

# 분산 클라우드 환경에서 서비스 트래픽 예측을 통한 에지 클라우드 활용 기법 연구

김도현<sup>o</sup>, 김윤곤, 강석원, 홍충선\*  
 경희대학교 컴퓨터공학과  
 { doma<sup>o</sup>, ykkim, dudntndud, cshong\* }@khu.ac.kr

## Edge Cloud Utilization Method through Service Traffic Prediction in Distributed Cloud Environment

Do Hyeon Kim<sup>o</sup>, Yunkon Kim, Seok Won Kang, Choong Seon Hong\*  
 Department of Computer Science and Engineering, Kyung Hee University

### 요 약

클라우드 컴퓨팅 환경은 실시간, 저 지연 등의 서비스 요구사항을 만족하기 위해 점차 분산화 형태로 변화하고 있다. 이에 따라 효율적인 클라우드 활용을 위한 네트워킹 기술로써, 각 클라우드 노드 간 서비스 배치 기술 또는 라우팅 기술이 분산 클라우드 환경에서 주요 기술로 활용될 수 있다. 본 논문에서는 분산 클라우드 환경에서 코어 클라우드 내 시계열 분석 모델인 Autoregressive Integrated Moving Average(ARIMA)를 활용한 트래픽 예측 시스템을 구성하고 예측 결과에 따른 서비스 배치 기법을 제안한다. 이는 실제 트래픽 데이터와 ARIMA 모델을 활용한 예측 시스템을 기반으로 특정 구간에 에지 클라우드로부터 서비스를 받을 수 있도록 미리 DNAT 테이블을 생성하여 서비스 인지 라우팅 모듈에 대한 지연을 줄임으로써 에지 클라우드를 보다 효율적으로 사용할 수 있도록 한다.

### 1. 서 론

네트워킹 관점에서의 분산 클라우드 모델은 네트워크 에지에 클라우드를 구성하여 사용자와 가까운 곳에서 실시간, 저 지연이 요구되는 서비스를 제공할 수 있는 효율적인 모델이다. 분산 클라우드와 같은 모델로써, Fog/Edge 컴퓨팅, MEC(Mobile Edge Computing)과 마이크로 데이터센터가 있으며 이를 활용한 다양한 연구가 진행되고 있다 [1]. 이러한 분산 클라우드 모델을 위하여, [2]에서는 Access Point 기능을 갖춘 프로토타입 상에 NAT(Network Address Translation) 기능을 기반으로 라우팅 모듈을 구현하였으며, 데이터베이스와 연동되어 등록되어 있는 서비스에 대한 정보를 토대로 에지 클라우드에서 서비스를 제공하도록 구현하였다. 하지만 라우팅 모듈의 경우, 패킷 검사 - 데이터베이스와 매칭 - 데이터베이스 정보 기반 DNAT PREROUTING 테이블 구성 - 패킷 전송과 같은 절차를 거쳐야 하기 때문에 딜레이가 발생하는 문제점이 있었다. 이에 본 논문에서는 라우팅 모듈 활용에 대한 딜레이를 줄이기 위해, 트래픽 예측 기법인 ARIMA를 기반으로 서비스 요청 패킷이 유입되기 전에, 미리 해당 서비스에 대한 DNAT 테이블을 생성하는 기법을 제안한다. 이는 코어 클라우드상에 배치된 ARIMA 트래픽 예측 시스템을 토대로 특정 서비스 요청에 대한 트래픽 패턴을 예측함으로써, 트래픽 양이

Threshold 값을 넘는 구간에서는 에지 클라우드에서 서비스를 제공할 수 있도록 하는 기법이다. 본 논문의 2장에서는 트래픽 예측 기법인 ARIMA에 대한 내용을 살펴보고, 3장에서는 본 논문에서 제안하는 시스템에 대한 내용과 서비스 제공 절차에 대해 설명한다. 성능평가의 경우 4장에서 기존 DNAT 라우팅 기법과 제안하는 시스템과의 RTT 결과값 비교를 통하여 지연시간에 대한 성능평가 내용을 다루며, 5장에서 본 논문을 정리하고 향후 연구방향을 제시한다.

