

지능형 에지 컴퓨팅 및 네트워킹 기술

Technologies of Intelligent Edge Computing and Networking

홍승우 (S.W. Hong, swhong@etri.re.kr)	네트워크연구본부 책임연구원
이창식 (C.S. Lee, cslee2624@etri.re.kr)	네트워크연구본부 연구원
김선철 (S.C. Kim, sckim007@etri.re.kr)	네트워크연구본부 책임연구원
강경순 (K.S. Kang, kskang@etri.re.kr)	네트워크연구본부 책임연구원
문 성 (S. Moon, seong@etri.re.kr)	네트워크연구본부 책임연구원
심재찬 (J.C. Shim, jcshim@etri.re.kr)	네트워크연구본부 책임연구원
홍성백 (S.B. Hong, sbhong@etri.re.kr)	네트워크연구본부 책임연구원
류호용 (H.Y. Ryu, hyryu@etri.re.kr)	네트워크연구본부 책임연구원/PL

In the upcoming post-app era, real-time, intelligent and immersive services such as autonomous vehicles, virtual secretaries, virtual reality, and augmented reality are expected to dominate. However, there is a growing demand for new networking and computing infrastructure capabilities because existing physical connection-oriented networks and centralized cloud-based service environments have inherent limitations to effectively accommodate these services. To this end, research on intelligent edge network computing technology is underway to analyze the contextual situation of human and things and to configure the service environment on the network edge so that the application services can be performed optimally. In this article, we describe the technology issues for edge network intelligence and introduce related research trends.

* DOI: 10.22648/ETRI.2019.J.340103

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송 연구개발 사업의 일환으로 수행하였음[2018-0-01502, 인공지능 기반 지능형 에지 네트워킹 기술개발].



본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

최신 네트워크 기술 동향 특집

- I. 서론
- II. 에지 네트워크 컴퓨팅
기술 동향
- III. 지능형 에지 컴퓨팅 및
네트워킹 기술
- IV. 결론

1. 서론

최근 가트너(Gartner)에서는 3가지 핵심 키워드, 지능(Intelligent), 디지털(Digital), 메시(Mesh)와 함께 각 키워드에 포함되는 2018년 10대 전략 기술을 발표하였다. 이중 메시(Mesh) 키워드는 메시-앱 및 서비스 아키텍처를 포함하는데, 향후 수년 이내에 기존의 웹 및 모바일 기반의 어플리케이션에서 가상비서, 증강현실, 상황인지 기반 동적 서비스 등의 포스트-앱 형태로 전환이 이루어지며, 이러한 서비스를 수용하기 위한 서비스 구조 및 인프라 개발을 강조하고 있다. 새로운 인프라 측면에서 의미 있는 전략 기술이 디지털(Digital) 카테고리에 포함되어 있는 ‘클라우드에서 에지(Cloud to Edge)’로 표현된 에지 컴퓨팅 기술이다[1]-[5].

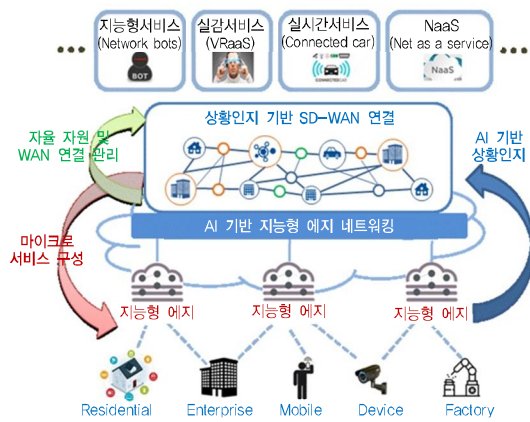
에지 컴퓨팅은 정보 처리와 콘텐츠 수집 및 전달이 정보 소스(Source)와 인접한 네트워크 에지에서 처리되는 컴퓨팅 방식을 의미하는데, IDC가 정의한 바에 따르면 중요한 데이터를 지역에서 처리하거나 저장하고, 정제되고 정보화된 데이터를 중앙 데이터 센터나 클라우드로 보내는 약 10평방미터 이하 규모의 마이크로 데이터 센터들로 구성된 메시 네트워크(Mesh Network)로 표현하고 있다[6].

에지 컴퓨팅 기술이 차세대 인프라 기술로 주목받고 있는 이유는 사람 및 사물과 근접한 네트워크 에지에 컴퓨팅 자원을 분산 배치함으로써 기존의 중앙 집중 클라우드 기반 서비스 인프라의 단점을 보완할 수 있기 때문이다. 우선, 사물인터넷의 확산과 스마트기기의 폭발적인 증가에 따른 데이터 처리와 네트워크 대역폭에 대한 부담을 줄여줄 수 있다. 씨게이트(Seagate) 자료에 따르면 2025년에는 전 세계 데이터 총량이 현재보다 10배 늘어난 163 제타바이트(ZB)에 달할 것으로 예측하고 있다[7]. 수많은 단말 및 경로로 수집된 방대한 데이터들을 중앙 집중 클라우드에서 처리할 경우 서비스 처리 시간의 증가뿐만 아니라 이를 전달하기 위한 네트워크 자

원을 빠르게 고갈시키는 문제를 초래하게 될 것이다. 또한, 에지 컴퓨팅은 사용자 서비스를 근접한 곳에서 처리하여 5G의 핵심 서비스로 분류되는 자율주행차, 촉각인터넷(Tactile Internet) 등 실시간 상호 작용이 가능한 저지연 서비스와 함께 가상현실(Virtual Reality), 증강현실(Augmented Reality)과 같은 클라우드 기반 실감형 광대역 서비스를 위한 인프라로 사용될 수 있다. 뿐만 아니라, 에지 컴퓨팅에 프라이버시 및 보안 정책 기능을 부여하여 자체적으로 높은 보안 수준을 제공하거나 중앙 클라우드에 문제발생시 독립적으로 서비스를 제공할 수 있는 가용성을 제공하는 등 다양한 문제를 해결할 수 있는 기반 기술로 사용될 수 있다.

이와 같은 에지 컴퓨팅 기술이 가지는 장점들로 최근 여러 표준화 단체에서 기존의 클라우드 개념을 확장하여 에지 컴퓨팅을 수용하는 움직임을 보이고 산업계에서도 많은 벤더들이 자체적인 솔루션을 개발하고 있다. 하지만, 아직은 단순히 위치적 측면에서 일부 기능이 구현되거나 표준화 작업이 진행되고 있을 뿐이며, 에지 컴퓨팅 인프라의 장점을 제대로 발현하기 위한 지능적인 개체로서 연구해야 할 부분들이 많이 남아 있다.

