

Chapter
01

컴퓨팅 컨티뉴엄: 임베디드-엣지-클라우드 컴퓨팅 연계 기술 동향

강성주_한국전자통신연구원 책임연구원
전인걸_한국전자통신연구원 책임연구원

컴퓨팅 컨티뉴엄(Computing Continuum)은 임베디드-엣지-클라우드 컴퓨팅 등 다계층의 컴퓨팅 요소를 하나의 '연속체'로 바라보고, 전체 컨티뉴엄 상의 최적의 위치에 서비스에 필요한 컴퓨팅 리소스를 정확히 배치하기 위한 개념이다. 본 고에서는 컴퓨팅 컨티뉴엄의 개념 및 관련된 기초 연구 동향을 살펴본다. 이와 함께 GigaMEC/CompactEdge/EdgeCPS의 특징과 연계 구조를 파악해봄으로써 컴퓨팅 컨티뉴엄 개념의 실현 가능성을 엿보고, 향후 6G 시대를 대비한 미래형 서비스 개발 전략을 공유한다.

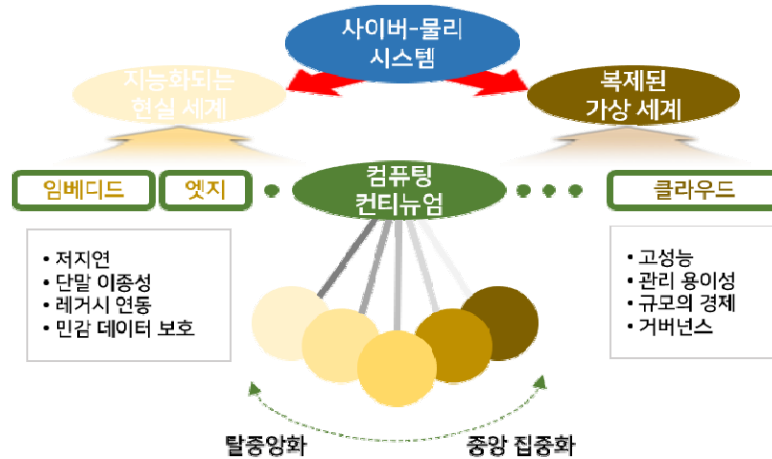
I. 서론

컴퓨터 공학의 역사에서 데이터를 어디에 저장하고, 어디에서 처리할지, 즉 컴퓨팅 플랫폼 배치에 대한 접근 방식은 기술의 진보와 시대 상황에 따라 반복적으로 발전해 오고 있다. 즉, 중앙집중형(centralized) 컴퓨팅과 탈 중앙형(decentralized) 컴퓨팅 패러다임은 메인 프레임에서 모바일 컴퓨팅으로, 그리고 클라우드 컴퓨팅을 거쳐 다시 엣지 컴퓨팅까지 [그림 1]의 진자처럼 움직이며 기술 발전을 이루어왔다[1]. 더불어 현실 세계의 가장 말단인 사람들은 스스로 학습하고, 자율적으로 행동하고, 인간과 소통하며 현실 세계의 지능화에 이바지하기 위한 기술을 탑재하기 시작하고 있다. 또한, 다른 한쪽에서는 현실 세계의 객체와 관계, 정보와 논리, 의사 결정 체계가 그대로 복제된 가상 세계(또는 디지털 세계)가 데이터센터 어딘가에서 성장해 나가고 있으며, 궁극적으로는 지능화되는 현실 세계와 동기화하여 자동

* 본 내용은 강성주 책임연구원(☎ 042-860-5881, sjkang@etri.re.kr)에게 문의하시기 바랍니다.

** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

***이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2020-0-00844, 엣지 서버 시스템 자원 관리 및 제어를 위한 경량 시스템 소프트웨어 기술 개발)



〈자료〉 한국전자통신연구원 자체 작성

[그림 1] 컴퓨팅 컨티뉴엄을 통한 사이버-물리 시스템의 실현

적, 지능적으로 현실 세계의 사물이 제어되는 사이버-물리 시스템(Cyber-Physical Systems: CPS)을 지향해 가고 있다.

지능화된 현실 세계와 가상 세계가 융합된 사이버-물리 시스템 환경에서는 서비스의 요청자인 현실 세계 속 사물(스마트 기기, 자율주행차, 드론 등)에게 제공될 서비스와 데이터를 처리할 최적의 위치를 결정하는 것이 매우 중요하다. 특히, 시스템의 구현을 위해 사물에 내장된 컴퓨터, 엣지 컴퓨팅 노드 또는 클라우드 컴퓨팅 환경 등을 구체적으로 결정하는 것은 아래와 같은 이유로 인해 어려움이 발생한다.

