

엣지 클라우드 환경 잉여 컴퓨팅 자원의 활용을 위한 가용성 확보 방법 연구

김동완^o, 신용태^{*}

^o충실대학교 컴퓨터학과,

^{*}충실대학교 컴퓨터학부

e-mail: kk5339@soongsil.ac.kr^o, shin@ssu.ac.kr^{*}

A Study on the Availability of Surplus Computing Resources in Edge Cloud Environment

Dong-Wan Kim^o, Yong-Tae Shin^{*}

^oDept. of Computer Science, Soongsil University,

^{*}Dept. of Computer Science, Soongsil University

● 요약 ●

최근 빅데이터 및 인공지능의 중요성이 커짐에 따라 클라우드 시스템을 효율적으로 설계하고 관리하기 위한 연구가 활발히 진행 중이다. 본 논문은 기술 발전으로 각 개인은 고성능의 컴퓨팅 자원을 소유하고 있지만, 이 자원이 대부분 잉여 자원으로써 낭비되고 있다는 점을 착안하여, 잉여 컴퓨팅 자원을 효율적으로 활용하기 위해 엣지 클라우드 환경에서 분산된 자원의 가용성을 확보하기 위한 방법을 제안한다.

키워드: 엣지 클라우드(edge cloud), 분산 컴퓨팅(distributed computing),
시계열 분석(time-series analysis),

I. Introduction

창조경제연구회(Korea Creative Economy Research Network, KCERN)의 35차 포럼 보고서에 의하면 오늘날 빅데이터와 인공지능의 중요성이 커짐에 따라 클라우드의 역할이 매우 중요해졌다[1]. 빅데이터의 수집, 저장, 분석 그리고 인공지능 개발에 필요한 컴퓨팅 자원을 개별 기업이 별도로 확보하기에 어렵기 때문이다. 한국산업은행의 조사에 따르면, 컴퓨팅 자원을 확보해도 개별 서버의 자원 사용률은 평균 10~15%에 불과하다[2]. 고비용을 투자하여 확보한 하드웨어와 소프트웨어가 대부분의 시간을 유휴 상태로 대기하는 것은 IT Resource 낭비다. 때문에 유휴 상태인 잉여 컴퓨팅 자원을 효율적으로 활용하기 위해 등장한 것이 가상화(virtualization) 기술이다. 한편, 잉여 컴퓨팅 자원의 발생은 기업뿐만 아니라 각 개인에게서도 나타나는 추세이다. 기술의 발전으로 인해 대중에게 보급된 고성능의 스마트폰, 각 개인이 소유한 고사양의 데스크탑pc는 소유자의 사용시간 외에는 잉여 컴퓨팅 자원으로써 방치되는 것이다. 연구[3]에 의하면 클라우드 시스템을 효율적으로 설계하고 관리하기 위한 연구 및 논의가 활발히 진행 중이지만, 이는 확보한 IT Resource를 효율적으로 운영하기 위한 연구이다. 따라서 [4]의 연구는 클라우드 생태계의 전체적인 인프라 구축의 효율성을 위해 한정된 인프라 자원을 최대한 확보해야 함을 지적한다. 따라서 본 논문은 유휴 상태로 낭비되고 있는 잉여

컴퓨팅 자원을 효율적으로 활용하기 위한 방안을 연구하고 계층 구조를 지닌 엣지 컴퓨팅 구조를 제안한다. 클라우드가 제공하는 서비스는 SLA(Service Level Agreement)의 규약을 만족해야 한다. 따라서 클라우드 인프라를 구성하는 데이터스(이하 노드)는 24시간의 서비스 지속성과 데이터의 무결성을 보장할 필요가 있다. 때문에 본 논문은 노드의 가용성을 판단하기 위해 시계열 데이터 분석 기법인 ARIMA(Autoregressive Integrated Moving Average)을 통해 가용성을 평가한다. 또한 서비스의 연속성을 확보하기 위해 평가를 마친 노드들의 나열한 활동시간 24시간이 되도록 집합을 구성하고 예기치 못한 상황에 의한 서비스 장애에 대비하여 데이터의 복제본(replica)를 노드의 그룹별로 분산하여 저장해 무결성을 확보한다. 위 과정을 통해 제안하는 시스템은 제공자로부터 신뢰성있는 컴퓨팅 자원을 확보하며, 클라우드 이용자에게 안정적으로 서비스를 제공한다. 저장 장치 사용을 위한 클라우드 스토리지 서비스나 디버닝 등 대량의 연산이 필요한 작업을 효율적으로 처리하기 위해 엣지 클라우드 기반의 분산 컴퓨팅 서비스 제공이 가능해지는 것이다.

