지며 네트워크 자원 모니터링은 SDN 컨트롤러를 통해 이루어진다. 모니터링된 정보는 딥 러닝에 기반한 알고리즘을 통해 분석된다. 이를 통해 VNF의 작업 특성을 파악하고 현재 VNF의 자원 수요와 파악된 특성을 통해 단기간에 VNF가 요구할 자원 수요를 예측하여 이를 스케일링 모듈로 보낸다. 스케일링 모듈은 현재 VNF가 할당 받은 자원에 비해 예측된 자원 수요가 많다면 VNF를 추가로 할당하여 예상되는 수요를 맞출 수 있도록 한다.

3. 모니터링 모듈 및 데이터 베이스

모니터링 모듈은 각 물리 머신에 설치되어 있는 OpenStack Telemetry 서비스인 Ceilometer와 SDN 컨트롤러로부터 VM의 자원 정보와 네트워크 자원 정보를 수집한다. VM의 자원 정보에는 CPU 사용률, 메모리 사용량, 디스크 I/O 등이 있다. 네트워크 자원 정보에는 네트워크 지연 시간, 네트워크 대역폭 등이 있다. 수집된 정보는 시계열 데이터베이스(Time series database)를 통해 저장되며 데이터베이스는 분석 모듈로부터 요청을 받으면 이에 해당하는 데이터를 전달하는 역할을 수행한다. 데이터베이스는 모니터링 정보를 저장할 뿐만 아니라 분석 모듈의 분석 결과를 저장하여 차후 분석에 활용할 수 있도록 한다.

4. 분석 모듈

분석 모듈은 딥 러닝 알고리즘을 사용하여 VNF의 자원 수요를 예측하는 역할을 수행한다. 작업 특성에는 CPU 위주의 작업, 메모리 I/O 위주의 작업, 디스크 I/O 위주의 작업 등 VNF의 서비스에 따라 다양한 작업 특성이 존재할 수 있다. 따라서 이는 자율 학습

(Unsupervised Learning)을 통해 VNF의 작업 특성을 클러스터링(Clustering)하고 같은 클러스터에 속한 VNF들은 서로 비슷한 자원 수요를 보일 것으로 추정한다. 이러한 분석 결과를 데이터베이스에 저장하고 모니터링 결과와 대조하여 작업 특성에 따른 실제 자원 수요 변화를 학습하는 알고리즘을 구현한다. 알고리즘은 예측된 자원 수요를 일정 시간이 지난 뒤 측정된 실제 자원 수요와 비교하여 오차를 수정하여 학습하는 형태로 정확도를 높인다.

5. 스케일링 모듈

VNF가 수행하고 있는 작업 부하가 VNF에 할당된 성능을 넘어 일정 한계(Threshold)를 초과했을 때 스케일링 작업을 수행할 경우 실제로 불필요한 자원 할당으로 인해 비효율적으로 자원을 소모할 수 있다. 또한 VNF의 작업 부하가 한계점 근처에서 진동하는 경우 이를 효과적으로 처리하기가 어렵다는 단점이 있다. 본 논문에서는 분석 모듈을 통해 VNF의 자원 수요를 예측하고 이에 따라서 스케일링 작업을 수행한다. 이를 통해 정적인 한계점 방식이 가지는 문제점을 해결하고 VNF 자원 수요에 따라 능동적으로 대처할 수 있게 된다.

IV. 결론

본 논문에서는 사용자가 NFV 스케일링 과정에서 겪을 수 있는 서비스 품질 저하를 최소화 하기 위해 VNF의 자원 수요 변화에 따라 선제적으로 스케일링할 수 있는 시스템을 제안한다. 이를 위해 VNF가 사용하고 있는 자원을 모니터링하고 시간 변화에 따른 자원 변화와 함께, 네트워크 서비스를 함께 구성하고 있는 다른 VNF를 함께 고려하여 VNF의 작업 특성을 파악하고 이에 따른 자원 수요 변화를 예측하는 알고리즘을 사용하여 이를 통해 NFV의 스케일링을 결정할 수 있다.

ACKNOWLEDGMENT

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2015-0-00575, 글로벌 SDN/NFV 공개소프트웨어 핵심 모듈/기능 개발)

참고 문헌

- [1] R. Mijumbi, J. Serrat, J. L. Gorricho, N. Bouten, F. D. Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 236–262, Firstquarter 2016.
- [2] Mijumbi, Rashid, et al. "Topology-aware prediction of virtual network function resource requirements." *IEEE Transactions on Network and Service Management*, 2017, 14.1: 106-120.
- [3] P. Tang, F. Li, W. Zhou, W. Hu, and L. Yang. 2015. "Efficient Auto-Scaling Approach in the Telco Cloud Using Self-Learning Algorithm." In *2015 IEEE Global Communications Conference (GLOBECOM)*. 1–6.