첫째, 사물이 속한 현실 세계는 시시각각 변화하는 동적 환경으로서 동일한 서비스 요청이더라도 상황에 따라 응답 속도와 체감하는 품질이 다르게 나타난다. 인구가 밀집된 도심지에서 통화 연결이나 멀티미디어 콘텐츠 재생이 지연되는 것이 대표적인 예이다.

둘째, 동일한 서비스를 요청하는 개별 사물의 규격이나 컴퓨팅 역량에 따라 제공할 서비스의 형상이 다르다. 단일 소프트웨어 인스턴스로 서로 다른 여러 사용자 그룹에 서비스를 제공할 수 있는 멀티테넌시(multi-tenancy) 개념을 예로 들 수 있다.

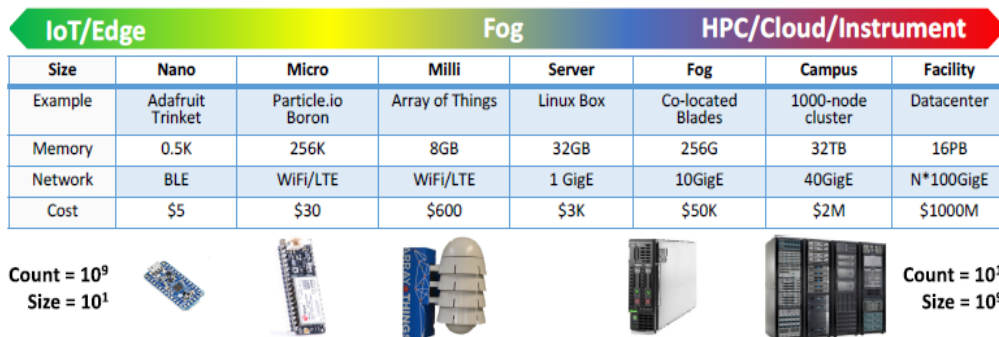
셋째, 서비스의 특성에 따라 이를 제공하기 위한 자원의 속성이 다르다. 예를 들어, 스마트 공장 내 CNC 설비의 모션 제어 서비스는 빠른 패킷 처리를 요구할 것이고, 과학적 연산을 위한 시뮬레이션 서비스는 고성능 컴퓨팅 인프라를 요구할 것이다.

세 가지 예시 모두 개별 사물 관점에서의 서비스 응답 시간과 같은 체감 품질(Quality of Experience: QoE)을 만족할 수 있도록 최적 위치에서 최적의 형태로의 서비스 제공이 필요하다. 이를 위해서는 서비스를 요청하는 사물과 사물이 속한 현실 세계의 운영 환경 그리고 서비스를 제공하기 위한 다양한 형태의 컴퓨팅 요소 간의 상호 연계와 실시간 관측이 가능해야 한다. 컴퓨팅 컨티뉴엄(Computing Continuum)은 다음과 같은 목적으로 임베디드-엣지-클라우드 컴퓨팅 등 다계층의 컴퓨팅 요소를 하나의 '연속체'로 보고, 실세계의 서비스 요청에 대해 분산된 컴퓨팅 자원과 서비스를 통합하여 동적으로 제공하는 개념이다[2],[3].

본 고에서는 임베디드-엣지-클라우드 컴퓨팅 연계를 표현하는 개념으로서 컴퓨팅 컨티뉴엄 관련 연구 동향에 대해 다룬다. II장에서는 컴퓨팅 컨티뉴엄 개념 정의와 기초 연구를 살펴보고, III장에서는 임베디드-엣지-클라우드 컴퓨팅 연계 플랫폼 기술 관련 연구 과제 소개를 통해 컴퓨팅 컨티뉴엄 개념의 실현 가능성을 논하고, IV장에서는 본 고의 결론을 제시한다.

II. 컴퓨팅 컨티뉴엄 기초 연구 사례

컴퓨팅 컨티뉴엄은 디지털 컨티뉴엄(Digital Continuum)이나 트랜스 컨티뉴엄(Trans Continuum)으로도 표현되며, 기술이 포함하는 범위나 적용 시나리오에 따라 다르지만, 공통적으로는 [그림 2]와 같이 계층적으로 구분되는 여러 컴퓨팅 요소들이 연결되어 하나의



〈자료〉 P. Beckman, J. Dongarra, N. Ferrier, G. Fox, T. Moore, D. Reed, and M. Beck, "Harnessing the Computing Continuum for Programming Our World," Wiley, Fog Computing: Theory and Practice, 2020, pp.215-230.