개념적으로 에지 컴퓨팅 노드는 사람 및 사물과 중앙 클라우드 사이의 위치에 존재하여 서비스를 제공하기 위한 저장, 컴퓨팅, 네트워킹 자원을 가진 개체이다. 단순히 자원을 제공하는 수동적인 개체가 아니라 능동적인 지능을 가지고 연결(Connect), 컴퓨팅(Compute), 제어(Control)를 수행하는 지능적인 개체로서 역할을 수행하여야 한다. 사람 및 사물과 근접한 위치에서 사람 및 사물들에 대하여 끊임 없는 상호 간의 연결과 함께 분산 배치된 수평적 에지 간의 연결을 고려하여야 한다(연결). 수많은 사람 및 사물에서 생성되는 데이터를 수집, 분류할 뿐만 아니라 자체적으로 분석하여 지능을 형성하여야 한다(컴퓨팅). 또한, 얻어진 지능을 통하여 사용자에게 적절한 서비스를 제공하고, 최적의 서비스가 유지될 수 있도록 자원의 최적 배치 및 제어가 이루어져야



(그림 1) 지능형 에지 네트워크 개념도

한다(제어). 이를 위하여 (그림 1)과 같이 사람 및 사람과 근접한 네트워크 에지에서 분산된 컴퓨팅 인프라들이 상호협력하여 상황을 분석하고, 응용 서비스들이 최적으로 수행될 수 있도록 한정된 컴퓨팅, 네트워크 자원을 최적으로 관리하는 지능형 분산 에지 네트워크 컴퓨팅 기술 연구가 진행되고 있다.

본고에서는 네트워크 에지 인프라의 지능화와 관련된 연구 기술들과 최신 연구 개발 동향에 대하여 분석한다.

II. 에지 네트워크 컴퓨팅 기술 동향

ETSI(European Telecommunications Standards Institute)는 5G 이동통신 시스템에 에지 컴퓨팅 기능을 추가하는 MEC(Mobile Edge Computing)와 관련된 표준화를 진행하였으며, 후속으로 Multi-access Edge Computing을 위한 문서를 준비 중이다. MEC 표준의 특징은 산업체의 서비스 시나리오와 PoC(Proof of Concept)를 병행하여 진행하는데 MEC-ISG(Industry Specification Group)로 시작되는 표준 문서가 산업규격 그룹에서 만든 문서이다. ETSI 표준들은 NFV와 클라우드 컴퓨팅의 결합 형태로 작업이 진행되고 있다. MEC 표준은 MANO(Management and Orchestration) 구조의 NFV(Network Function Visualization)와 유사한 형태이지만, NFV 인프라에 모바일 중심의 응용 플랫폼

또는 미들웨어가 결합된 구조를 사용하고 있다.

ITU-T는 에지 컴퓨팅에 인공지능 지원을 목표로 기계학습(ML: Machine Learning) 및 고급 네트워킹 기능을 갖춘 IEC(Intelligent Edge Computing)를 작업 중에 있다. 이는 정보 기술(IT) 및 운영 기술(OT) 산업이 네트워크 에지에 집중되면서 실시간 네트워크, 사이버 보안을 보장하는 보안 기능, 자체 학습 솔루션 및 개인화된 연결성 기술이 중요해진다. 에지 컴퓨팅의 핵심인 컨테이너 기술과 ML 기술은 새로운 비즈니스 및 시장의 요구에 대한 선도적인 기술로 볼 수 있다. 그러나 이들 기술은 Edge Intelligence에서는 중요하지만, 표준으로는 존재하지 않고 Linux Foundation 주도의 오픈 소스 진영에서 먼저 개발되어 적용되고 있다.

포그 컴퓨팅(Fog Computing) 기술은 (주)시스코에 의해 소개된 기술로 초기에는 클라우드 서버와 네트워크 에지 디바이스들 사이의 연산, 저장 및 네트워킹 서비스를 제공하기 위한 클라우드 컴퓨팅의 확장 기능으로 소개되었다. 포그 컴퓨팅 기기는 저사양 기기에서부터 고사양의 클라우드 디바이스까지 다양하게 사용할 수 있다. 포그 컴퓨팅 기술은 저지연 서비스에 사용할 수 있는 장점을 가지며 증강 현실(AR) 및 실시간 비디오 분석, 사이버 물리적 시스템, 공유 주차 시스템, 다양한 차량간 통신 서비스 등에 적용 가능하다. 2015년 (주)시스코는 마이크로소프트, 인텔, ARM, Dell, AT&T 등이 참여하는 OpenFog 컨소시엄을 발표하였고 현재 협력사를 중심으로 IoT 서비스를 위한 분산 클라우드 컴퓨팅 환경 프로젝트를 수행 중이다.

또한, 카네기멜론 대학을 중심으로 하는 Open Edge Computing Initiative가 있다. 2009년 카네기멜론 대학은 미 국방성으로부터 국방 혁신에 대한 연구 용역을 받아 'Cloud Computing at the Tactical Environment', 'Advanced Mobile Systems' 등의 프로젝트를 수행하면서 Cloudlet이라는 소형 데이터센터 기술을 구현하였다. 이후 2016년 6월 Cloudlet 플랫폼을 기반으로 서비스

활성화를 위하여 Open Edge Computing Initiative를 구성하였다. 이 Initiative에는 화웨이, 인텔, 노키아, 보다폰 등의 산업체와 NTT, 도이치텔레콤, 버라이즌 등의 통신사업자가 참여하였으나 최근에는 활동이 정체된 모습을 보이고 있다.

에지의 지능화를 위해 선도 기업들이 데이터 수집, 분석 및 추론을 위한 에지 플랫폼 기술을 선보이고 있다. 마이크로소프트는 Azure IoT Edge를 개발하였는데 클라우드에서 데이터의 고급 분석, 기계 학습 및 인공지능 모델을 학습한 후 IoT Edge에 컨테이너를 활용하여 배포하고, Edge에서는 인공지능을 활용한 고급 데이터 분석을 수행한다. 서비스로는 Azure 머신 러닝, Azure 스트림 데이터 분석, 클라우드 기반 기기 관리 등이 있다. IoT Edge는 컨테이너 기반의 모듈과 컨테이너의 설치, 실행, 업데이트를 위한 런타임 모듈 그리고 기기 관리 및 모니터링을 위한 클라우드 인터페이스로 구성된다.

아마존은 AWS Greengrass 솔루션을 보유하고 있으며 AWS 클라우드 기능을 로컬 디바이스로 확장하여 로컬 컴퓨팅, 메시징, 데이터 캐싱, 동기화 및 인공지능 추론 기능을 구현하였고 AWS Lambda와 AWS IoT를 사용하여 AWS 클라우드 및 로컬 디바이스에서 애플리케이션을 수행한다. 아마존은 딥러닝을 지원하는 개발자용 비디오 카메라인 AWS DeepLens도 개발하였다.

