

데이터 중심 초분산 자율 인프라 기술

Data-Centric Hyper-distributed Autonomous Infrastructure Technologies

김선미 (S.M. Kim, kimsunme@etri.re.kr)	네트워크연구본부 책임연구원
김상기 (S.K. Kim, kimsang@etri.re.kr)	네트워크연구본부 책임연구원
변성혁 (S.H. Byun, shbyun@etri.re.kr)	네트워크연구본부 책임연구원
정희영 (H.Y. Jung, hyjung@etri.re.kr)	네트워크연구본부 책임연구원
강세훈 (S.H. Kang, skang@etri.re.kr)	네트워크연구본부 책임연구원
임종철 (J.C. Lim, hektor@etri.re.kr)	네트워크연구본부 책임연구원
윤승현 (S.H. Yoon, shpyoon@etri.re.kr)	네트워크연구본부 책임연구원
신용윤 (Y.Y. Shin, uni2u@etri.re.kr)	네트워크연구본부 선임기술원

Various hyper-intelligent and ultra-realistic data-driven services are being increasingly developed with the goal of achieving a hyper-connected intelligent society. To sustain this trend, our research focuses on the integration and optimization of data-driven applications from several aspects such as delivery, storage, execution, and sharing of data and software, beyond the limitations of the existing network infrastructure. In this paper, we present important research issues of data-centric hyper-distributed autonomous infrastructure technologies.

* DOI: 10.22648/ETRI.2019.J.340102

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송 연구개발 사업의 일환으로 수행하였음[2017-0-00045, 초연결 지능 인프라 원천기술 연구개발].



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

최신 네트워크 기술 동향 특집

- I. 서론
- II. 자율 분산 프레임워크 기술
- III. 데이터 중심 전달 기술
- IV. 데이터 중심 분배 기술
- V. 데이터 중심 인-네트워크 프로세싱 기술
- VI. 공유/거래 마켓 플레이스 기술
- VII. 자율 제어/관리 기술
- VIII. 결론

1. 서론

미래에는 수많은 정보 기기들이 인터넷에 접속될 것이다. 2025년까지 500억 개 이상의 디바이스가 접속될 것이며, 이들과 연계된 새로운 응용 서비스들이 산업 환경은 물론 사람들의 일상생활을 바꿔놓을 것으로 예상된다. 이러한 인터넷 환경 변화의 중심에는 데이터가 있다. 2020년에는 매일 2.3조 기가바이트의 데이터가 유통될 것이며, 데이터의 양과 이 데이터를 활용한 서비스는 기하급수적으로 늘어나고 다양해질 것으로 예측되고 있다.

그런데 현재의 연결 중심의 IP 프로토콜 인터넷과 그 기반에서 이루어지는 클라우드 중심의 컴퓨팅 모델, 그리고 글로벌 클라우드 사업자 중심의 중앙집중적이며 폐쇄적인 서비스 인프라는 미래의 서비스(예, IoT) 환경이 창출하는 막대한 데이터의 양과 빠른 증가 속도, 급증하는 데이터 중심의 서비스 변화에 대처하기 불가능할 것으로 예측되며, 개방적인 혁신 생태계의 출현을 막는 부작용을 낳고 있다.

이러한 문제점들을 해결하고, 미래에 도래할 초연결 지능 사회의 중추 신경망으로서의 패러다임의 변화를 이끌 미래 인프라가 가져야 할 특성 및 기술들은 다음과 같다.

첫째, 시간/공간 민감(time/space-sensitive)한 데이터 중심 서비스의 데이터를 현장에서 즉시 저장/처리할 수 있도록 필요한 서비스 실행 환경(컴퓨팅/스토리지/네트워크 자원)을 동적으로 구성하여 실시간으로 제공하는 자율 분산 프레임워크 기술이 필요하다.

둘째, 대용량 멀티미디어 콘텐츠와 실시간 이벤트 등 다양한 형태의 데이터들을 서비스 요구사항에 맞게 최소 비용과 최적 연결 방식을 통하여 전달 및 분배할 수 있는 데이터 중심 네트워킹 기술이 필요하다.

셋째, 데이터 처리를 위한 응용소프트웨어들이 서비스 수요에 따라 온 디맨드로 실행이 가능하여야 하며,

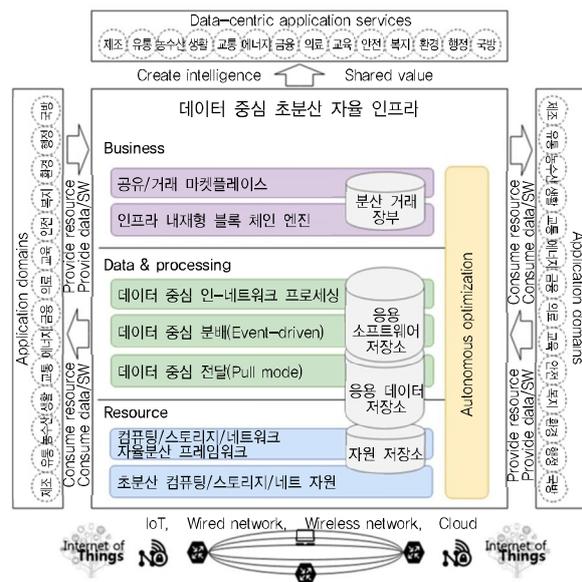
서비스가 효과적으로 처리될 수 있도록 네트워크 인프라 내 최적의 실행 위치를 결정하고 실행시킬 수 있는 데이터 중심 인-네트워크 프로세싱 기술이 필요하다.