본 논문은 2장에서 관련 연구로 클라우드 스토리지 기술 분석, 모바일 엣지 컴퓨팅, 시계열 분석 기법에 대해 설명한다. 3장에서 잉여 컴퓨팅 자원을 활용한 엣지 클라우드의 가용성 확보 방법에

대해 자세히 기술하고 4장에서 결론 및 향후 연구를 제시한다.

II. Preliminaries

1. Related works

1.1 모바일 엣지 컴퓨팅 (Mobile Edge Computing: MEC)

클라우드 컴퓨팅, 모바일 컴퓨팅, 무선 네트워크를 결합해 모바일 사용자, 네트워크 운영자, 클라우드 컴퓨팅 제공 업체에게 풍부한 컴퓨팅 리소스를 제공하는 모바일 클라우드 컴퓨팅의 확장 개념이다. 5G 등 모바일 네트워크에서 IoT의 확산 및 신규 서비스 개발 지원을 위한 엣지 컴퓨팅이 주목된 개념이다. 클라우드(Cloud)-포그(Fog)-엣지(Edge) 계층으로 구성된다[5].

1.2 클라우드 스토리지 기술 분석

클라우드 스토리지 시스템은 대량의 데이터 저장 및 클라우드 컴퓨팅 처리의 핵심 구성요소로써 스토리지 미들웨어 시스템과 분산 파일시스템에 네트워크로 연결된 클러스터화로 수천대의 스토리지 장치들로 구성되어 사용자에게 클라우드 서비스를 제공한다. 리소스 자원, 네트워크 장치, 분산파일시스템을 포함하며 태블릿, 스마트폰 등 기기종 장치에서 데이터를 공유하며 쉽고 빠르게 네트워크를 통해 대량의 데이터를 전송하거나 접근한다[6].

1.3 ARIMA (Autoregressive Integrated Moving Average)

ARIMA는 시계열 분석 연구에서 중요한 기법이다. 자기 회귀 모형인 AR(Autoregression) 모형과 예측 오차를 이용하여 미래를 예측하는 MA(Moving Average) 모형을 합친 모형으로 데이터에서 복잡한 패턴을 식별하여 시계열에서 오는 값을 과거 값의 선형 조합으로 미래를 예측한다[7].

III. The Proposed Scheme

Fig. 1.은 제안 시스템 모델의 프로세스다. 제안 모델은 잉여 컴퓨팅 자원을 공유하고자 하는 제공자의 온 디맨드(On-Demand) 방식에 의해 플랫폼에서 디바이스를 가상화하여 엣지분산 클라우드 인프라로써 활용한다. 가상화된 노드는 데이터센터의 프로세스에 의해 ARIMA 분석을 거쳐 가용성을 진단하고 서비스 연속성을 보장하기 위한 노드의 집합을 구성한다. 그룹화된 노드는 데이터 무결성을 보장하기 위해 레플리카를 생성하여 인프라 구성을 마치고, 클라우드 서비스 이용자에게 적절한 서비스를 제공한다.

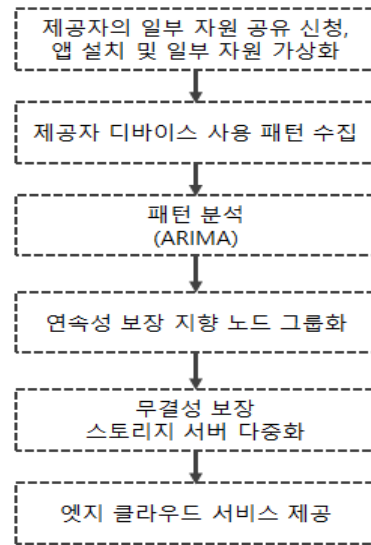


Fig. 1. The system's process for efficient utilization of surplus computing resources