구글은 2018년 7월에 자사 Cloud IoT Core™ 에 연동하는 에지 플랫폼의 알파 버전을 제공하고 있다. 이 플랫폼은 Android 및 Linux를 지원하며 TensorFlow Lite 모델을 로컬 에지에서 동작(학습은 Cloud에서)하도록 개발한 Google Kubernetes Engine(GKE)을 지원하고 있으며, 기계학습에 특화된 자사 MLU인 Edge TPU™ 과 공동 프로모션을 추진 중이다.

이외에 Linux Foundation이 주도하는 벤더 중립 오픈 소스 프로젝트인 Edge X Foundry가 있으며 이는 IoT 에지 컴퓨팅을 위한 개방형 공동 프레임워크다. Edge X는 Plug-and-play 컴포넌트 생태계를 가지며

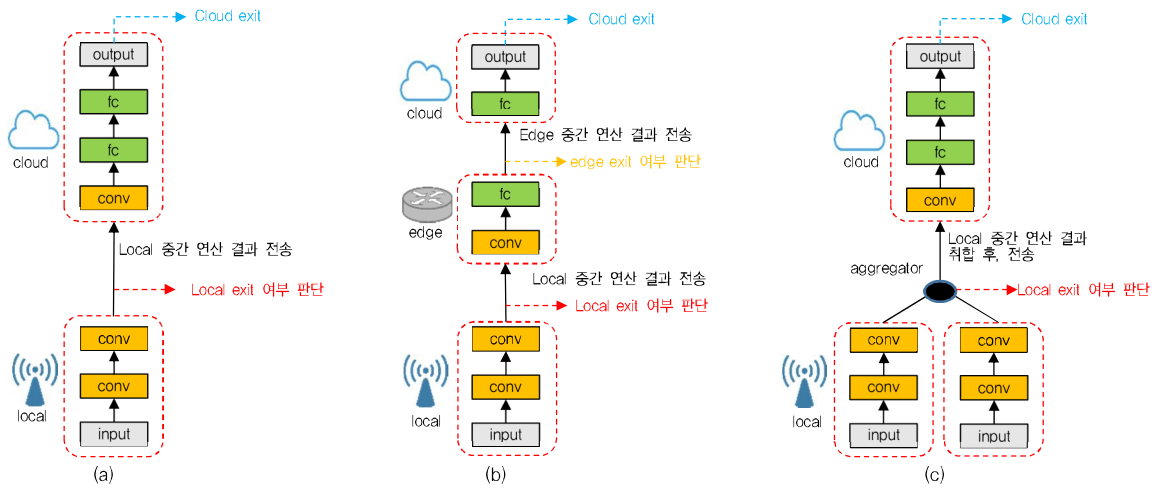
IoT 솔루션 개발을 가속화하고 시장을 통합할 수 있는 OS 중립적인 레퍼런스 소프트웨어 플랫폼이다. 이 플랫폼은 다양한 실리콘(예, x86, ARM)과 운영체제(예, Linux, Windows, Mac OS) 그리고 응용 환경(예, Java, JavaScript, Python, Go Lang, C/C++)을 지원하며 2017년 4월에 시작하여 약 6개월마다 오픈 소스를 Release하고 있다. 2018년 6월 California 버전 이후 2018년 10월 말 Delhi 버전을 Release 하였다.

III. 지능형 에지 컴퓨팅 및 네트워킹 기술

1. 분산 인공 지능 처리 기술

DNN(Deep Neural Networks)은 최근 주목받고 있는 ML 기법 중 하나이며, 매우 높은 정확도를 특징으로 하여 이미지 분류, 음성 인식, 자연어 처리 등의 많은 어플리케이션에 활용되고 있다. DNN은 입력 Layer와 출력 Layer, 그리고 그사이에 존재하는 여러 개의 Hidden Layer로 이루어져 있으며, 최종 결과를 도출해 내기 위해 수많은 Layer에서 복잡한 연산 과정을 거친다. 이러한 연산 과정에서 많은 계산량과 저장공간, 에너지 소모를 필요로 하며, 이들 요구사항은 제한된 컴퓨팅 자원과 저장 공간을 가지고 빠른 서비스를 제공해야 하는 에지 컴퓨팅 환경에서 해결해야 할 하나의 큰 과제이다.

이러한 문제를 해결하기 위해서는, (그림 2)와 같이 클라우드와 에지 노드 간에 지능적 협력을 통한 분산 처리가 가능해야 한다[8]. 우선적으로, 에지 노드에서 DNN의 초반 Layer 연산 과정을 거치다가, 특정 Layer에 도달했을 때 중간 처리 결과를 기반으로 정확도를 계산한다. 계산된 정확도는 미리 정해진 임계치와 비교를 하여 정확도가 임계치보다 높을 경우, 처리된 결과를 출력하고 이후 Layer 연산 단계를 멈춘다(Local Exit). 만약, 정확도가 임계치보다 낮을 경우, 해당 Layer까지 처리된 중간 결과를 클라우드로 전송하고 클라우드에서는 이후 Layer부터 다시 연산을 진행하여 최종 Layer까지



(그림 2) DNN 협력적 분산 처리 구조

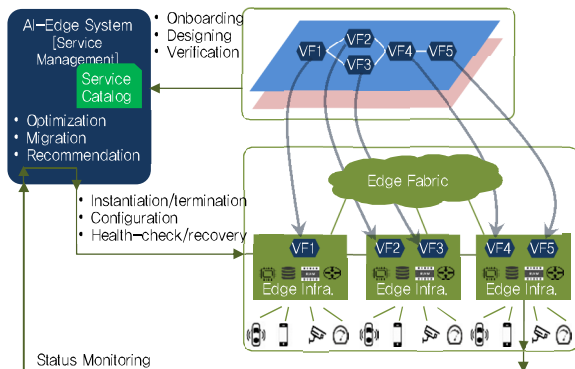
처리된 결과를 출력한다(Cloud Exit). 이러한 Exit 구조는 Local Exit, Cloud Exit과 같이 2단계 Exit 구조뿐만 아니라, 경우에 따라서 3단계(Local, Edge, Cloud) 또는 그 이상의 Exit 구조로 확장될 수도 있다. 또한, 에지 기기가 여러 개 존재할 경우에는, 중간의 Aggregator를 두어 각 에지 노드에서 처리된 결과를 취합하고 Aggregator에서 Exit 여부를 판단할 수 있다.