넷째, 서비스 실행에 활용되는 자원, 데이터, 이를 처리하는 소프트웨어들은 더 많은 지능이 창출될 수 있도록 공유/거래되어야 하며, 이를 위한 안전하고 신뢰성 있는 인프라 내재 마켓플레이스 기술이 필요하다.

다섯째, 컴퓨팅/스토리지/네트워크 자원들은 스케일업/다운이 유연하여야 하며, 복잡한 서비스들과의 관계를 인간의 개입 없이 최적화하여 서비스 효율성과 인프라 효율성을 동시에 최대화할 수 있도록 자율 제어/관리 기술이 필요하다.

즉, 이를 위하여 네트워크 연결을 중심으로 하는 기존 인프라 범위를 넘어서, (그림 1)과 같이 데이터 중심 서비스들이 원활하게 수행되고 지속적으로 발전할 수 있도록 지원하는 데이터 중심 초분산 자율 인프라가 연구되고 있다.

본고에서는 데이터 중심 초분산 자율 인프라를 실현하기 위하여 필요한 기술들과 이와 관련한 최신 기술 동향에 대하여 분석한다.



(그림 1) 데이터 중심 초분산 자율 인프라

II. 자율 분산 프레임워크 기술

AR(Augmented Reality)/MR(Mixed Reality), 자율자 동차나 로봇, 디지털 트윈과 같은 데이터 중심의 초실감 초지능 서비스들은 현장에서 현장의 데이터를 중심으로 물리 세계와 사이버 세계를 연계하여 서비스를 제공한다. 이러한 서비스들은 사용자의 움직임에 따라 주변의 임의의 대상들과 일시적으로 데이터를 교환/공유(proximity 특성)하고, 즉시적 데이터 처리를 통해 의사결정(plug & process 특성)하여 제어한 후, 이동 및 상황변화에 따라 데이터 교환 대상을 변경하거나 처리 방식을 수시로 변경(variety & velocity 특성)하게 된다.

이러한 데이터 교환/처리를 위해서는 필요한 컴퓨팅/스토리지/네트워크가 실행 환경으로 구축되어야 하는데, 빈번하게 생성/소멸되고 변화하는 서비스들을 위한 실행 환경을 서비스 관리자가 그때그때 개별 구축하고 관리하는 것은 불가능하며, 서비스 발전을 위해서 해결되어야 하는 부분이다. 한편, 기술의 발전에 따라 에지 디바이스 및 네트워크 인프라 내 장비들은 놀라울 정도로 성능이 좋아지고 있지만, 고성능 기기를 사용하는 작업은 인터넷 검색, 게임 등이 일반적이어서 대부분 자원은 유향 상태이다. 자율 분산 프레임워크 기술은 이런 유향 상태인 분산된 자원들을 공유하여 서비스가 요구하는 실행 환경을 필요에 따라 동적으로 제공하는 기술이다.

자율 분산 프레임워크의 초기 흐름으로 글로벌 분산 컴퓨팅 및 분산 스토리지 기술을 생각해볼 수 있다. 탈중앙화(Decentralized)와 공유(Share)라는 새로운 가치 기준으로, 유향상태인 수많은 개인 컴퓨팅 및 스토리지 자원을 활용하여 저비용으로 고성능의 컴퓨팅과 스토리지를 제공하는 분산 컴퓨팅 및 분산 스토리지 기술 프로젝트들이 분산 환경의 P2P(Peer-to-Peer) 개념과 분산 신뢰성이 특징인 블록체인 기술을 활용하여 활발히 추진되고 있다.

분산 컴퓨팅 분야의 SONM 프로젝트는 완전 분산된

슈퍼컴퓨터를 제공하여 에지/포그 컴퓨팅 기반 범용 컴퓨팅 제공을 목표로 추진되고 있으며[1], iExec는 블록체인 기반 분산 클라우드 컴퓨팅 마켓을 제공하며 탈중앙화 어플리케이션(DApps: Decentralized Applications)에 대한 오픈체인 컴퓨팅을 목표로 추진되고 있다[2].

분산 스토리지 분야의 IPFS(Inter Planetary File System)는 P2P 컨텐츠 기반의 하이퍼미디어를 이용하여 전세계 네트워크에 연결된 모든 컴퓨터를 하나의 파일시스템으로 제공하며[3], Cassandra는 확장성과 가용성에 최적화된 분산 스토리지 기술, 데이터 센터까지 고려한 데이터 복제 정책으로 안정성 측면에서 장점을 가지고 있다[4].

이러한 새로운 분산 컴퓨팅 및 스토리지 기술들은 주변의 분산된 수많은 컴퓨팅 장비에서 유향 리소스를 활용하여 서비스에 필요한 컴퓨팅/스토리지/네트워크가 통합된 가상 실행 환경을 제공하는 분산 프레임워크의 기반 기술로서, 유향 자원 활용 및 사용에 대한 보상을 통한 공유 경제를 실현하고, 분산 특성에 따른 프라이버시 보호와 각종 장애 및 공격에 대해 SPF(Single Point Failure) 문제 해소 등 보안성 강화에 활용될 수 있다.