### 2. 관련 연구

#### ◎ARIMA(Autoregressive Integrated Moving Average)

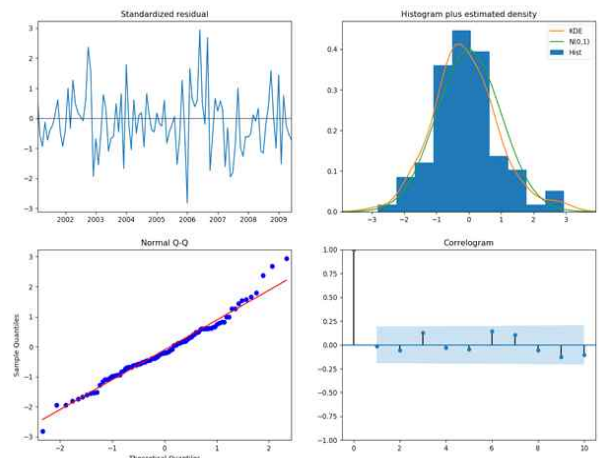


그림 1. ARIMA 모델 분석 결과 예시

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT연구센터육성 지원사업의 연구결과로 수행되었음(IITP-2018-2013-1-00717) \*Dr. CS Hong is the corresponding author

ARIMA는 과거의 데이터의 가중합(AR: 자기회귀)과 과거의 예측 오차의 가중합(MA: 이동평균)을 혼합해서 예측하는 기법으로써, 원칙적으로는 안정적인 데이터를 기반으로 활용되지만 데이터가 불안정(Non-stationary)할 경우 차분을 통해서 시계열 예측이 가능하다 [3].

일반적인 ARIMA(p, d, q) 모델에 의한 시계열 분석은 관찰된 시계열의 안정화 단계, 모형의 식별 단계를 거쳐 추정, 진단 그리고 예측 단계로 진행된다 [4]. 그림 1은 ARIMA 모델을 분석한 예시으로써, 하단 왼쪽의 그래프를 통하여 잔차의 규칙적인 분포가 N(0,1)을 사용하는 표준 정규 분포에서 취한 선형 추세를 따르고 있음을 알 수 있고 이는 잔차가 정상적으로 분포되어 있음을 나타낸다.

아울러 가장 최적의 p, d, q 값을 찾기 위해서는 AIC(Akaike's Information Criterion)값이 가장 적은 p, d, q 조합을 찾고 이를 적용한다. 본 논문에서 활용된 조합의 경우, (0, 2, 1) 값을 활용하였으며 각 조합 중 가장 적은 AIC 값을 갖는다. 이러한 ARIMA 모델을 통하여 특정 서비스에 대한 트래픽의 변화를 예측할 수 있다. 본 논문에서는 에지 클라우드를 효율적으로 활용하기 위해, 코어 클라우드 상에서 ARIMA 기법을 활용하여 미래의 트래픽 패턴을 예측함으로써 어느 시간대에 또는 어느 날짜 구간에 에지 클라우드를 통해 서비스를 제공할 것인지를 결정하게 된다.

### 3. 제안 사항

아래의 그림 2는 제안하는 시스템 모델을 나타낸 것으로써, 코어 클라우드 상에 ARIMA 모델 기반 예측 시스템을 배치하고 특정 서비스 트래픽에 대한 예측 결과를 통해 Threshold 값을 넘어가는 시간대에서는 에지 클라우드에서 DNAT 테이블을 생성할 수 있도록 테이블 Build 메시지를 전송함으로써 에지로부터 서비스를 제공받을 수 있도록 한다.