분산 Exit 구조와 더불어, 서비스의 End-to-End 지연 또는 에지 노드의 에너지 소모를 최소화하기 위해서는 DNN의 어느 Layer를 기준으로 에지 노드와 클라우드가 나누어 연산할 것인지에 대한 방안이 필요하다[9], [10]. 이러한 Layer Partitioning 문제를 해결하기 위해서는, 우선적으로 특정 DNN 구조에서 각 Layer 들의 Computational Latency와 Output Data Size에 대한 특성을 파악하여야 한다. 이러한 Layer 별 특성은 DNN 아키텍처에 따라 달라지며, 각 Layer마다 Input Size, Filter 개수, Filter 크기 등의 다양한 Configuration에 대해 Profiling하여, Layer 별로 Computational Latency 또는 에너지 소모 예측 모델을 만들어야 한다. 예측 모델은 설계단계에 만들어지며, Profiling한 데이터를 바탕으로 Regression 모델을 사용하여 만든다. 이후, 실제 운용단계에서는 만들어 놓은 Layer 별 성능 예

측 모델과 함께, 현재 에지 노드의 자원 사용 상황, Bandwidth, Application Load 등을 고려하여 최적화하고자 하는 Metric(Latency 또는 Energy Consumption)에 맞는 Best Partition Point를 찾아낸다. 찾아낸 Partition Point를 기준으로 하여, 에지 노드에서는 앞쪽 Layer에 대한 연산 과정을 거치고 중간 계산 결과를 클라우드로 전송한다. 이를 수신한 클라우드에서는 Output Layer까지 연산 과정을 거친 후, 최종 결과를 유저에게 전송하는 방식으로 동작한다.

2. 가상화 기반 마이크로 서비스 기술

현재의 서비스들은 자원 사용의 효율화 및 비용 절감을 위해서 COTS(Commercial Off the Shelf) 하드웨어 상에 가상화된 형태로 동작하는 것이 일반적이다. 즉, 에지 인프라를 구성하고 있는 에지 노드에는 COTS 하드웨어들이 배치되고 실제로 서비스를 수행하는 단위 기능들은 Virtual Machine이나 Container 형태로 가상화되어서 이들 물리 인프라 상에서 동작하고, 이들 기능이 서로 네트워킹을 하기 위해서 SDN(Software Defined Network) 등으로 제어가 된다. 또한, 사용자들이 느끼고 경험하는 특정 서비스는 실제로 하나 이상의 마이크로 서비스로 구성되어서 동작한다. 마이크로 서



(그림 3) 분산 에지 마이크로 서비스 관리 구조

비스란 서비스를 구성하는 최소 기능 단위로써 이들 기능이 서로 밀접한 관계를 가지지 않도록 설계되어서 빠른 배포 및 수정이 가능하도록 서비스를 구성하는 것을 의미한다. 이들 마이크로 서비스는 실제로 물리 인프라 상에서 가상화된 VF(Virtual Function)로 동작한다(그림 3) 참조].

특정 서비스를 에지 인프라 상에서 동작시키는 것은, 그 서비스와 서비스를 구성하는 VF들의 요구사항을 만족하도록 인프라 상에 분산 배치하는 것을 의미한다. 궁극적으로 서비스를 에지 인프라에 분산 배치해서 수행시키기 위해서는 먼저 서비스를 설계하는 단계와 실제로 서비스를 운영하는 단계가 필요하다. 서비스를 설계하는 단계에서는 서비스를 구성할 VF들을 등록하는 단계, VF들이 인프라 상에서 올바르게 수행될 수 있는가를 검증하는 단계, 등록된 VF들을 이용해서 하나의 서비스를 제작하는 단계, 제작된 서비스를 검증하는 단계, 서비스에 적용할 정책을 정의하는 단계와 정의된 정책을 검증하는 단계 등으로 구성된다. 서비스 설계 단계를 통해서 등록된 서비스들은 자원의 요구사항과 정책 요구사항을 기준으로 에지 인프라 상에 배치되어서 실행된다. 이 단계가 서비스를 운영하는 단계로서 서비스 시작 및 종료 제어, 자원들의 형상 제어 기능을 수행한다. 또한 실시간으로 서비스를 구성하는 기능들이 정상적으로 수행되는지를 모니터링해서 장애 발생 시 복구하는

기능을 수행한다.

여러 서비스가 에지 인프라에서 동작하고 있는 동안에도 실시간 최적화 단계가 필요하다. 최적화 단계에서는 서비스와 정책의 요구사항과 자원의 상황을 기준으로 서비스들의 배치를 결정하는 단계이다. 서비스를 구성하는 VF들은 요구사항과 자원의 상황에 따라서 에지 인프라에 배치될 수도 있고, 상위 클라우드에 배치될 수 있다. 이들 VF의 배치는 동적 마이그레이션을 통해서 상황에 따라서 변경된다. 또한, 새로운 서비스는 관리자의 명시적인 요청에 의해서 에지 인프라에 배치될 수도 있고, 사용자나 주변 이벤트의 상황을 기준으로 자동으로 새로운 서비스가 에지 인프라에 배치될 수도 있다. 이런 자동 서비스 추천 기능은 다양한 상황 정보를 활용하는 AI/ML(Artificial Intelligence/Machine Learning) 기술을 사용해서 처리된다.

위와 같은 가상화 서비스 관리 기술의 대표적인 예로는 ETSI MANO와 Linux 재단의 ONAP(Open Network Automation Platform) 등이 있다[11], [12]. ETSI MANO에서는 NFVI 위에서 VNF(Virtualized Network Function)들의 동작을 제어하는 기능의 표준화를 담당하고 있다. MANO는 전체적인 Network Service를 관리하는 NFVO, 각 VNF를 관리하는 VNFM(Virtualized Network Function Manager) 그리고 실제로 그 NFVI를 제어하는 VIM(Virtualized Infra-structure Manager)으로 구성된다. MANO는 네트워크 서비스 및 VNF에 대해서 온보딩, 라이프사이클 관리, 성능 관리, 장애 관리 및 포워딩 그래프 관리 등에 집중하고 있다. ONAP는 그 기능이 크게 설계 단계와 운용 단계로 구분되어 있다. 설계 단계에서는 VNF 온보딩, 서비스 설계, 정책 생성, 워크플로우 설계 및 분석 응용 온보딩 기능들을 수행하고 운용 단계에서는 MANO의 기능과 모니터링을 통한 서비스 보장 기능을 수행한다. 또한 이 두 단계는 서로 협력하여 실시간 상황 모니터링 및 분석을

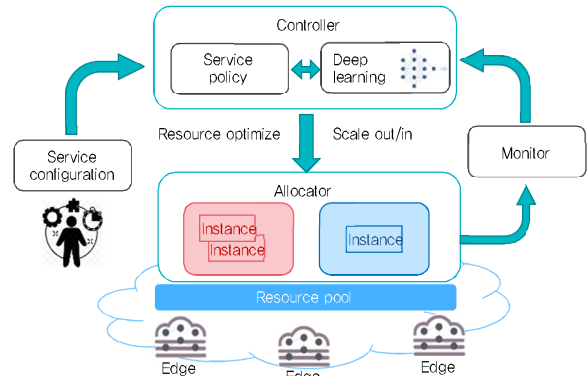
통해서 Closed-loop Automation 기능을 수행한다.