그러나 현재까지는 컴퓨팅과 스토리지 자원이 개별적인 기술로 개발 진행되고 있는 단계로서, 컴퓨팅/스토리지 및 네트워킹 분야까지 통합되어 빈번하게 생성/소멸되고 변화하는 서비스를 위한 가상 실행 환경을 언제 어디서든지 즉시적으로 제공할 수 있는 연구가 필요하다. 또한, 서비스의 특성에 따라 지역성(Locality) 보장, 지연(Latency) 보장, 프라이버시 보장, 이동성까지 보장할 수 있도록 운용자 개입 없이 자율적으로 서비스 실행 환경을 제공할 수 있는 자율 분산 프레임워크 기술에 대한 연구가 필요하다.

III. 데이터 중심 전달 기술

데이터 중심 초분산 자율 인프라의 네트워크 계층은

기존 호스트 연결중심의 IP 네트워크와 달리 데이터 자체를 얻어 오는 것을 목적으로 하는 ICN(Information-Centric Networking) 기술을 기반으로 한다. ICN은 데이터의 위치 정보에 상관없이 오직 이름을 통해 데이터를 교환하며, 안전성, 이동성 지원 등 데이터 중심 서비스를 위한 많은 장점을 가지고 있다.

ICN 개념은 2001년 TRIAD에서 처음 제시되었으며 [5], 2006년 Van Jacobson의 CCN(Content-Centric Networking) 연구로 본격화되었다[6]. 2006년부터 미국 NSF가 FIND 프로그램으로 지원한 수십 개의 미래인터넷 프로젝트는 CCN에 기반한 NDN(Named Data Networking)[7], Mobility First, XIA(eXpressive Internet Architecture)의 3개의 FIA-NP 프로젝트로 수렴되어 이어졌는데, NDN과 Mobility First는 ICN 구조이며, XIA도 ICN을 수용하고 있어 미래인터넷 연구가 ICN으로 수렴되었다고 할 수 있다. 유럽에서도 EU의 PSIRP(Publish-Subscribe Internet Routing Paradigm)와 후속 PURSUIT(Publish-Subscribe Internet Technology) 및 POINT(iP Over ICN-the betTer IP), NetInf(Network of Information) 및 SAIL(Scalable and Adaptive Internet Solutions) 등을 필두로 매우 다양한 구조를 갖는 ICN 과제들이 수행되어 왔다[8].

IRTF(Internet Research Task Force)의 ICNRG(Information-Centric Networking Research Group)에서는 다양한 ICN 연구 기반 표준 개념정립과 서비스 시나리오, 연구 이슈 등을 논의하고 있다[9].

최근 Cisco는 PARC로부터 CCN S/W 및 IPR를 인수하여 CICN(Community ICN) 오픈소스 프로젝트를 운영하고, IPv6 헤더구조를 갖는 변형된 CCN인 hICN(Hybrid ICN)을 개발하여 Verizon과 함께 IPv6 망에서 순차적으로 ICN을 도입할 수 있는 방안을 연구하고 있다[10].

제시된 ICN 구조들은 세부 사항은 차이가 있지만 다음과 같은 공통점을 특징으로 한다.

- 데이터 이름 기반 통신: 응용이 교환하는 데이터 자체에 부여한 이름을 기반으로 라우팅/포워딩을 수행한다.
- 내재된 보안 기능: 데이터 생산자에게 할당한 공개키 기반으로 데이터 생산자 및 데이터 자체의 변조 여부를 데이터 수요자가 검증할 수 있는 signature를 데이터에 추가하여, 데이터 무결성 및 생산자 인증을 네트워크 계층에서 보장한다.
- 인-네트워크 캐싱: 전달되는 데이터를 네트워크 내에서 자동 캐싱하여, 수요자가 요구한 데이터를 생산자가 아닌 인-네트워크 캐시로부터 보내줄 수 있도록 함으로써, 콘텐츠의 효율적 전달 및 멀티캐스트, 애니캐스트의 효율적 수행을 가능케 한다.

2018년 현재 다양한 ICN 구조 연구 중에서 NDN/CCN이 유력한 구조로서 가장 많은 지지를 받고 있으며, NSF/Intel 공동후원의 ICN-WEN(ICN-Wireless Edge Network) 프로그램의 3개 프로젝트 및 EU/Japan 공동프로젝트인 ICN2020 등을 포함한 다양한 ICN 프로젝트가 NDN/CCN 기반으로 진행 중이다.

그러나 NDN/CCN으로 대표되는 ICN이 인터넷에 도입되기 위해서는 네임 기반 라우팅 및 포워딩 성능문제라는 커다란 이슈를 해결해야 한다. 현재 $O(10^8)$ 인 IPv4 호스트 수에 비해 구글이 인덱싱한 URL 수는 $O(1.012^t)$ 이어서 ICN에서 다룰 데이터 수는 현재보다 10,000배 이상 많다고 할 수 있다[11]. 따라서 데이터 이름 기반의 라우팅을 인터넷 스케일로 하기 위해서는 라우팅 엔트리 수를 적정 수준으로 줄이기 위한 구조적 연구가 필요하며, NDN/CCN의 계층적 데이터 이름은 고정길이의 IP 주소와는 달리 이름 길이에 제한이 없어서 데이터 이름기반 포워딩 고속화를 위한 연구가 필요하다. 또한, NDN/CCN에서는 라우터의 인터레스트 패킷 애그리게이션 및 인-네트워크 캐시 기능 때문에