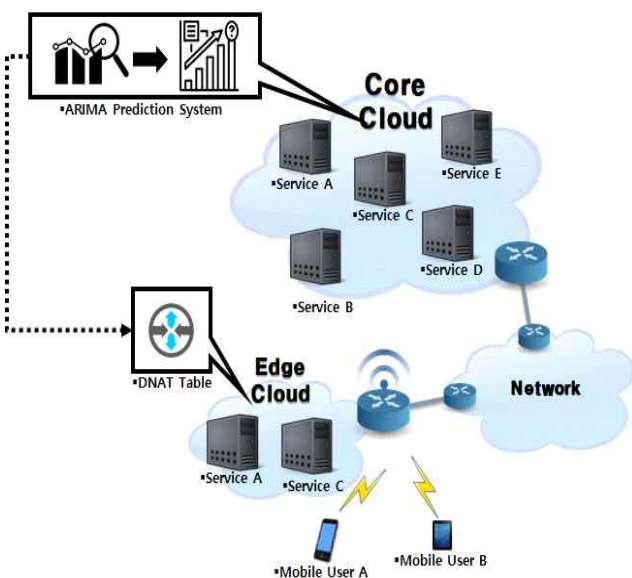


그림 2. 시스템 모델

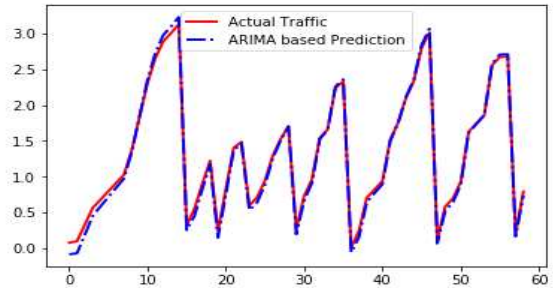


그림 3. 비디오 트래픽 예측 결과 (p, d, q)=(0, 2, 1)

그림 3은 Crowdad에서 제공하는 비디오 트래픽 데이터셋[5]을 활용하여 트래픽을 예측한 결과 그래프로써, x 축은 일(Day)을 나타내고 y 축은 트래픽 양(Gb)을 나타내며 RMSE (Root Mean Squared Error) 값의 경우 0.066이고 실제 트래픽과 유사한 예측값을 보여주고 있다.

$$Threshold = (TVol_{max} + TVol_{min}) \times 0.7 \quad (1)$$

수식 (1)은 코어 클라우드에서 예측되는 트래픽 양에 대한 Threshold 값으로 본 값을 상회하는 트래픽 양이 예측되는 구간에 대해서, 미리 에지 클라우드에 DNAT 테이블을 생성하도록 한다.

### 4. 성능 평가

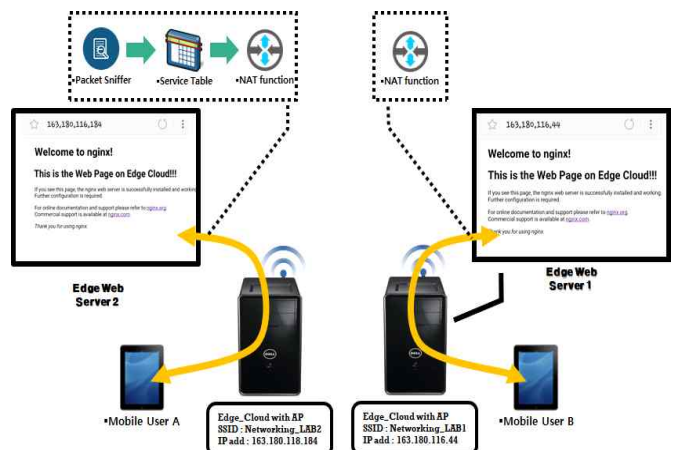


그림 4. 성능평가 환경 구성

그림 4는 성능평가를 위한 테스트 환경으로써, 코어 클라우드 상의 트래픽 예측 시스템 기반 DNAT 테이블을 생성하는 Pre-Build 방법과 기존에 제안했던 실시간 패킷 검사, 데이터베이스 매칭을 통하여 DNAT 테이블을 생성하는 Post-Build 방법 간의 패킷 전송에 대한 지연 시간 확인을 목표로 한다. 에지 클라우드의 경우 둘 모두 Ubuntu 14.04 기반의 Desktop 환경을 기반으로 구축하였으며, 무선 Access Point 환경을 구성하기 위해 Atheros AR9485 모듈을 탑재하였다. 이를 통해 모바일 기기가 에지 클라우드

드 노드에 접속이 가능하며 에지 클라우드를 거쳐 인터넷 또한 사용이 가능한 환경을 마련하였다.