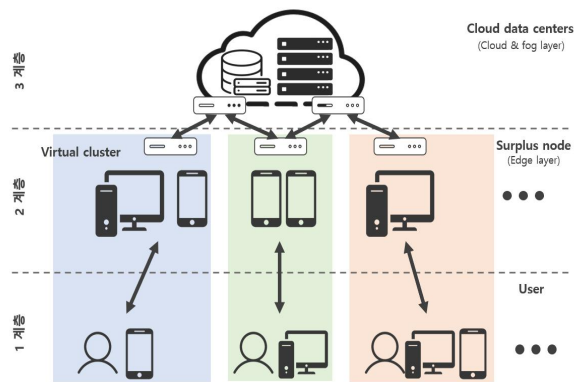


Fig. 2. Proposed System Edge Cloud Hierarchy Plot

Fig. 2.는 본 논문에서 제안하는 잉여 컴퓨팅 자원의 활용을 위한 엣지 클라우드의 계층 구조도다. 3 계층은 MEC 네트워크 구조에서 대량의 데이터를 저장 및 처리하는 데이터 센터를 소유한 클라우드 계층과 저장, 처리, 데이터 수집 및 전달과 같은 기능이 더 스마트한 장치로 구성된 포그 계층이다. 2 계층은 제공자에 의해 하이퍼바이저를 통해 가상화된 잉여 컴퓨팅 자원을 공유하는 노드들의 집합으로 이루어진 계층이다. 3 계층과 클라우드 서비스 이용자의 디바이스로 이루어진 1 계층과 통신을 하는 계층으로, 2계층의 노드들은 다음 표와 같은 데이터를 3계층에 전달하고 3계층은 데이터를 분석하여 가용성을 평가한다.

Table 1. Data Collection Items for Availability Diagnosis

Device	Battery	Network status	Storage	Computing power
D 1	x_1 %	y_1 Mbps	z_1 GB	r_1
D 2	x_2 %	y_2 Mbps	z_2 GB	r_2
...
D n	x_n %	y_n Mbps	z_n GB	r_n

개인PC 또는 모바일 디바이스의 배터리 잔량, 네트워크 상태, 컴퓨팅 자원의 상태 등은 소유주의 사용유무 즉, 사용자의 생활 행동 패턴과 밀접한 관련이 있다. 가령 모바일 디바이스의 배터리 잔량의 수치는 연속성을 띄며, 일상 및 수면 시간에 따라 일정한 패턴을 보이므로 과거의 값이 미래의 값에 영향을 끼치는 자기회귀 모형(AR)을 따른다. 또한 소유주의 디바이스 사용에 의한 배터리 소모 및 컴퓨팅 자원의 소모는 과거의 충격이 현재의 값에 영향을 주므로 이동 평균(MA) 모델을 따른다 볼 수 있다. 또한 충전 상태 유무에 따라 상승 또는 하락의 추세를 보인다(I). 따라서 표 1 항목을 시계열 데이터 분석을 통해 가상화된 잉여 컴퓨팅 자원을 미래 클라우드 이용자가 서비스를 요청했을 때 할당 가능 여부를 예상할 수 있다.

$$y_t' = c + \phi_1 y_{t-1}' + \dots + \phi_p y_{t-p}' + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad \text{①[7]}$$

수식①은 ARIMA(p,d,q)모델의 정의이며 는 차분을 구한 시계열이다. p는 AR부분의 차수, d는 차분의 정도, q는 MA부분의 차수를 나타낸다.

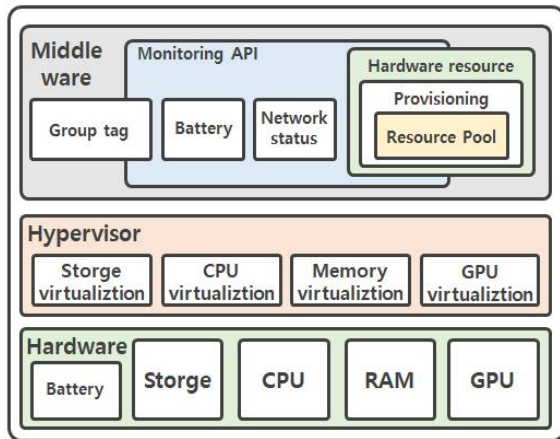


Fig. 3. Virtualization System Architecture

데이터의 수집, 분석과 가용성 진단 이후 가상화를 통해 잉여 컴퓨팅 자원을 공유하기 위해서는 각 디바이스를 하이퍼바이저(Hypervisor)를 실행할 시스템이 필요하다. 본 논문에서 제안하는 하드웨어 가상화와 분산 클라우드 구축을 위한 시스템의 구조는 다음 Fig. 3과 같다.

가용성 진단평가를 통과한 제공자의 디바이스는 하이퍼바이저를 통해 스토리지, CPU, 메모리, GPU 등의 컴퓨팅 자원을 가상화하여 클라우드 인프라로써 활용한다. 시스템의 마들웨어는 하드웨어 리소

스를 바탕으로 프로비저닝(Provisioning)된 리소스 풀(Resource pool)과 배터리, 네트워크 상태등을 모니터링하여 노드의 가용성을 실시간으로 점검하여 클라우드 서비스 이용자에게 자원을 할당한다. 또한 시스템에서 제공하는 가상화 기능을 통해 클라우드 인프라가 된 노드들은 서비스 연속성 및 무결성을 확보하기 위해 다음 예시와 같이 그룹화한다.