3. 동적 자원 자율관리 기술

클라우드 컴퓨팅 구조에서는 다양한 어플리케이션들이 공통된 하드웨어자원을 공유하고, 가상화 기술을 통하여 자원을 상호 분리하며 각각 할당된 자원 속에서 어플리케이션들이 수행된다. 기존 클라우드 컴퓨팅에 비하여 분산 에지 컴퓨팅은 상대적으로 자원이 제한적이며, 서로 다른 자원을 가지는 다수의 에지 컴퓨팅 노드들을 통하여 여러 어플리케이션이 수행되어야 하기 때문에 서비스 어플리케이션이 배치되는 경우뿐만 아니라 수행 중에도 각 어플리케이션의 요구사항을 만족하기 위한 정교한 동적 자원 관리 기술이 요구된다. 일반적으로 동적 자원 관리 기술은 최적 자원 구성을 위한 자원 최적화 기능, 서비스 품질 만족을 위한 자동 스케일링 기능, 서비스 요구사항 구성을 위한 서비스 정책 기능, 실시간 서비스 상태 감시 기능 등을 포함한다.

하지만, ETSI MEC, OpenFog, Open Edge Computing 등과 같은 에지 컴퓨팅을 다루는 표준화 단체에서는 서비스 구조와 인터페이스 및 Use Case들을 주로 다루고 있으며, 기존의 구조를 확장하여 에지 컴퓨팅 기능을 수용하려는 오픈소스 단체인 OpenStack++, OPNFV 등에서도 오케스트레이션 기능의 하부기능으로 포함되어 있을 뿐 자원관리 기능에 대해 자세히 다루지는 않는다. 반면에 최근 에지 컴퓨팅에 대한 중요성이 부각되면서 학계를 중심으로 분산 컴퓨팅 환경에서 인공지능을 활용한 동적 자원 관리 기능에 관한 많은 연구 논문이 제시되고 있다[13], [14].

(그림 4)는 인공지능을 활용하여 어플리케이션마다 서로 다른 성능 요구사항을 가지고 있는 상황에서 제한된 자원을 최적으로 할당하여 성능 요구사항을 만족하기 위한 동적 자원 자율관리 구조를 나타낸다. 구조는 크게 Controller, Monitor, Allocator로 구성된다. Controller



(그림 4) 분산 에지 동적 자원 자율관리 구조

는 어플리케이션 요구사항과 현재 자원 사용 상황, 서비스 품질 등을 입력으로 받아 알고리즘 저장소에서 적합한 스케줄링 알고리즘을 선택하고, 해당 알고리즘을 수행한다. Monitor는 시스템의 자원사용 상황과 서비스 품질 등의 정보를 수집하고 이를 Controller에게 전달하여 Controller의 정보를 계속해서 업데이트하는 역할을 수행한다. Allocator는 Controller 알고리즘에 의해 결정된 자원 할당량을 바탕으로 실제 응용서비스를 배치한다. 사용자의 개입 없이 인공지능을 활용하여 반복학습을 수행하고 입력으로 들어오는 어플리케이션의 요구사항을 만족시키면서 최소한의 자원을 할당하도록 동작하게 된다.

분산 에지 컴퓨팅 환경에서는 위치 정보를 고려하여 자원을 할당하고 서비스를 배치하는 것도 중요하다. 예를 들어 에지 컴퓨팅을 활용하여 서비스 체인(Service Function Chain)을 구성할 경우 위치 정보를 고려하지 않는다면 서로 상이한 위치의 컴퓨팅 자원으로 서비스 체인을 구성하여 성능 문제를 야기할 수 있다. 따라서, 위치 기반으로 근거리의 에지 컴퓨팅 환경을 클러스터로 구성하고 하나의 클러스터로 들어온 서비스 플로는 최대한 같은 클러스터 내에서 서비스받도록 유도되고 같은 클러스터에 존재하지 않는 서비스 기능(Service Function)을 사용해야 할 경우 최대한 가까운 클러스터에 존재하는 기능을 할당하여 전송 지연을 줄이는 방법

을 사용할 필요가 있다[15].

상기 기술한 내용 이외에도 분산 에지 컴퓨팅 환경에서 동적 자원 자율 관리에 관한 연구는 다양하게 이루어지고 있다[16]~[18]. 분명한 것은 스마트 공장, 스마트 빌딩, 스마트 시티, 자율 주행차, 가상현실, 증강현실 등 갈수록 복잡해지는 서비스들을 효율적으로 수용하기 위해서는 인프라 차원에서 자원 수요를 미리 예측하고 서비스 상태를 감시하여 선제적으로 대응하는 자율 제어 및 관리 기술에 대한 요구가 커지고 있다는 것이다. 따라서, 서비스 요구사항을 만족하면서 제한된 자원의 효율성을 최대화 하기 위해 인공지능을 적극 활용하는 연구가 지속될 것으로 예상된다.

4. 분산 에지 네트워킹 기술

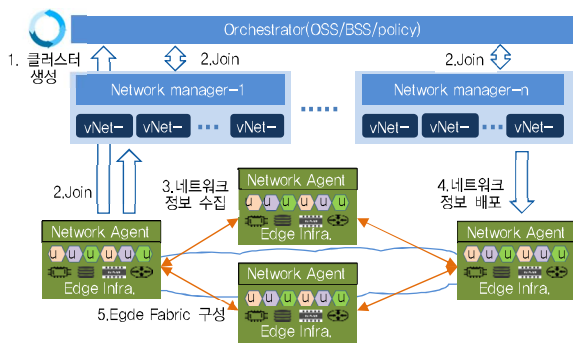
WAN(Wide Area Network)의 에지에 연결된 장치인 CPE(Customer Premises Equipment)는 가입자의 맥내(구내)에 설치되어 통신 사업자의 종단 지점과 연결되는 모든 단말과 관련 장비이다. CPE가 위치하는 곳은 통신 서비스 제공자의 분배 기반 구조나 중앙 시설에 위치한 장비와 고객의 장비가 분리되는 건물 내부 지점이다. CPE는 통신 서비스 사업자의 서비스에 연결하고 LAN을 통하여 집안 또는 기업의 건물 내에서 서비스를 사용할 수 있게 하는 제품으로 보통 라우터, 스위치, 셋톱 박스, 홈 네트워킹 어댑터, 인터넷 접근 게이트웨이 등의 장비 형태로 존재하며, 가입자가 직접 구매하거나 서비스 사업자에 의해 제공되기도 한다[19].