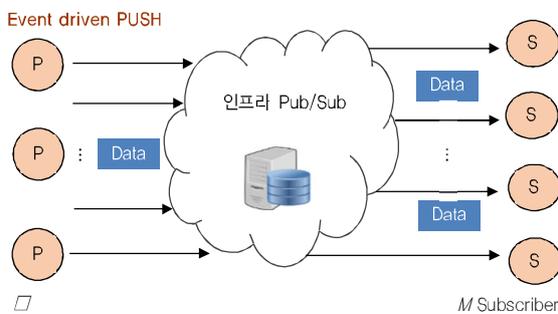
DDoS 공격을 원천적으로 차단할 수 있지만, 인터레스트 상태정보를 가지는 라우터를 공격할 수 있기에 이를 위한 네트워크 보호 방안이 필요하고, 데이터 암호화 및 권한 검증을 위한 확장성 있는 키 분배 연구가 필요하다.

IV. 데이터 중심 분배 기술

최근 데이터 기반 사물 인터넷 자동화 서비스, 소셜 미디어와 같은 이벤트 기반 응용이 활성화되고 있다. 또한 빅데이터나 대규모 시스템 로그 처리, 마이크로 서비스 등에서도 Pub/Sub 통신 방식을 기본적으로 요구하고 있다. 이를 지원하기 위해서는 Pull 방식 데이터 전달뿐만 아니라, (그림 2)와 같은 이벤트 기반 Push 방식의 Pub/Sub(Publish/Subscribe) 통신 기술들에 대한 수요가 늘고 있다는 것을 주목할 필요가 있다.

현재 Pub/Sub 시스템에서 사용하고 있는 TCP/IP 기술이 송신자 주도 네트워킹 기술인 것에 반해, NDN/CCN과 같은 데이터 중심 네트워킹 기술은 기본적으로 데이터 수요자의 요청에 의해 데이터 생산자가 데이터를 제공하는 수요자 주도 Pull 방식이다. 이는 이벤트 기반 서비스를 제공하기에 한계가 예상되며, 따라서 발생한 이벤트를 기반으로 데이터 생산자가 Push 방식으로 데이터를 전달하는 기술들이 추가로 연구되고 있다.

현재 NDN에서의 가장 대표적인 Pub/Sub 기술은 PSync라고 할 수 있다[12]. PSync는 데이터 전체에 대한 싱크만을 지원하는 기존의 NDN Sync 프로토콜을



(그림 2) 이벤트 구동형 데이터 Pub/Sub 분배 기술

Pub/Sub 서비스 지원을 위해 확장한 것으로서 업데이트되는 데이터에 대한 부분적인 싱크를 지원할 수 있도록 재설계한 기술이다. PSync에서 수요자는 자신의 구독 목록을 Bloom필터 형태로 표현한다. 데이터 생산자는 발행 데이터 상태를 IBF(Invertible Bloom Filter)로 관리하며, 이전 IBF와 현재의 IBF 간의 비교를 통해 새로운 데이터의 생성을 확인하는 구조를 가진다. 데이터 수요자는 데이터 생산자에게 주기적인 싱크 패킷을 통해 자신의 구독 목록을 보낸다. 구독 목록을 수신한 데이터 생산자는 자신의 IBF를 검사하여 수신한 구독 목록 중 새롭게 생성된 데이터가 있는 경우 해당 데이터의 네임을 Data 패킷을 통해 알려준다. 이후 응답된 네임을 기반으로 일반적인 NDN 요청/응답 절차를 통해 데이터가 전달된다.

NDN에서 Pub/Sub 기술로 최근에 제안된 또 다른 기술은 HoPP(HoP and Pull)이다[13]. HoPP은 HoPP 프로토콜을 지원하는 네트워크 노드들로 구성되며 미리 설정된 CP(Content Proxy) 노드가 생산자와 수요자 간의 라데부 포인트 역할을 수행한다. 네트워크를 구성하는 노드들은 CP 방향으로 DODAG(Destination Oriented Directed Acyclic Graph)를 구성한다. 생산자는 이벤트가 발생하였을 때 이를 CP 방향으로 광고하며 광고를 수신한 각 노드는 이를 CP 방향으로 전달함과 동시에 인터레스트를 이용하여 발행된 이벤트 데이터를 요청하여 캐시한다. 수요자는 자신이 필요한 이벤트 데이터에 대해 인터레스트 패킷을 통해 CP 방향으로 구독을 광고한다. 만일 중간의 노드가 구독 요청에 해당하는 데이터가 캐시되어 있는 경우 바로 응답하고, 아니면 이 요청은 CP까지 전달되어 CP가 요청에 최종적으로 응답한다.

전술한 PSync나 HoPP과 같은 기존 기술은 NDN의 요청/응답 방식을 그대로 따르고 있으므로 이벤트 기반 데이터 전달에는 비효율적이다. 또한, 종단 노드 간 방식이거나 LoWPAN과 같은 특정한 영역에서 사용을 위

해 설계되어 인프라 규모 확장성과 성능, QoS 등 다양한 서비스 요구사항 수용에 한계를 갖고 있다. 따라서 이벤트 데이터를 기반으로 하는 응용 서비스들이 네트워크 접속 위치에 제약 없이 다양한 데이터를 즉시 전달 받을 수 있도록 규모 확장성과 기능의 유연성을 내재한 네트워크 계층의 데이터 분배 구조 및 기술에 대한 연구가 추진되고 있다.