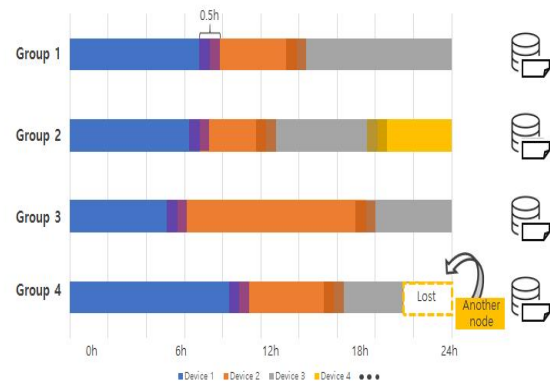


Fig. 4. Grouping nodes to ensure service continuity

본 논문에서는 가용성 및 무결성을 보장하기 위해 약식으로 n개의 노드로 구성된 4개의 그룹을 구성하였다. 차트의 길이는 ARIMA 분석에 따라 클라우드 이용자로부터 서비스 요청 시, 자원이 가능할 것으로 예상되는 노드의 가용 시간이며, 각 그룹은 예상 가용 시간의 집합이 24시간의 연속성을 갖도록 구성된다. 또한, 시스템에 의한 실시간 모니터링 결과 예기치 못한 상황으로 엷지 클라우드의 일부 노드가 작동하지 않을 경우, 이를 대체할 수 있는 다른 노드를 기존 그룹에 추가하여 가용성을 확보한다. 시스템의 마들웨어는 그룹 태그를 통해 노드가 속한 그룹을 구분하고, 클라우드 서비스 이용자가 자신의 데이터에 접근하기 위한 식별자로써 작용한다. 4개의 그룹은 DB서버 다중화 구현의 일환으로 데이터의 무결성을 위해 데이터의 복제본을 저장한다.

IV. Conclusions

본 논문은 시계열 데이터 분석을 통해 사용 가능한 잉여 컴퓨팅 자원 확보하고 확보한 자원을 그룹화하여 안정적으로 이용자에게 서비스를 제공하는 방법을 제시하였다. 분산 클라우드의 일종인 엷지 클라우드 구현에 있어 데이터의 가밀성 문제가 제기되나, 잉여 컴퓨팅 자원의 공유를 플랫폼을 거치지 않고 유저간 거래 모델을 구현하기 위해 프라이빗 블록체인을 도입하으로써 인증되지 않은 사용자의 접근을 막거나, 컨테이너 가상화 기반의 컨피덴셜 컴퓨팅을 구현하으로써 자원 제공자의 접근을 막는 등의 보안 기법을 고려할 수 있을 것이다. 추후 연구에서는 ARIMA 분석에 의한 가용성 판단 시스템의 실제 성능을 검증해 볼 것이다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심 대학사업의 연구결과로 수행되었음(2018-0-00209)

REFERENCES

- [1] Minhwa Lee, "Prerequisites for the Fourth Industrial Revolution Cloud Data Innovation Strategy," KCERN Forum Report, Vol. 35, Apr. 2017
- [2] Maengsoo Kang, "Cloud Computing Market Trends and Prospects," KDB Monthly, Vol. 758, No. 1, Jan. 2019.
- [3] Kyung Choi, Mihui Kim, "Research on Convergence of Internet-of-Things and Cloud Computing", The Journal of the Korea Contents Association, Vol. 16, No. 5, May 2016.
- [4] Yongho Kang, Changbok Jang, Wanjik Lee, Seokyeol Heo, Jooman Kim, "Efficient Method to Support Mobile Virtualization-based Cloud Resource Management", Journal of Digital Convergence, Vol. 12, No. 2, pp.273-283, 2014.
- [5] Hoon Kim, Kanperbez, Dae Hyeong Kim, "Mobile Edge Computing: Candidate Technology for 5G Transmission Issue Resolution", The Proceedings of the Korean Institute of Electromagnetic Engineering and Science, Vol. 28, No. 5, pp.10-16, Sep 2019
- [6] JeongSu Park, YuMi Bae, SungJae Jung, "Technical analysis of Cloud Storage for Cloud Computing", Journal of the Korea Institute of Information and Communication Engineering, Vol. 17, No. 5, pp. 1129-1137, May 2013
- [7] Yapeng Zhou, Miaohua Huang, "Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model," Microeletrincs Reliability Vol. 65, pp. 265-273, 2016