일반적으로 CPE를 구성하는 라우터, WAN 가속기, Firewall 등 네트워크 장비들은 물리적으로 분리되어 있고 각각의 장비들을 설치, 네트워크 연결, 설정, 운영, 업데이트 및 장비 교체 등 구축 비용과 운영 비용이 많이 든다는 단점이 있다. 따라서, NFV의 출현과 함께 서버 기반 네트워킹 소프트웨어 성능이 높아짐에 따라 기존 CPE의 단점을 해소할 수 있는 VNF 기반의 vCPE(Universal Customer Premises Equipment) 기술이 많

은 관심을 받고 있다. vCPE는 각 네트워크 장비 기능을 가상화하여 전용 하드웨어 없이 소프트웨어만 범용 서버에 올려 네트워크 기능을 서버에서 소프트웨어(VNF)로 대체하는 것으로 기존의 장비들의 네트워크 기능을 VNF가 수행한다[20], [21]. 즉, vCPE 환경에서는 기업 고객이 Firewall 서비스를 요청하면 통신사 설치기사가 Firewall 장비를 들고 기업에 방문하여 설치해주는 것이 아니라 NFVO가 Virtual Firewall vCPE를 범용 서버로 원격 설치한다. 중앙 제어 방식을 사용하여 OPEX 비용을 최소화하고, 관리 방식을 단순화하여 배포 시간을 혁신적으로 줄일 수 있다.

SD-WAN(Software Defined WAN)은 고비용의 MPLS(Multi-Protocol Label Switching)와 저비용의 초고속 인터넷 회선을 이용해 기업이 사설 오버레이 네트워크를 생성할 수 있게 해주는 네트워크 기능을 의미하는데, 일반적으로 중앙의 SD-WAN 컨트롤러와 각 에지 장치로 구성된다. 최근에는 MPLS를 사용하는 것이 비용측면에서 비효율적인 이유로 저비용의 초고속 인터넷 회선을 증설하고 MPLS를 줄여가는 추세이다. 그러나 화상회의나 VoIP와 같은 리얼타임 어플리케이션을 사용하는 기업의 경우에는 패킷 손실 및 지연에 장점을 갖는 MPLS를 사용한다[22], [23].

분산 에지 네트워킹 기술은 원격지 에지 상호 간 WAN 구간을 연결하기 위한 SD-WAN 기술과 vCPE와 같은 가상화 기반의 마이크로서비스를 구성하는 에지 노드 내부에 수행되는 컨테이너들을 상호 연결할 수 있는 컨테이너 네트워킹 기술로 이루어진다. (그림 5)와 같이 SD-WAN을 구성하는 측면에서 중앙의 OSS(Operations Support System), BSS(Business Support System), Policy Manager에는 사용자 및 그룹 관리 기능 등이 수행되고, Fabric Manager를 통하여 각 Fabric Agent로부터 Network 접속 정보를 수집, 배포, 터널 제어 기능 등이 동작되며, 각 에지 노드의 Fabric Agent는 Fabric Manager와 메시지 송수신을 통하여 Peer 에지



(그림 5) 분산 에지 네트워킹 구조

노드들과 Pull Mesh Fabric을 구성하기 위하여 Peer-To-Peer 터널 설정 기능을 수행한다[24], [25].

컨테이너 기반의 마이크로 서비스는 여러 노드를 연결한 클러스터 기반으로 만들어지며, 각 노드 내부의 컨테이너 기반 마이크로 서비스가 실행된다. 노드 내부에서 실행되는 각 컨테이너는 사설 주소를 사용하게 되며, 서비스를 위한 각 마이크로 서비스 간에 연결성을 제공하는 것이 컨테이너 네트워킹 제어기술이다. 마이크로 서비스들이 실행되는 컨테이너 간 연결성을 제공하려면 사설 주소 사용에 따른 문제를 해결해야 한다. 서로 다른 노드 내부의 컨테이너 간에 연결성을 제공하기 위하여 가상 인터페이스 및 가상 네트워크 개념을 적용한다. 이를 위하여 각 노드 사이에 가상 네트워크와 가상인터페이스를 구성하고 이를 연결하기 위한 터널을 설정하고 각 노드 내부의 컨테이너 연결성을 제공하기 위하여 라우팅 테이블을 설정하는 방식으로 연결성을 보장한다[26].

5. 실시간 Telemetry 측정 및 감시 기술

에지 노드의 자원을 관리하기 위해서는 운영과 관리에 대한 정보로 Compute, Storage, Network뿐만 아니라 GPU, FPGA(Field Programmable Gate Array), MLU, Smart-NIC(Network Interface Card) 등 다양한 가속기에 대한 정보가 사용된다. 텔레메트리 기능은 에이전트(Agent)와 콜렉터(Collector) 구조로 에지 노드의 자원 상태를 수집하여 이를 필요로 하는 컴포넌트에 배포

하여 에지의 마이크로 서비스 아키텍처에서 필요한 상황 정보를 제공하게 된다.

전통적 Collector는 Pull 기반으로 동작하였으며 주로 SNMP(Simple Network Management Protocol), CLI show 명령어, 시스템 로그를 분석하여 시스템을 관리하는데 필요한 정보로 사용하였다. 그러나 이러한 방법들은 정보의 완전성(Complete)과 확장성(Scale), 그리고 효율성(Efficient)에서 문제가 되며 실시간성이 떨어지고, 비구조적(Unstructured)이다. 이러한 단점을 극복하기 위하여 구글 등의 대형 네트워크의 운영자들은 OpenConfig라는 벤더 중립적인 일관된 데이터 모델을 개발하였으며 Streaming 텔레메트리를 네트워크 모니터링에 적용하고 있다. 이를 통하여 네트워크의 성능 데이터 수집, 실시간성 지원, 모델 기반의 확장성 확보, 분석에 바로 사용이 가능한 데이터 수집이 가능하도록 하였다.

sFlow는 고속 스위칭 네트워크의 모니터링을 위한 산업표준으로 어플리케이션, 가상 서버, 가상 네트워크, 서버, 네트워크에 대한 데이터 수집이 가능하다. 전달하려는 정보는 여러 개로 구분된 구조체 형태를 사용하며 전송은 UDP 프로토콜로 인코딩은 XDR(RFC 4506)을 따른다[27].

P4는 데이터 평면을 프로그래밍하기 위한 언어로 스위치가 패킷을 어떻게 처리할지 프로그램이 가능한 프로토콜 독립성, 고성능의 ASIC부터 소프트웨어 스위치까지 적용이 가능한 타겟 독립성 그리고 장비가 적용되고 난 이후에도 스위치의 패킷 처리 방식 변경이 가능한 특징이 있다[28]. 이러한 특징은 OpenFlow의 단점인 데이터 평면에 대한 부족함을 메워줄 대안으로 제시되고 있다. 특히, 제어 평면의 개입 없이 데이터 평면에서 네트워크 상태를 수집하고 보고하는 In-band Network Telemetry(INT) 프레임워크는 P4의 특징 중 하나로 SAI에서도 INT의 표준화가 이루어지고 있다.

이 외에도 Insitu OAM(iOAM)으로 대표되는 IETF의



(그림 6) 실시간 Telemetry 측정 및 감시 구성도

Streaming 텔레메트리는 패킷이 네트워크에서 전달될 때 패킷에서 플로우의 경로 추적과 검증 정보, 성능 지표 정보를 직접 취득하며 어플리케이션에 특화된 메타데이터도 포함하고 있다. Flow Information Export(IPFIX) 프로토콜은 관리나 다른 목적으로 사용이 가능한 트래픽 플로우 정보를 제공한다.