V. 데이터 중심 인-네트워크 프로세싱 기술

NDN/CCN 기술이 네트워크상에서 미리 만들어진 데이터 전달에 주목하고 있다면, 데이터 중심 인-네트워크 프로세싱 기술은 네트워크가 단순 데이터 전달 기능뿐만 아니라 네트워크 내의 적절한 위치에서 데이터에 대해 필요한 연산을 수행하여 그 결과를 전달하는 기능을 수행하도록 하는 기술이다.

NFN은 NDN 기반의 인-네트워크 프로세싱 개념을 처음으로 소개한 기술로서 ‘사용자는 데이터가 아니라 결과를 원한다’는 개념에서 시작한다[14]. NFN은 원하는 데이터 이름을 직접적으로 사용하는 대신 원하는 데이터를 어떻게 얻을 수 있는지에 대한 일종의 데이터 처리를 위한 레시피를 네트워크에 요청하여 그 결과를 획득하게 한다. 이 레시피는 필요한 함수와 다른 데이터에 대한 참조가 가능하도록 함수 이름과 파라미터 이름의 조합으로 구성되어 λ -expression 형식으로 기술된다. NFN에서 네트워크는 필요한 데이터를 확보하고, 적절한 위치를 선택하여 함수를 실행한 후 그 결과를 클라이언트에게 반환하는 역할을 수행한다.

인-네트워크 프로세싱의 또 다른 기술로서 NFaaS 기술이 있다[15]. NFaaS는 NDN 기반의 인-네트워크 프로세싱을 서버리스 컴퓨팅 개념과 결합하였다. NFaaS 기술에서는 보다 다양한 응용을 지원하기 위해 경량의 가상 머신인 unikernel을 사용하였으며, 함수를 어느 위치에서 실행할 것인지를 결정하기 위해 각 노드에서는

함수의 인기도와 응용의 타입(지연 민감형 혹은 네트워크 대역 소모형)에 따라 unikernel score를 계산하여 함수의 다운로드 여부를 결정한다. 그 결과로서 지연 민감형 응용은 클라이언트와 가까운 네트워크 에지에서 실행되고, 네트워크 대역 소모형 응용은 네트워크 코어에 근접한 위치에서 실행되도록 유도된다. NFaaS는 응용 타입과 함수 인기도에 따라 함수 실행 위치가 네트워크에 골고루 분산되는 장점을 지니고 있다.

NDN과 인-네트워크 프로세싱을 결합한 또 다른 연구로는 RICE가 있다[16]. RICE에서는 RMI와 같은 원격 함수 호출을 지원하기 위해 사용자 인증 및 권한 인증 방법을 제공하고, 원격 함수에 대한 대용량의 데이터 전달 방법에 대해 정의한다. 또한, 기존 NDN 방식에서 함수 실행 시간이 긴 경우, PIT(Pending Interest Table) 엔트리 만료를 방지하기 위한 반복 인터레스트나 long-lived 인터레스트로 네트워크 부하가 증가하는 단점이 있다. RICE에서는 이러한 문제를 해결하기 위해 함수 실행 요청에 사용되는 이름(Function Name)과 그 결과를 반환받기 위해 사용되는 이름(Thunk Name)을 분리하여 해결하였다.

데이터 중심 인-네트워크 프로세싱에 있어 가장 중요한 이슈 중 하나는 응용의 데이터 프로세싱 요청을 받아 최적의 실행 위치를 결정하고 전달하기까지의 과정을 어떻게 오케스트레이션 하는가의 문제이다. 특히 최적의 실행 위치를 결정하는 문제는 단순히 네트워크의 자원 상황이나 응용의 특성만으로 결정하는 것이 아니라 입력 데이터의 위치, 서비스 요구 사항, 함수 실행을 위해 필요한 자원 요구량 및 제한 조건 등을 종합적으로 고려하는 방향의 연구를 필요로 한다. 또한, ICN 기반 환경에서 인-네트워크 프로세싱을 수행함에 있어 사용자 인증 및 함수 사용 권한, 데이터 사용 권한 등을 체크할 수 있는 효율적 방법과 이를 데이터 공유/거래 마켓플레이스와 연계할 수 있는 방안에 대한 연구가 추진되고 있다.

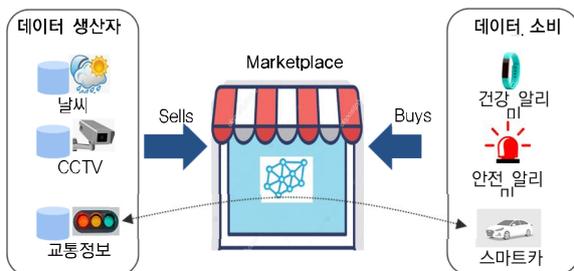
VI. 공유/거래 마켓플레이스 기술

데이터의 활용을 극대화하기 위해서는 데이터의 생산자와 소비자를 잘 연결해 주는 플랫폼이 필수적인데, 현재 데이터 활용 모델은 고립된 silo 형태의 모델로써 데이터 사업자가 데이터를 수집한 후 이를 가공하여 데이터 소비자에게 판매하는 형태의 모델이 대부분이다. 또한, 데이터를 소비하는 주체는 주로 사람이며, 데이터를 이용하기 위한 사전 계약이 이루어져 있어야 한다.

