

# 과학빅데이터 고속전송을 위한 유무선 네트워킹 적용방안 연구

석우진, 김기욱, 궤재승

한국과학기술정보연구원, 과학기술연구망센터

wjseok@kisti.re.kr, wowook@kisti.re.kr, jskwak@kisti.re.kr

## A Study on Wired and Wireless Networking for Science Big Data Transfer

Woojin Seok, Kiwook Kim, Jaiseung Kwak

KISTI, KREONET center

### 요약

본 논문에서는 과학빅데이터를 전송하는 방법으로 과학실험장비에서 발생하는 빅데이터를 유선네트워크에서의 고속전송하는 기술 방안과 과학실험장비 내부에서의 데이터를 송수신하기 위한 근거리 고속 무선네트워크 기술에 대한 적용기술을 살펴보고자 한다. 이러한 유무선 네트워킹 기술이 해결하고자 하는 기술적 요소 등을 살펴보고 적용가능한 기술방안을 제안하고자 한다.

### 1. 서론

과학의 발전은 실험과학, 이론과학을 거쳐 데이터를 기반으로 계산을 수행하는 데이터 과학으로 발전하고 있다. 이러한 현상은 고에너지물리 연구, 핵융합에너지 연구, 천문학 연구, 염기서열 연구 등의 대다수의 과학 분야에서 나타나고 있다. 고에너지물리 연구의 경우, 2007년부터 스위스입자물리연구소(CERN)의 입자가속기로부터 발생하는 연간 12~14 Petabyte 규모의 실험데이터를 전송하고 있다.

핵융합 연구분야에서는 K-STAR 핵융합 실험로에서 나오는 진단 데이터 분석, ITER 국제 핵융합 실험로의 원거리 진단 데이터 등의 빅데이터가 발생하고 있으며, 천문학 분야는 가상천문대(국내 여러 지역에 흩어져 있는 천과 망원경의 신호를 통합하여 거대한 천과망원경으로 관측하는 것과 같은 효과를 내는 가상천문대)를 구축하여 여러 지역의 천과망원경에서 수집된 천문 데이터를 실시간으로 데이터 통합 기지로 전송이 요구되고 있다. 또한, 기후 변화 관찰 및 예측을 위한 컴퓨터 시뮬레이션과 기상 장비로부터 얻어진 기상 데이터들의 분석하는 연구 분야로써, 기상예측의 정밀도를 높이기 위해서 시·공간 격자 해상도를 점점 더 높일수록 더 많은 데이터가 생성되고 있다. 염기서열 데이터, 단백질의 아미노산 서열 데이터, 단백질 구조와 관련된 X-선 회절 데이터 등 바이오 데이터와 임상 이미지 데이터의 저장 및 열람, 그리고 이미지 데이터 및 의료 데이터 등이 발생하고 있다.

이러한 과학빅데이터를 전송하는 방법으로 백본에서의 고속전송으로 유선네트워크 기술 등이 적용되고 있으며, 또한 최근 실험장비에서 발생하는 데이터의 수집에 무선 네트워크 기술 등이 추

가로 요구되고 있다. 본 논문에서는 과학빅데이터의 고속전송을 위한 유선 네트워크 기술과 실험장비에서의 데이터를 송수신하기 위한 무선네트워크 기술에 대하여 기술적 문제점과 해결방안에 대해서 제안하고자 한다.

### 2. 유선 고속전송을 위한 ScienceDMZ 전송 방안

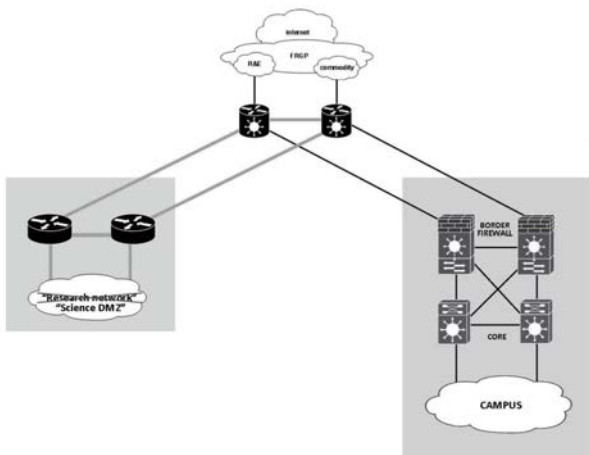
과학 실험장비에서 발생하는 데이터는 테라급 혹은 페타급의 큰 데이터 스케일을 보여주고 있다. 이러한 과학빅데이터의 전송을 위하여 유선 백본 네트워크에서는 TCP 성능을 최적화하여야 한다. 또한 최근 연구가 진행되고 있는 고속 무선통신 기술을 활용한 실험장비내 데이터 전송은 근거리 통신을 중심으로 높은 주파수대역 중심으로 수행되어야 한다.