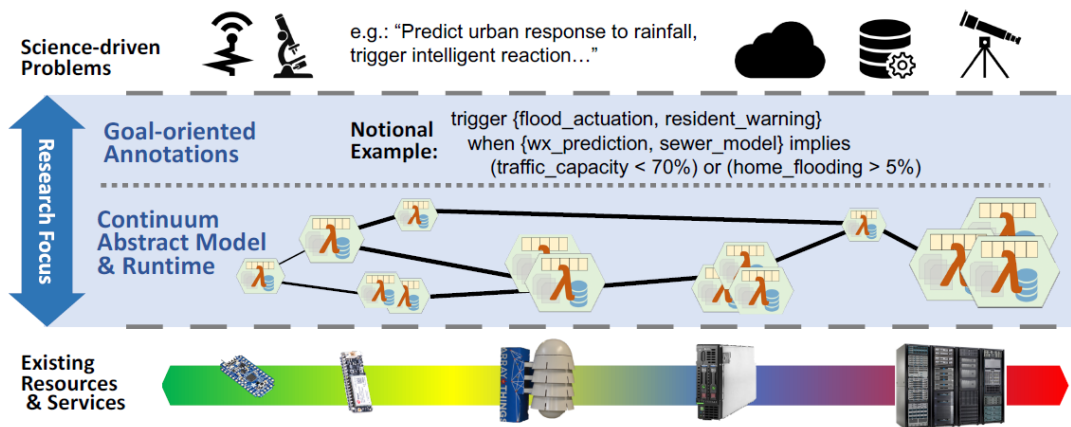
[그림 2] 컴퓨팅 컨티뉴엄의 개념

구조를 이루는 형상을 의미한다.

1. 아르곤 연구소

미국의 국립아르곤연구소(Argonne National Laboratory)는 환경 정보를 획득하는 각종 센서 입력부터 HPC에서의 예측 시뮬레이션까지 연계를 바탕으로 하는 의사결정 지원 방법으로서 컴퓨팅 컨티뉴엄의 활용 가능성을 연구하고 있다. 미국 국립과학재단(NSF)에서 지원받는 대규모 사이버인프라스터럭처(Large Facilities Cyberinfrastructure) 관련 과제 [4]들을 대상으로 관련 연구 방법론과 개념 실현을 위한 요소 기술들을 제시하였다. [그림 3]은 대표적인 예시로서, “현재부터 8시 후에 강우가 예상될 때, 사회 기반 시설의 지능형 대응 체계”의 동작 구조를 보여준다.

사회 기반 시설인 하수도와 고속도로에 설치된 IoT 장치부터 기상 예측과 고성능 연산을 위한 HPC까지 하나의 컴퓨팅 컨티뉴엄을 이루고, 컴퓨팅 컨티뉴엄 상에서는 상황 발생 시 조건에 따른 대응 전략을 기술한 목표 지향형 명세(Goal-oriented Annotations)에 따라 분산된 응용(Continuum Abstract Model & Runtime)이 동작한다. 동작 시나리오에 따르면 컴퓨팅 컨티뉴엄 상의 HPC는 강우 예상과 인접 도시로의 유입 예측, 하수도 범람 가능성 등을 주기적으로 시뮬레이션하고, 포그/엣지 컴퓨팅 시스템에서는 영상과 센서 기반의 하천



〈자료〉 P. Beckman, J. Dongarra, N. Ferrier, G. Fox, T. Moore, D. Reed, and M. Beck, “Harnessing the Computing Continuum for Programming Our World,” Wiley, Fog Computing: Theory and Practice, 2020, pp.215-230.

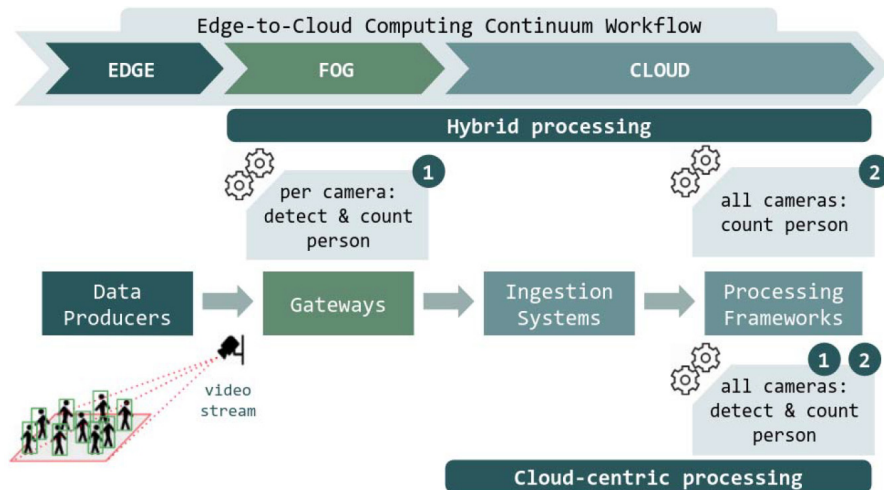
[그림 3] 컴퓨팅 컨티뉴엄의 활용 예시

/도로 감시와 펌프/밸브의 강인성 확인을 수행하다가 실제 강우 발생 시 도시의 대응 체계와 연계하는 등 다양한 활용 가능성을 제시한다.