그러나 미래사회에서는 다양한 형태의 수요가 많은 데이터 생산자와 소비자가 등장할 것이고, 이러한 생산자와 소비자는 사람이 아닌 자율적인 디지털 기기일 수 있으며, 데이터의 이용 측면도 쌓여있는 데이터를 가공하여 이용하는 것뿐만 아니라 필요에 따라 실시간 데이터를 직접 이용하는 형태로 진화할 것이다.

즉, 현재의 특정 영역의 데이터를 가진 데이터 플랫폼 사업자 구조에서, 미래에는 (그림 3)과 같이 공개적이고 안전한 마켓플레이스를 통해 데이터 사업자라는 중간 매체 없이도 데이터 생산자와 소비자가 필요할 때 데이터를 즉시적으로(on the fly) 거래할 수 있는 구조로 진화할 것이다.

이러한 추세에 따라 최근 블록체인 기술을 결합한 다양한 데이터 마켓플레이스들이 등장하고 있다. Datum은 사용자가 인터넷 기반의 서비스(예: 페이스북, 트위터 등)를 사용하면서 생성되는 데이터를 거래할 수 있는 시스템이다[17]. 사용자가 생성한 데이터는 인터넷상에 분산된 안전한 스토리지나 사용자의 장치에 저장되고,



(그림 3) 데이터 마켓플레이스

데이터에 대한 판매 계약이 블록체인의 스마트 컨트랙트에 의해 명시되어, 데이터 구매자가 블록체인상에서 해당 데이터를 찾고 데이터 구매를 요청하는 트랜잭션을 스마트 컨트랙트에 보냄으로써 자동적으로 거래가 이루어질 수 있도록 한다. PikcioChain도 Datum과 유사한 방식으로 동작하지만, 기관 데이터 거래가 가능하도록 하였으며, 프라이버시를 강화하여 데이터 저장은 데이터 소유주의 기기에서만 이루어지도록 하였다[18]. 더불어 별도의 검증 노드를 통해 데이터 검증을 수행하고, 그 검증 결과 자체도 판매하는 것이 가능하다는 특징을 가지고 있다.

IOTA는 IoT 디바이스에서 생성되는 데이터를 사고팔 수 있는 개방형 마켓플레이스에 대한 파일럿 프로젝트를 수행하고 있다[19]. IOTA에서는 센서들에 대한 정보를 블록체인에 암호화하여 저장할 수 있는데, 마켓플레이스 기능을 통해 지도에서 관심 있는 센서를 선택하고, 그 센서의 데이터에 대한 지불(payment) 트랜잭션을 블록체인에 전송함으로써 구매가 이루어지며, 암호화된 데이터를 해독할 수 있는 키가 사용자에게 제공되어 블록체인에 저장되어 있는 데이터를 읽을 수 있게 된다. Databroker DAO도 IoT 데이터에 대한 네트워크 사업자 기반 마켓플레이스를 제공한다[20]. 센서의 소유주는 네트워크 사업자가 제공하는 네트워크에 연결하여 센서에서 발생하는 데이터를 데이터 silo에 저장한다고 가정한다. 데이터 매매의 자동 처리를 위해 DAO 스마트 컨트랙트를 사용하며, 데이터 소비자가 API 게이트웨이를 통하여 데이터를 이용하고자 할 때의 접근 제어 기능을 수행한다.

특정 도메인에 특화된 데이터를 대상으로 하는 데이터 마켓플레이스로서는 의료 데이터용 NanoVision, AI 데이터용 Ocean과 같은 공유 프레임워크 기술도 등장하고 있다[21], [22].

데이터 공유-거래의 주요 이슈인 프라이버시 보장을 위해 GDPR과 같은 보호법뿐 아니라 개인정보 보호 기

반 저장/연산을 위한 MPC(Multi Party Computing), 동형 암호(Homomorphic Cryptography), Zero Knowledge Proof 기술 등이 연구되고 있다. Enigma는 데이터를 여러 조각으로 나누어 오픈체인 형태의 스토리지 저장하고, MPC를 통해 데이터를 처리함으로써 보안성이 강화된 데이터 마켓플레이스를 구현하였다[23].

데이터의 중요성이 증가하면서 활용을 높이기 위한 마켓플레이스 기술에 대한 지속적인 연구가 요구되고 있다. 즉, 방대한 데이터 속에서 의미적 문맥을 파악하여 사람뿐만 아니라 자율 기기와의 효율적 데이터 교환을 위한 시맨틱 기술, 사람/기기 간 실시간의 자동화된 거래 및 접근 제어를 위한 블록체인의 스마트 컨트랙트 기술, 모든 응용도메인이 보편적으로 접근하여 데이터를 공유할 수 있는 인프라 내재형 마켓플레이스에서 분산 신뢰성과 고성능을 동시 확보할 수 있는 기술에 대한 연구가 추진되고 있다.