서버 시스템에서는 CPU, 펌웨어 그리고 운영체제에 무관하게 시스템 관리와 모니터링을 수행하기 위한 IPMI가 있으나 보안 문제가 언급되고 있으며, 이에 Facebook의 BMC(Baseboard Management Controller) 결과물을 활용하는 OpenBMC가 제안되었다. SDN/NFV가 활성화되면서 ATIS에서는 가상화 환경에서 유형별 가상화된 리소스와 어플리케이션 서비스, 가상 저장소에 대한 모니터링이 필요한 주요 인자들에 대해 열거하고 있다.

Collector에서 수집된 데이터들은 실시간 혹은 분, 시, 일 등 주기적으로 배치 처리된다. 전달되는 데이터와 처리된 데이터는 향후 이용을 위해 저장되며, 이때 대량의 비정형 데이터를 정형화하여 처리할 수 있는 빅데이터 플랫폼이 고려된다. 외부 컴포넌트들은 지속적인 모니터링을 위하여 Collector에서 발행(Publish)하는 이벤트를 구독(Subscribe)하게 되며 임계값을 벗어난 이벤트에 대해서는 알람을 받게 된다(그림 6) 참조].

6. Zero Touch Provision(ZTP) 기술

데이터 센터를 구축하는 기존 방법들에서는 사람이 서버에 일일이 운영체제를 설치하고 설정한 다음 서비스 데몬을 설치하고, 서비스 환경을 설정하는 과정을 거친다. 이러한 과정은 가상화를 사용하기 위한 하이퍼바이

저 혹은 컨테이너를 구성하는 과정과 장비에 문제가 발생하여 새로운 장비로 대체되어야 할 경우에도 동일하게 반복된다. 이러한 반복 작업은 운용 비용의 증가를 가져오므로 사람의 개입을 최소화하기 위한 방안이 모색되었다. 소프트웨어 설치와 구성을 자동화하기 위한 자동 권한 부여(Zero Touch Provision) 기술은 인증된 소프트웨어 혹은 이미지로 장치 브링-업(Bring up), 원격 접속을 위한 초기 설정 혹은 스크립트 다운로드, 장치의 NMS나 컨트롤러 혹은 오케스트레이터 연결까지를 담당한다.

ZTP를 구성하는 주요 기술 중 하나가 부트 펌웨어이다. 부터(Booter), 부트 로더(Boot Loader) 등으로 불리는 부트 펌웨어(Boot Firmware)는 시스템 전원이 켜지면, 하드웨어 셀프 테스트(POST: Power On Self-Test)를 수행한 다음 설치된 운영체제에 제어를 넘겨주는 역할을 수행한다. 만약 설치된 운영체제가 없는 경우에는 운영체제를 설치하는 역할을 수행한다.

부트 펌웨어는 주로 서버에 iPXE가 네트워크 장비에는 ONIE(Open Network Install Environment)가 사용된다. iPXE는 상용 소프트웨어인 PXE의 모든 기능에 HTTP 프로토콜을 이용한 부팅, iSCSI SAN을 이용한 부팅, 무선 네트워크를 이용한 부팅, 광대역 네트워크를 이용한 부팅이 추가되었으며 스크립트를 이용한 부트 과정 제어도 가능하다.

OCP(Open Compute Project)에서 ONL(Open Network Linux)과 함께 사용되는 부트 펌웨어는 ONIE로 네트워크 운영체제를 설치해주는 환경이다. 이 과정은 로컬 디스크, USB 혹은 네트워크 연결을 통하여 설치할 NOS(Network Operating System)의 인스톨러 이미지를 장치에 설치하고 재부팅 후 ONIE 과정 없이 설치된 NOS로 장치가 구동되게 된다. ONIE를 사용함으로써 ONIE의 규약을 따르는 어떠한 NOS도 설치가 가능하며, 많은 수의 네트워크 장비를 갖는 데이터센터에서 각 장치에 대한 프로비저닝을 가능하게 한다. 이로써 서버를 관

리하는 것과 유사한 방식으로 네트워크 장치를 제어할 수 있다. 이를 위하여 ONIE에서는 여러 종류의 장치에 대한 설치를 지원하기 위해 설치할 운영체제 이미지 파일 이름에 ARCH, VENDOR, MACHINE, REVISION을 포함하도록 하고 있다. 부트 펌웨어가 가동되고 난 이후 운영체제를 설치하는 과정은 IP 주소 획득과 설치 파일 다운로드, 그리고 운영체제로 부팅되고 난 후에 초기 설정을 위한 환경 설정 파일을 다운로드 한 후 설정 과정을 거침으로써 자동 권한 부여 기능을 완료하게 된다.

대개의 경우 장치의 외부 관리 포트에 관리 네트워크에 연결하여 사용하며, 다운로드를 위하여 주로 DHCP (Dynamic Host Configuration Protocol) 옵션을 이용한다. 기본적으로 Host를 구분하기 위해서는 MAC 주소가 사용되는데 DHCP 옵션 61번 Client Identifier에 장치의 시리얼 번호와 같은 고유 번호를 이용할 경우 YANG(Yet Another Next Generation) 모델과 결합하여 대단위의 권한 부여가 가능하다.

IV. 결론

과거 웹(WWW)과 개인 컴퓨터의 시대를 거쳐 스마트 디바이스의 폭발적인 증가로 접어든 모바일 시대에는 다양한 앱들과 멀티미디어 데이터를 수용하기 위하여 네트워크 분야에서 양적, 질적 성장을 이루어 왔으며, 컴퓨팅 영역에서도 클라우드 컴퓨팅의 급진적인 발전을 견인해 왔다. 하지만, 모바일 시대를 넘어 도래하고 있는 지능, 실감, 실시간 서비스 위주의 포스트-앱 시대에는 단순히 양적, 질적 성장을 넘어서서 네트워크와 컴퓨팅 인프라를 새롭게 구성하는 패러다임의 전환이 이루어지고 있다. 기존에는 IP 기반 연결 중심의 네트워킹과 중앙 클라우드 중심의 컴퓨팅이 상호 별개로 발전해 왔으나, SDN/NFV의 발전으로 네트워크가 소프트웨어 중심으로 전환되고, 에지 컴퓨팅의 등장으로 네트워크 자체가 컴퓨팅을 포함하는 융합된 형태의 인프라로 진

화하고 있다. 이와 함께 사람과 사물이 직접적으로 접속하는 네트워크 에지의 역할의 중요성은 갈수록 증대되고 있으며, 컴퓨팅 능력을 가진 네트워크에 인공지능을 결합하여 지능화된 형태로 발전해 나가고 있다.