2. 국립 전산 및 자동화연구소

프랑스의 국립 전산 및 자동화연구소(INRIA)는 컴퓨팅 컨티뉴엄을 과학적 실험에 활용하기 위한 프레임워크인 E2CLab을 제안하였다[5]. E2CLab은 엣지와 클라우드에 걸친 컴퓨팅 컨티뉴엄 상에서 다양한 환경을 구성하고, 대규모 응용 서비스의 품질 속성을 3R(Repeatability, Replicability, Reproducibility)에 걸쳐 다차원 분석하는 방법을 제안하였다.

[그림 4]의 영상 기반 감시 시스템을 예로 들면, “영상 장치를 통해 획득한 영상을 컨티뉴엄 상의 어느 위치에서 처리하는 것이 가장 적합할까?”라는 질문에 대해 ① 엣지/포그 노드에서 사람의 존재와 인원수를 확인하고, 클라우드에서는 각각의 엣지/포그 노드에서 확인한 인원수를 합하는 방법과 ② 클라우드에서 사람의 존재와 인원수를 모두 확인하는 방법을 비교할 수 있다. 성능 비교를 위해서는 엣지-클라우드 간의 처리량/지연시간, 종단 간 대역폭, CPU/메모리 활용량 등 다양한 지표가 필요한데, 이를 컴퓨팅 컨티뉴엄 상에서 다루기 위한 다양한 관리 모듈을 정의하였다.

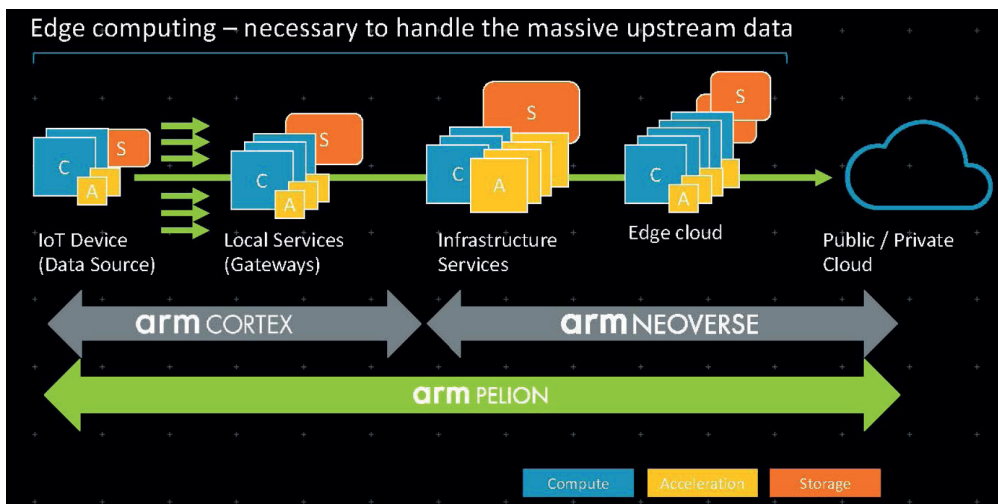


〈자료〉 P. Beckman, J. Dongarra, N. Ferrier, G. Fox, T. Moore, D. Reed, and M. Beck, “Harnessing the Computing Continuum for Programming Our World,” Wiley, Fog Computing: Theory and Practice, 2020, pp.215-230.

[그림 4] 컴퓨팅 컨티뉴엄 상에서 비교 실험 예시

3. ARM 홀딩스

영국의 반도체 회사 ARM은 IoT 장치부터 클라우드까지 필수적으로 탑재되는 프로세서 수준에서부터 컴퓨팅 컨티뉴엄을 구성하고, 이를 다양한 영역에서 활용하기 위한 플랫폼을 확산시키고 있다[6]. ARM은 컴퓨팅 컨티뉴엄을 “엔드포인트부터 클라우드까지 전체 시스템 상 최적의 위치에 필요한 컴퓨팅 리소스를 정확히 배치하기 위한 전략”으로 정의하고 있다. [그림 5]는 ARM의 컴퓨팅 컨티뉴엄 전략을 표현하고 있다.



〈자료〉 ARM, “The New Compute Continuum is Powering the Future of IoT,” ARM research White Paper, 2019, pp.1-14.