Ⅶ. 자율 제어/관리 기술

네트워킹 분야의 자율 제어/관리는 2001년 IBM에서 제안한 자율컴퓨팅의 self-CHOP(Self-Configuration, Healing, Optimization, Protection)기반 참조 모델인 MAPE-K(Monitor, Analyze, Plan, Execute, and Knowledge)를 기반으로 시작되었다[24]. 초창기 연구는 2000년대 중반에 EU FP(European Union Framework Programme)에 의하여 주도되었으나 2010년 이후에는 클라우드, SDN/NFV, AI/ML 기술의 발전으로 자율 네트워킹에 예측 및 최적 제어를 수행하는 연구 사례가 늘어나고 있다. 2017년까지 진행된 5G PPP phase 1에서는 CogNet, SELFNET과 같은 프로젝트에서 집중적으로 자율 제어 및 관리에 대한 연구가 진행되었다. 네트워킹 분야의 권위 있는 학회 중 하나인 SIGCOMM 2018의 논문 동향을 살펴보면, 인프라 기술 대부분의 영역에서 AI 기반 자율 제어/관리 연구가 진행되고 있

음을 살펴볼 수 있다. 발표된 논문을 AI/ML 기술 분류에 따라 매핑해보면 다음과 같다.

- Supervised Learning: 플로우 레벨 네트워크 모니터링, 콘텐츠 캐싱 예측, 네트워크 이벤트 분류, 온라인 비정상 탐지, 인터넷 구조 특성 추정(네트워크 지연 추론), 프로토콜 자동 설계 등
- Reinforcement Learning: 온라인 트래픽 최적화(스케줄링, 로드 밸런싱), 비디오 ABR 자동튜닝, 데이터 센터 토폴로지 동적제어, 무선환경 최적화, NDN 포워딩 정책, TCP 혼잡제어 등
- Unsupervised Learning: 암호화된 플로우 식별, 모바일 노드 이동 예측, 트래픽 데이터로부터 QoS 분포 예측, log 데이터로부터 비정상 상황 탐지, 데이터센터 서버 그룹핑 등

또한, 2018년부터는 연구뿐만 아니라 산업계, 표준화 단체에서도 인프라 관리에 있어서 AI/ML이 접목된 자율 제어/관리에 대한 관심이 급격히 높아지는 것으로 파악된다. 시스코는 암호화된 트래픽에서 악성코드를 탐지하는데 AI/ML 기술을 적용하고 있으며, 또 다른 사례로 자동으로 액세스 목록을 생성하는 등에 활용하고 있다[25]. 국내의 경우 KT는 2018년 인공지능 기반의 네트워크 운용 플랫폼 ‘뉴로플로우’ 개발을 완료/적용한다고 발표한바 있다. 이는 방대한 데이터를 딥러닝을 이용하여 분석하여 네트워크 운용 의사결정을 지원하는 플랫폼이다[26]. 삼성전자는 인공지능을 이용한 root cause analysis 등의 기술을 보유한 Zhilabs 기업을 인수하여 5G 인프라 운용에 인공지능 적용을 시사하고 있다.

표준 동향으로 IETF ANIMA WG과 IRTF NMRG에서는 autonomic networking을 위한 표준화 및 참조 모델에 대한 논의를 일찍부터 시작한 상태이다[27], [28]. ETSI에서는 2018년 ZSM(Zero touch network & Service

Management) ISG community 활동을 시작하여 NFV와 SDN 환경에서 완전 자동화된 관리 및 운용에 관한 논의를 시작하였다[29].

이와 같이 인프라에 대한 자율 제어 연구는 다양하게 추진되고 있다. 그러나 점점 더 복잡 다양해지는 서비스들의 지속적 발전이 가능한 인프라를 제어하기 위해서는 스스로 학습하여 적응적이고 선제적으로 대응이 가능한 자율 제어/관리 기술이 필요하다. 데이터 중심 초분산 자율 인프라는 수많은 이종 자원과 다중 계층 자원(물리 자원, 논리 자원, 소프트웨어 등)을 동시에 고려해야 하고, 외부 상황 변화(수요 변화, 이동성, 자원 변화)에 선제적으로 대응하면서 서비스 요구사항을 만족하고 동시에 인프라 비용 최적화를 달성해야 하는 등 그 제어기가 매우 복잡할 것으로 예상된다. 따라서 AI/ML 기술을 바탕으로 하여 운용자의 개입 없이 자율적으로 운용되고 미래 예측을 내재하여 선제적으로 제어 및 관리가 가능한 기술 연구를 추진하고 있다.

VIII. 결론

본고에서는 연결을 중심으로 하는 기존 네트워크 인프라의 한계를 넘어서, 초실감 초지능의 데이터 중심의 응용 서비스들이 가장 최적으로 수행될 수 있도록 컴퓨팅/스토리지/네트워크 자원 통합 제공 및 응용 데이터/소프트웨어의 저장, 실행 및 공유를 자율 최적화하는 초분산 자율 인프라 구조 및 기술들을 정리하였다. 살펴본 바와 같이 기존 네트워크의 영역과 수준을 탈피하여 미래 서비스의 지속적 발전을 뒷받침할 수 있는 패러다임 전환이 시도되고 있다. 본 연구는 전반적인 구조 밑그림이 그려진 단계로서, 향후 세부 핵심 기술들에 대한 본격적인 연구 개발과 구조적 기술적 검증이 필요한 시점이라고 할 수 있다.