본고에서는 이러한 지능형 에지 네트워킹 및 컴퓨팅에 대한 연구들과 이와 관련된 최신 기술 동향에 대해 기술하였다. 앞으로도 연구소, 학계, 산업계가 협업하여 다가오는 새로운 시대에 맞는 새로운 지능형 인프라에 대한 혁신이 성공적으로 이루어지길 기대해 본다.

약어 정리

AI	Artificial Intelligence
ATIS	The Alliance for Telecommunications Industry Solutions
BMC	Baseboard Management Controller
BSS	Business Support System
COTS	Commercial Off the Shelf
CPE	Customer Premises Equipment
DHCP	Dynamic Host Configuration Protocol
DNN	Deep Neural Networks
ETSI	European Telecommunications Standards Institute
FPGA	Field Programmable Gate Array
GKE	Google Kubernetes Engine
IEC	Intelligent Edge Computing
IPMI	Intelligent Platform Management Interface
iPXE	open-source implementation of the PXE
iSCSI	Internet Small Computer System Interface
MANO	Management and Orchestration
MEC	Mobile Edge Computing
ML	Machine Learning
MPLS	Multi-Protocol Label Switching
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure
NFVO	Network Function Virtualization Orchestrator
NIC	Network Interface Card
NOS	Network Operating System
OCP	Open Compute Project

ONAP	Open Network Automation Platform
ONIE	Open Network Install Environment
ONL	Open Network Linux
OSS	Operations Support System
POST	Power On Self-Test
PXE	Preboot eXecution Environment
SD-WAN	Software Defined WAN
SDN	Software Defined Network
SNMP	Simple Network Management Protocol
vCPE	Universal Customer Premises Equipment
VF	Virtual Function
VIM	Virtualized Infrastructure Manager
VNF	Virtualized Network Function
VNFM	Virtualized Network Function Manager
WAN	Wide Area Network
XDR	External Data Representation
YANG	Yet Another Next Generation

참고문헌

- [1] M. Resnick, N. Jones, and M. O'Neill, "Preparing for a World Beyond Apps," Gartner, May 18, 2017.
- [2] K. Panetta, "Gartner's Top 10 Strategic Technology Trends for 2017," Gartner, Oct. 18, 2016.
- [3] D. Cearley, B. Burke, and M. Walker, "Top 10 Strategic Technology Trends for 2016," Gartner, Feb. 29, 2016.
- [4] Frost & Sullivan, "Future Applications of Virtual Reality (Immersive Computing)," Frost & Sullivan, Dec. 30, 2016.
- [5] LG CNS, "Creative & Smart Welcome to LG CNS official blog," LG CNS, Apr. 2018.
- [6] B. Butler, "에지 컴퓨팅의 이해와 네트워크의 변화," ITWorld Korea, Sept. 26, 2017.
- [7] D. Reinsel, J. Gantz, and J. Rydning, "The Digitization of the World from Edge to Core, DATA AGE 2025," IDC, Nov. 2018.
- [8] S. Teerapittayanon, B. McDanel, and H. Kung, "Distributed Deep Neural Networks over the Cloud, the Edge and End Devices," in *Proc. IEEE Int. Conf. Distr. Comput. Syst.*, Atlanta, GA, USA, June 5-8, 2017, pp. 328-339.
- [9] E. Li, Z. Zhou, and X. Chen, "Edge Intelligence: On-Demand Deep Learning Model Co-inference with Device-Edge Synergy," in *Proc. Workshop Mobile Edge Commun.*, Budapest, Hungary, Aug. 20, 2018, pp. 31-36.
- [10] Y. Kang et al., "Neurosurgeon: Collaborative Intelligence between the Cloud and Mobile Edge," in *Proc. Int. Conf. Archit. Support Program Lang. Oper. Syst.*, Xi'an, China, Apr. 8-12, 2017, pp. 615-629.
- [11] ONAP Project, <https://onap.readthedocs.io/en/latest/>
- [12] A. Xiang et al., "NFV Release 2: Management and Orchestration: Functional Requirements Specification," ETSI GS NFV-IFA 010 V2.3.1, Aug. 2017.
- [13] K. Tocz and S. Nadjim-Tehrani, "A Taxonomy for Management and Optimization of Multiple Resources in Edge Computing," *Wireless Commun. Mobile Comput.*, vol. 2018, May 2018, pp. 7476201:1-7476201:23.
- [14] B.J. Ko and K.K. Leung, "Machine Learning for Dynamic Resource Allocation at Network edge," in *SPIE Defense+ Security*, Orlando, FL, USA, May 2018.
- [15] Y. Nam, S. Song, and J.M. Chung, "Clustered NFV Service Chaining Optimization in Mobile Edge Clouds," *IEEE Commun. Lett.*, vol. 21, no. 2, Feb. 2018, pp. 350-353.
- [16] W. Tärneberg et al., "Dynamic Application Placement in the Mobile Cloud Network," *Future Generation Comput. Syst.*, vol. 70, May 2017, pp. 163-177.
- [17] Y. Zhang, J. Yao, and H. Guan, "Intelligent Cloud Resource Management with Deep Reinforcement Learning," *IEEE Cloud Comput.*, vol. 4, no. 6, Oct. 2017, pp. 60-69.
- [18] W. Tärneberg et al., "Resource Management with Deep Reinforcement Learning," *HotNets '16, Proc. ACM Workshop Hot Topics Netw.*, Atlanta, GA, USA, Nov. 9-10, 2016, pp. 50-56.
- [19] Wikipedia, "Customer-Premises Equipment," Available: https://en.wikipedia.org/wiki/Customer-premises_equipment
- [20] AT&T, "AT&T Universal Customer Premises Equipment (uCPE)," AT&T Whitepaper, June 2016.
- [21] Netmanias Blog, "SDN/NFV-based Enterprise Service (vCPE) Commercialization - AT&T and Verizon," Aug. 2016.
- [22] M.C. Smith, "Is SD-WAN better than MPLS?" Network World, Sept., 2017. Available: <https://www.networkworld.com/article/3226371/wide-area-networking/is-sd-wan-better-than-mpls.html>
- [23] J. Ruparel, "How SD-WAN as a Service Solves MPLS Limitations," sdx Central, Oct., 2018. Available: <https://www.sdxcentral.com/articles/contributed/sd-wan-as-a-service-solves-mpls-limitations/2018/10/>
- [24] Riverbed, "Key Requirements for SD-WAN: Rethink

Networking for the Cloud Era,” Riverbed Whitepaper, 2017.

[25] ZeroTier Project, “ZeroTier Manual,” Available: <https://my.zerotier.com/>

[26] K8S Project, “Cluster Networking Documentations,”

Available: <https://kubernetes.io/docs/concepts/cluster-administration/networking/>

[27] sFlow Project, <https://sflow.org/>

[28] P4 Project, <https://p4.org/>