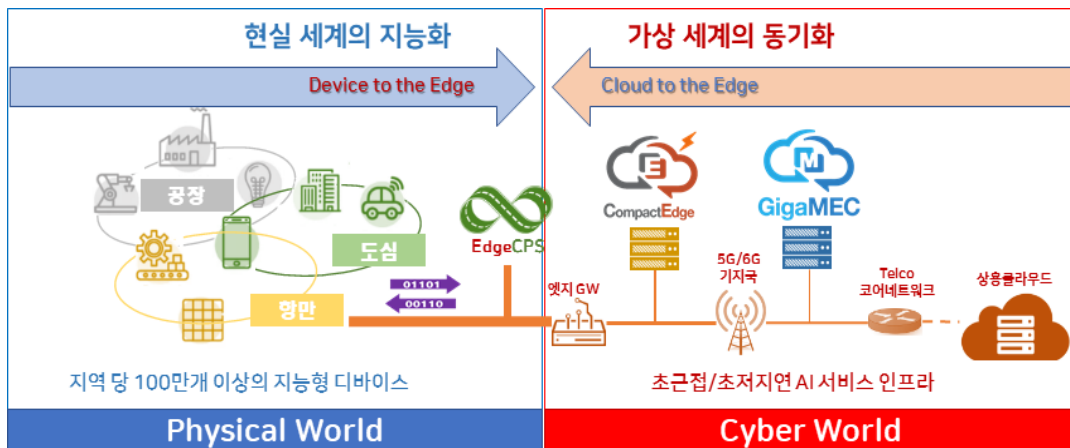
[그림 5] ARM 홀딩스의 컴퓨팅 컨티뉴엄 전략

ARM은 컴퓨팅 컨티뉴엄의 구성 요소를 크게 클라우드-엣지 컴퓨팅-엔드포인트로 정의한다. 특별히 ARM이 바라보는 엣지 컴퓨팅은 엔드포인트 장치가 연결되는 종단 네트워크부터 클라우드 직전까지의 모든 장비를 아우른다. ARM은 엣지 컴퓨팅을 클라우드로부터의 오프로드 계층(offload layer)으로 정의하고 있으며, 향후 AI 관련 대부분의 워크로드(학습/추론)를 다루게 될 가장 성장 가능성이 큰 시장으로 판단하고 있다. 따라서 현실 세계의 다양한 서비스 요구사항이 엣지 컴퓨팅 영역에 형성된 컴퓨팅 컨티뉴엄에서 처리될 수 있도록 플랫폼(ARM PELION) 역량을 집중하고 있다.

III. 임베디드-엣지-클라우드 컴퓨팅 연계 플랫폼

본 장에서는 컴퓨팅 컨티뉴엄을 실현하기 위한 요소 기술로서 MEC-클라우드, 엣지-클라우드, 임베디드-엣지를 연계하는 세 개의 서비스 플랫폼, GigaMEC, CompactEdge, EdgeCPS를 소개한다. 각 플랫폼은 서비스 응용을 “컨테이너로 가상화(containerized)”하고, 가상화된 컨테이너의 배포, 확장, 관리를 자동화해주는 오케스트레이션 기술을 바탕으로 임베디드-엣지-클라우드 컴퓨팅 플랫폼 간의 연계 목표로 개발되었다.

서론에서 기술하였듯이 현실 세계가 복제된 디지털 세계는 [그림 2]의 컴퓨팅 컨티뉴엄 상 우측의 데이터센터에서 성장하고 있다. 이러한 디지털 세계의 완성은 하루아침에 이루어지기 보다는 점진적으로 이루어지고 있으며, 그 방법은 클라우드의 역할을 현실 세계와 가까운 엣지로 오프로딩(Cloud to the Edge)하는 방향과 임베디드 컴퓨팅과 엣지 컴퓨팅과의 분업과 협업을 통해 디바이스를 지능화(Device to the Edge)하는 양방향에서 이루어져야 한다. GigaMEC와 CompactEdge는 클라우드의 역할을 오프로딩하고, EdgeCPS는 디바이스 임베디드 시스템부터 엣지 서버를 거쳐 클라우드 컴퓨팅까지를 하나의 컴퓨팅 컨티뉴엄으로 연계하고, 현실 세계 속 사물(스마트 기기, 자율주행차, 드론 등)에게 제공될 서비스를 위해 필요한 컴퓨팅 리소스를 최적의 위치에 정확히 배치하기 위한 전략을 제공한다. [그림 6]은 컴퓨팅 컨티뉴엄 상에서 세 가지 플랫폼의 포지셔닝을 보여준다.