<표1> ScienceDMZ 네트워크 구조 설계 제안

항목	기존 네트워크	Science DMZ
제안	일반 트래픽과 과학 빅트래픽 충돌, 연결거리 및 지연시간 차이에 따른 물리적 전송성능 저하	네트워크 구조적 분리, 분리된 네트워크에 대한 보안기술적용 수준의 조정을 통한 성능저하이슈 해결
속도	보안장비(IDS, 방화벽)으로 인한 성능저하, TCP, 시스템 버퍼 및 MTU 크기 등을 관해 적으로 설정을 통한 성능저하	성능저하가 없는 보안기술(ACL, SDN Flow) 적용, 개선된 TCP 혼잡제어를 통한 대용량 전송환경 구성, 장/단거리 네트워크 구조(버퍼, MTU 등) 설계
관리	트래픽 중심(SNMP, Netflow)의 모니터링의 한계	성능 중심의 단대단 모니터링 가능, 실제 전송시간 중심의 성능특정 스케줄로 정확한 성능 및 안정성

과학빅데이터의 원거리전송을 위하여, 백본 대역은 40G/100G/600G 등으로 지속적으로 발전하고 있다. 또한, 대용량 데이터 전송 시 활용되는 TCP 프로토콜의 성능을 개선하기 위하여, 대용량 데이터 전송에 최적화된 BIC TCP, FAST TCP를 사용되고 있다. 과학빅데이터의 원거리 전송시, 높은 BDP(Bandwidth Delay Product)가 형성되며, 이를 최적으로 활용하는 TCP 혼잡제어 방식이 요구되기 때문이다. 이러한 개선과 더불어, 네트워크의 구조적인 혁신을 위하여, 네트워크 성능저하의 큰 요소인 방화벽, 침입탐지시스템 등의 성능저하 요인을 제거하기 위한 구조적 변형을 제안한다. ScienceDMZ 라는 이름의 혁신적인 네트워크 구조 설계 방안으로써, 네트워크 전송성능을 미치지 못하는 방화벽, 침입탐지시스템(IDS) 등의 보안장비 패킷처리 성능으로 인한 대용량 데이터 전송 성능 저하 문제를 해결하기 위한 것이다. 구조적으로는, 과학데이터전송(Science Zone)과 일반데이터전송(Working Zone)로 분리하여, 기존 네트워크 환경에서 과학빅데이터 전송 네트워크를 분리하여, 이를 바탕으로 보안장비를 경유하는 네트워크 경로를 제거하는 것이다.

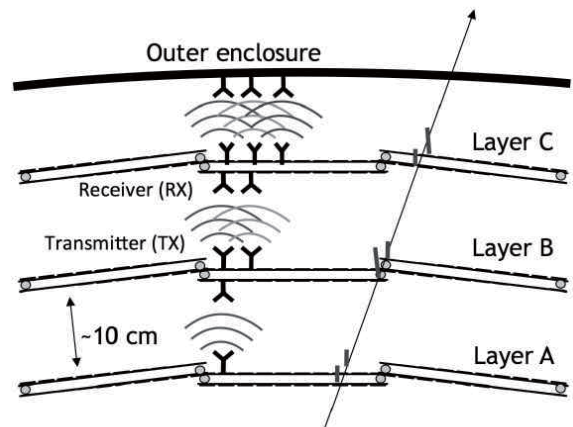
그림1의 사례처럼, 분리하는 방법으로, 광패스(Light-Path/Layer1), 가상랜(Vlan 및 Trunk/Layer2), Virtual Circuit 및 SDN (Software-Defined Networking) 슬라이스 등의 기술로써 2계층 망분리를 수행한다. TCP 전송최적화를 위하여, 높은 BDP(Bandwidth Delay Product)에 맞는 시스템 버퍼 적용, MTU(Maximum Transmit Unit) 적용이 요구된다. 또한 방화벽 등의 보안장비를 사용하지 않으면서 기본적인 보안정책을 수행하기 위하여, 네트워크 트래픽 접근제어(ACL), 위협 트래픽 모니터링 등 성능저하가 발생되지 않은 보안기술을 적용하게 된다. 그리고, 분리된 네트워크 상에서의 전송성능을 유지관리하기 위하여, 성능측정기술(PerfSONAR)를 적용하여 모니터링 하는 것이 추가로 필요하다.