Pre-Build 방식의 경우, 본 논문에서 제안하고자 하는 라우팅 과정을 축소시키는 방법에 대한 효율성을 확인해야 하기 때문에, 코어 클라우드로부터 Threshold 값을 넘는 구간에서 미리 DNAT 테이블을 에지 클라우드에 배치했다고 가정하였으며, Post-Build 방식의 경우 데이터베이스 상에 코어로 전송되는 서비스 트래픽이 있을 경우 116.184를 IP 주소로 갖는 에지 클라우드로 맵핑 할 수 있도록 정보를 등록하여 진행하였다.

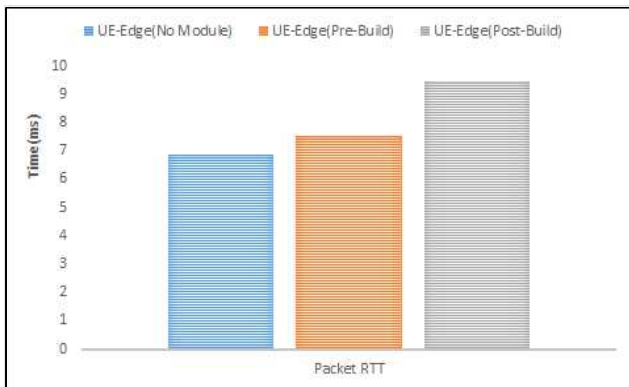


그림 5. DNAT 테이블 Build 방식에 따른 RTT 결과값

그림 5는 DNAT 테이블 생성 방법에 따른 RTT(Round Trip Time)을 측정된 결과로써 각 End-to-End에서 100번의 패킷을 보내고 그에 대한 평균 RTT 값을 통해 결과를 도출하였다. 트래픽 예측결과에 따른 Pre-Build 방식은 모듈을 활용하지 않았을 때 보다 DNAT 테이블 활용에 대한 지연시간이 추가되었지만, 기 제안하였던 Post-Build 방식에 비해 RTT 시간이 약 20.1% 감소하였다. 이는 분산 클라우드 환경에서 예측을 통한 Pre-Build 방식이 라우팅 모듈 활용에 대한 딜레이를 줄일 수 있음을 보여준다.

## 5. 결론 및 향후 연구

본 논문에서는 분산 클라우드 환경에서 에지 클라우드를 보다 효율적으로 활용하기 위하여 트래픽 예측 시스템 기반의 DNAT 테이블 Pre-Build 방식을 제안하였다. 이는 코어 클라우드와 에지 클라우드 간 Control 메시지를 전송하여 테이블 생성 및 삭제에 대한 기능을 수행하도록 함으로써 에지 클라우드의 라우팅 모듈 사용에 대한 지연시간을 줄일 수 있었다.

향후 연구로써 코어 클라우드 내 트래픽 예측 시스템과 관련하여 ARIMA 방식 외 다른 예측 기법과의 성능평가를 통한 최적의 예측 시스템 구축과 더불어 Threshold 값에 대한 최적화 연구를 진행할 계획이다.

## 6. 참고 문헌

- [1] Jing Zhang, Weiwei Xia, Feng Yan, Lianfeng Shen, "Joint Computation Offloading and Resource Allocation Optimization in Heterogeneous Networks With Mobile Edge Computing", IEEE Access, Vol.6, March. 2018
- [2] 김도현, 김윤곤, 변준영, 손동영, Tien-Dung Nguyen, 홍충선, "분산 클라우드 환경에서 에지 클라우드 활용을 위한 서비스 인지 라우팅 모듈 구현," 2017 Korea Software Congress(2017 KSC), 12월 20일 ~ 22일, 부산, 2017
- [3] youngji's blog, "ARIMA를 통한 월간 전력수요 예측", Online Available : <https://youngji-cho.github.io/2017/01/20/ARIMA를-통한-월간-전력수요-예측-연습하기.html>
- [4] 우경, 이성석, "Time Series Modeling for Forecasting Land Price Change Rate -Focusing on the Intervention ARIMA Model," 부동산학회, Vol. 60, 2015(in Korean)
- [5] Fraida Fund, Cong Wang, Yong Liu, Thanasis Korakis, Michael Zink, Shivendra Panwar, CRAWDDAD dataset nyupoly/video (v. 2014-05-09), traceset: nyupoly-dash, downloaded from <https://crawdad.org/nyupoly/video/20140509/nyupoly-dash>, <https://doi.org/10.15783/C7W30R>, May 2014