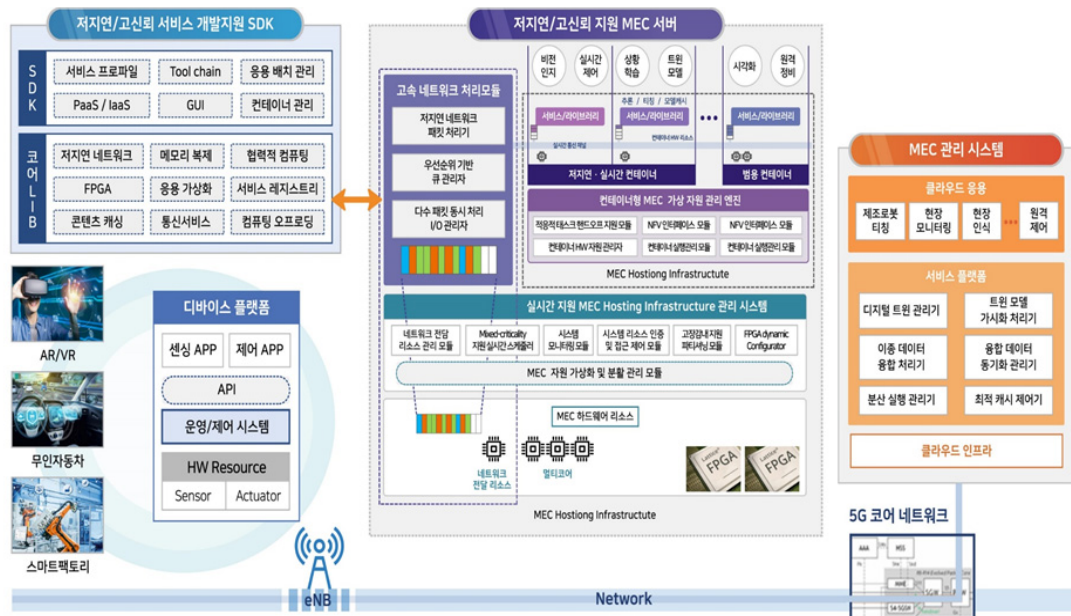
〈자료〉 한국전자통신연구원 자체 작성

[그림 6] 컴퓨팅 컨티뉴엄 상에서 GigaMEC/CompactEdge/EdgeCPS의 포지셔닝

1. GigaMEC

GigaMEC는 이동통신사의 5G 네트워크의 말단부인 기지국/집중국에 배치되어 초저지연 서비스를 제공할 수 있는 플랫폼으로서 범부처 GigaKorea 사업 과제인 “저지연 융합 서비스를 위한 모바일 에지 컴퓨팅 기술 개발” 과제를 통해 개발되었다. GigaMEC는 저지연/고신뢰 서비스를 지원하는 MEC 서버와 GigaMEC 플랫폼용 응용 서비스 개발을 지원하는 SDK, 그리고 통신사의 관리 및 오케스트레이션 시스템(MANO)과 연계하여 지역적으로 분산된 GigaMEC 플랫폼과 서비스를 관리할 수 있는 MEC 관리 시스템으로 구성된다[7]. [그림 7]은 GigaMEC의 구조를 보여준다.

GigaMEC 플랫폼은 유럽통신표준협회(ETSI)의 MEC를 준수하여 MEC 프레임워크 및 참조 아키텍처, MEC 플랫폼 관리, MEC 응용 관리를 위한 플랫폼 핵심 기술을 개발하였으며, 위치 기반 서비스, 트래픽 상태 기반 서비스 등 이동통신사의 네트워크로부터 획득할 수 있는 정보를 바탕으로 적응형 MEC 응용 서비스를 개발할 수 있는 관련 API 기술을 제공하고 있다.