데이터 중심 초연결 지능 사회에서 신산업/신서비스 활성화를 위한 국가 지능화 기반을 마련하기 위해서는

데이터 중심 초분산 자율 인프라로의 혁신이 필수적이라고 판단된다. 지속적인 검증과 보완으로 혁신을 성공적으로 완성하기 위해서는 연구소뿐만 아니라 학계와 산업계의 협력이 요구된다.

약어 정리

AI	Artificial Intelligence
CCN	Content-Centric Networking
CDN	Content Distribution Network
CP	Content Proxy
DODAG	Destination Oriented Directed Acyclic Graph
IBF	Invertible Bloom Filter
IPFS	Inter Planetary File System
ICN	Information-Centric Networking
LoWPAN	Low Power Wireless Personal Area Network
ML	Machine Learning
NDN	Named Data Networking
NFN	Named Function Networking
NFaaS	Named Function as a Service
P2P	Peer to Peer
PIT	Pending Interest Table
Pub/Sub	Publish/Subscribe
RN	Rendezvous Node
ZSM	Zero touch network & Service Management

참고문헌

- [1] SONM, "SONM Project Documentations," 2017. Available: <https://docs.sonm.com/>
- [2] iExec, "iExec Project Documentations," 2016. Available: <https://docs.iex.ec/>
- [3] IPFS, "Protocol Labs: IPFS Documentations," 2017. Available: <https://docs.ipfs.io/>
- [4] CASSANDRA, "What is Cassandra," Apache Software Foundation, 2016. Available: <http://cassandra.apache.org/>
- [5] Stanford University TRIAD Project. Available: <http://www-dsg.stanford.edu/triad/>
- [6] V. Jacobson, "A New Way to Look at Networking," Google Tech Talk, Aug. 8. 2006.
- [7] NSF Named Data Networking Project. Available: <http://www.named-data.net/>

- [8] B. Ahlgren et al., "A Survey of Information-Centric Networking," *IEEE Commun. Mag.*, vol. 50, no. 7, July 2012, pp. 26-36.
- [9] IRTF, "IRTF ICN Research Group," Available: <https://trac.ietf.org/trac/irtf/wiki/icnrg>
- [10] D. Kurschner, "Cisco Announces Important Steps Toward Adoption of Information-Centric Networking," Cisco blog, Feb. 23, 2017. Available: <https://blogs.cisco.com/sp/cisco-announces-important-steps-toward-adoption-of-information-centric-networking>
- [11] M. Bari et al., "A Survey of Naming and Routing in Information-Centric Networks," *IEEE Commun. Mag.*, vol. 50, no. 12, Dec. 2012, pp. 44-53.
- [12] M. Zhang, V. Lehman, and L. Wang, "Scalable Name-Based Data Synchronization For Named Data Networking," in *INFOCOM 2017-IEEE Conf. Comput. Commun.*, Atlanta, GA, USA, May 1-5, 2017, pp. 1-9.
- [13] C. Gündoğan et al., "HoPP: Robust and Resilient Publish-Subscribe for an Information-Centric Internet of Things," arXiv preprint, arXiv: 1801.03890, 2018.
- [14] V. Jacobson et al., "Networking Named Content," in *Proc. Int. Conf. Emerg. Netw. Experiments Technol.*, Rome, Italy, Dec. 1-4, 2009, pp. 1-12.
- [15] M. Sifalakis et al., "An Information Centric Network for Computing the Distribution of Computations," *Proc. Int. Conf. Inform.-Centric Netw.*, Paris, France, Sept, 24-26, 2014, pp. 137-146.
- [16] M. Król et al., "RICE: Remote Method Invocation in ICN," *Proc. ACM Conf. Inform. Centric Netw.*, Boston, MA, USA, Sept. 21-23, 2018, pp. 1-11.
- [17] R. Haenni, "Datum Network the Decentralized Data Marketplace," DATUM Whitepaper V15, 2017.
- [18] PIKCIOCHAIN, "PIKCIOCHAIN the Personal Data Chain," PIKCIOCHAIN Whitepaper v2.0, 2018.
- [19] iota, <https://www.iota.org/>.
- [20] DAO, "Databroker Dao," Databroker DAO Whitepaper, 2018.
- [21] Nanovision, <https://nanovision.com/>.
- [22] Ocean Protocol Foundation, "Ocean Protocol: A Decentralized Substrate for AI Data & Services Technical Whitepaper," Ocean White Paper Version 0.9.3, 2018.
- [23] Enigma, <https://enigma.co/>.
- [24] 신승재 외, "자율네트워킹 연구동향," 전자통신동향분석, 제 32 권 제1호, 2017. 2.1, pp. 25-34.
- [25] M. Cooney, "'AI와 머신러닝으로 네트워크 지형도 바꾼다' 시스코의 미래 인프라 전략," IT WORLD, 2018.9.20.
- [26] 김윤희, "KT, AI 기반 네트워크 관제 플랫폼 개발," ZDNet Korea, 2018.1.17.
- [27] IETF Datatracker, "Autonomic Networking Integrated Model and Approach (anima)," 2018. Available: <https://datatracker.ietf.org/wg/anima/about/>
- [28] ETSI Datatracker, "Network Management Research Group," Available: <https://irtf.org/nmrg>.
- [29] ETSI, "Zero Touch Network & Service Management," Available: <https://www.etsi.org/technologies-clusters/technologies/zero-touch-network-service-management>