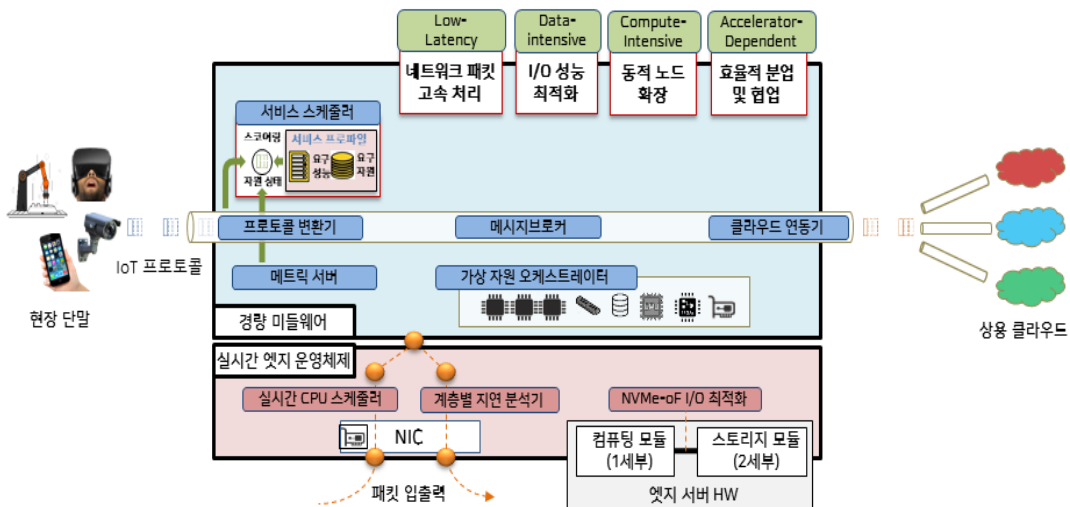
〈자료〉 한국전자통신연구원 자체 작성

[그림 7] GigaMEC의 구조

2. CompactEdge

CompactEdge는 과학기술정보통신부의 차세대 엣지 컴퓨팅 기술 개발 사업의 “엣지 서버 시스템 자원 관리 및 제어를 위한 경량 시스템 소프트웨어 기술 개발” 과제를 통해 개발 중인 시스템 소프트웨어 및 운영 플랫폼이다. CompactEdge는 열악한 산업현장에서도 운용이 가능하고 수요 기업의 요구에 따라 재구성이 가능한 맞춤형 엣지 플랫폼으로서 실제 산업 현장에서 다양한 엣지 서비스를 제공하는 범용 서버 장비의 소프트웨어로 활용할 수 있다. [그림 8]은 CompactEdge의 구조를 보여준다.

CompactEdge 플랫폼은 실시간 엣지 운영체제, 경량 미들웨어, 운영 플랫폼으로 구성된다. 실시간 엣지 운영체제는 현장의 엣지 디바이스의 서비스 요청에 해당하는 네트워크 패킷이 해당 요청을 처리할 수 있는 응용 프로세스에 도달하기까지의 시간을 수 밀리초 이내로 보장한다. 경량 엣지 미들웨어는 물리적 엣지 서버를 가상화하고, 엣지 응용 서비스를 효율적으로 운영하며, 이종의 IoT 프로토콜에 대한 처리와 다수의 상용 클라우드와의 클러스터 구성을 통한 동적 노드 확장 기능을 제공한다. CompactEdge 플랫폼은 현장에서 요구되는 다양한 엣지 서비스의 실행을 위해 필요한 연산·가속기 자원에 대한 실시간 관측을 바탕으로 최적의 자원 스케줄링을 통해 다중 서비스의 안정적 동시 실행을 위한 오케스트레이션



〈자료〉 한국전자통신연구원 자체 작성

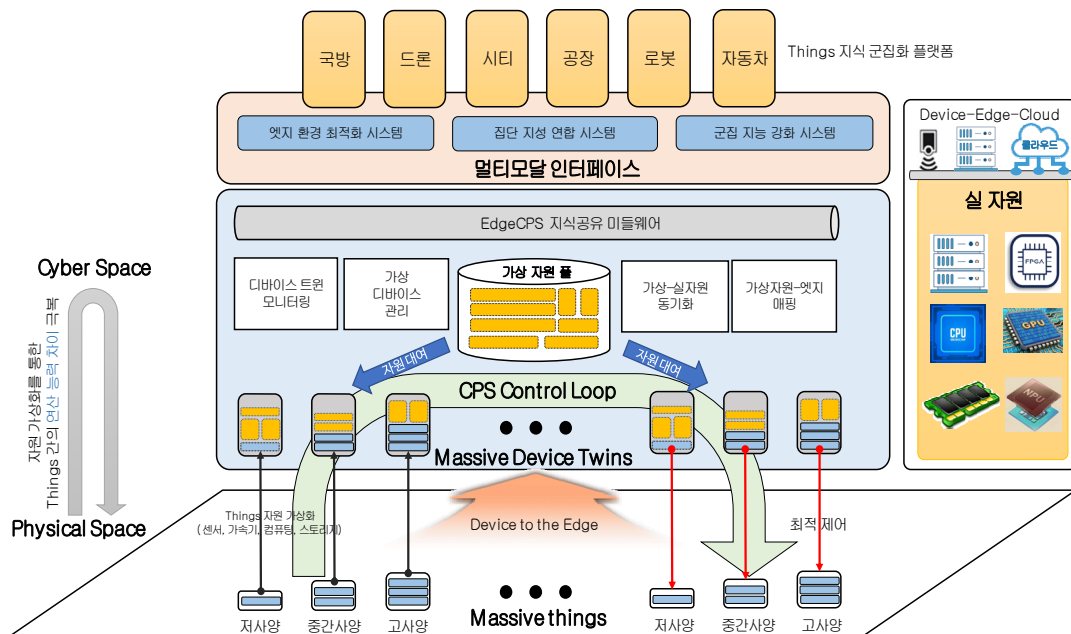
[그림 8] CompactEdge의 구조

기술을 특징으로 한다.

3. EdgeCPS

EdgeCPS는 ETRI의 원천기술 연구인 “인공지능 처리성능 한계를 극복하는 고성능 컴퓨팅 기술 연구” 과제를 통해 개발 중인 소프트웨어 플랫폼이다. 향후 2020년대 후반 6G 시대가 도래하면 70%의 IoT 단말이 셀룰러 기술을 사용할 것으로 예상되며, 단말의 수는 최대 100만 개/km²에 이를 것으로 예상된다[8]. 이와 같은 환경에서 현실 세계에 배치되는 IoT 디바이스의 사양은 [그림 9]와 같이 저사양부터 고사양까지 다양하게 존재할 것이며, 이를 연계한 수많은 혁신적인 미래형 서비스가 등장할 것으로 예상된다.

[그림 9]의 EdgeCPS는 이와 같은 초대규모 IoT 단말 간의 기능적/성능적 능력 차이를 극복하는 방법으로서 디바이스리스 컴퓨팅(Deviceless Computing) 기술을 제시한다. 디바이스리스 컴퓨팅을 통해 현실 세계의 IoT 단말이 요구하는 서비스 실행을 위한 자원과 기능을 담을 수 있는 ‘틀’인 디바이스 트윈(Device Twin)을 생성하고, 이를 실제 단말 및



〈자료〉 한국전자통신연구원 자체 작성

[그림 9] EdgeCPS 플랫폼의 구조

서버/클라우드 자원과 논리적으로 통합하여 “디바이스 규격에 대한 고려 없이” 원하는 서비스를 실행할 수 있다. 또한, EdgeCPS는 서비스 실행 시 GigaMEC, CompactEdge와 컴퓨팅 컨티뉴엄을 이루어 이음새 없는 자원 증강 및 공유를 지향하고 있다.

IV. 결론

컴퓨팅 컨티뉴엄은 임베디드-엣지-클라우드 컴퓨팅 등 다계층의 컴퓨팅 요소를 하나의 ‘연속체’로 보는 개념이며, 이를 바탕으로 컴퓨팅 요소 간 연계성을 고려한 응용 및 서비스 개발이나 과학적 실험을 효율적으로 하기 위한 기초 연구가 진행되고 있다. 시시각각 변화하는 동적 환경인 현실 세계는 동일한 서비스에 대해 운영 환경, 사물의 역량, 서비스 자원의 유휴 여부에 따라 제공할 수 있는 응답 속도와 체감 품질이 변화할 수밖에 없다.

GigaMEC와 CompactEdge, 그리고 EdgeCPS 플랫폼은 상호 연계가 가능한 자원 가상화 기반 서비스 플랫폼으로서, 단위 면적(km²) 당 서비스를 요청하는 IoT 단말의 수가 최대 100만 개에 달할 것으로 예상되는 6G 시대를 대비하여 임베디드, 엣지, MEC, 사설/상용 클라우드 컴퓨팅 등 컴퓨팅 컨티뉴엄의 자원을 효율적으로 관리하고, 대규모 지능형 IoT 단말에게 제공될 다양한 서비스를 위해 필요한 컴퓨팅 리소스를 최적의 위치에 정확히 배치하기 위한 전략을 제공한다. 이를 통해 무인/자율 시스템 제어, 메타버스, 빅데이터 분석, AI 응용, 텔레프레즌스/텔레오퍼레이션 등을 연계한 수많은 혁신적인 미래형 서비스 개발이 더욱 활발해질 것으로 기대되고 있다.

● 참고문헌

- [1] B. Gill, and D. Smith, “The Edge Completes the Cloud,” A Gartner Trend Insight Report, ID(G00360847), 2018, pp.1-26.
- [2] P. Beckman, J. Dongarra, N. Ferrier, G. Fox, T. Moore, D. Reed, and M. Beck, “Harnessing the Computing Continuum for Programming Our World,” Wiley, Fog Computing: Theory and Practice, 2020, pp.215-230.
- [3] D. Balouek-Thomert, I. Rodero, and M. Parashar, “Harnessing the Computing Continuum for Urgent Science,” ACM SIGMETRICS Performance Evaluation Review, Vol.48, No.2, 2020, pp.41-46.

- [4] S. Anderson, E. Deelman, M. Parashar, V. Pascucci, D. Petravick, and E. M. Rathje. "Report from the NSF Large Facilities Cyberinfrastructure Workshop," <https://facilitiesci.github.io/2017.html>, 2017.
- [5] D. Rosendo, P. Silva, M. Simonin, A. Costan, and G. Antoniu, "E2CLab: Exploring the Computing Continuum through Repeatable, Replicable and Reproducible Edge-to-Cloud Experiments," IEEE International Conference on Cluster Computing (CLUSTER), 2020, pp.176-186.
- [6] ARM, "The New Compute Continuum is Powering the Future of IoT," ARM research White Paper, 2019, pp.1-14.
- [7] 강성주, 전재호, 전인걸, "URLLC를 위한 MEC 기반의 디지털 트윈 기술," 한국통신학회지(정보와통신), Vol.35, No.11, pp.40-45.
- [8] Statista, "IoT and non-IoT connections worldwide 2010-2025," 2021.