(그림 1) Science DMZ 기술적 적용: 게이트웨이 라우터 중심으로 과학실험장비 데이터 분리

### 3. 실험장비에서의 60Ghz 무선전송 방안

과학실험장비는 대형화되고 많은 데이터를 생산하고 있다. 입자가속기는 그 대표적인 예로써, 생성되는 데이터는 전세계로 전송되어 물리학 연구에 활용된다. 이러한 대형실험장비에서 생성되는 데이터는 유선 케이블을 통하여 생성 및 전달되었으나, 최근 비면허 무선 주파수를 활용하여 더 많은 데이터를 케이블없이 생성 및 전달되도록 고안되고 있다. 스위스 CERN의 고광도 입자가속기(HL-LHC)를 활용하는 차세대 화소검출기(ATLAS Phase II Inner Tracker Pixel Detector) 연구장비에서 적중 클러스터를 모두 판독하기 위해 필요한 대역폭은 50-100 Tbps이며 10,000 여개의 상향링크를 필요로 한다[1]. 점점 높아지는 대역폭 요구사항을 만족하기 위해서 유선 연결을 위한 케이블의 증가가 수반되며, 이는 케이블링 공간으로 인한 측정불가영역(dead-zone)을 발생시키며, 또한 케이블 주위 공간에도 산란 및 상호작용, 열 발생 등으로 측정 데이터에 영향을 끼친다.



(그림 2) WADAPT의 대형연구장비 내 무선 연결 시나리오[1]

또한 케이블 경로로 인해 연구장비의 설계가 제한되어 제작의 효율성에도 영향을 주며, 케이블이나 커넥터의 파손 및 다수의 관리인력 운영 등으로 직접적인 비용 상승 요인이 되기도 한다. 이와 같이 과학 연구장비에 요구되는 고 대역폭의 제공에 더불어 실험장비 내 케이블 감소 및 전송 경로 최적화를 통한 설치비용 감소와 관리 간소화를 위해서, 60Ghz 기반의 근거리 초고속 무선통신 기술을 도입하게 되었다. 입자가속기 및 검출기 기술 발전을 위한 AIDAinnova 프로젝트(EU의 Horizon 2020 프로그램 지원)의 일환으로, WADAPT(Wireless Allowing Data and Power Transmission) 연구그룹에서 비면허대역 주파수를 활용하여 기존 검출기의 유선통신을 대체하기 위한 연구를 진행하고 있다.

이들의 목표는 그림 2와 같이 기존 검출기의 축 방향(Axial) 데이터 전송을 위한 수십 미터의 유선통신 대신, 약 1미터 거리의 무선통신으로 대체하여 전송 경로를 효율적으로 개선하는 것이다. 일반적으로 높은 주파수 대역을 사용 시 다음과 같은 장점들을 활용 가능하다. 먼저 높은 데이터 전송량의 달성이 가능하다. 고주파

수 대역은 허용 주파수 대역폭이 수GHz까지도 가능하므로 초고속 데이터 전송에 적합하다. 반면에, 주파수 대역이 높아질수록 투과력이 낮아지고 높은 신호감쇠에 의해 전파 도달거리가 좁아지는 단점이 있어 기존 이동통신 등의 상용 무선통신에서는 고주파수 영역을 선호하지 않았다. 하지만 연구장비간 근거리 통신 환경에서는 높은 신호감쇠를 오히려 장점으로 활용 가능하므로(주파수 재사용성 및 보안성 측면), 최대한 높은 주파수 대역을 활용하는 것이 유리하다.

60GHz 주파수 대역은 인근 주파수 대역들에 비해 공기 중 신호감쇠가 유독 높은 특징을 가지며, 이로 인해 근거리 통신을 위한 전파 도달거리 조절이 용이하다. O<sub>2</sub> 감쇠에 영향을 받는 주파수와 아닌 주파수를 비교해보면, 똑같이 1km 지점에서 30dB로 수신되도록 송신했을 때 O<sub>2</sub> 감쇠에 영향을 받는 경우엔 2.5km 지점까지만 가도 수신 신호세기가 0dB가 되지만, O<sub>2</sub> 감쇠에 영향을 안 받으면 32km 지점까지 가야 0dB가 된다. 이렇게 전파 도달거리의 조절의 용이함은 무선 링크 간 간섭의 영향을 줄이기 위해 활용 가능하다. 이러한 특성은 과학실험장비의 근거리 데이터 전송에 다양하게 활용이 가능하다.

#### 4. 결론

과학실험장비는 다양한 분야에서 구축되고 대형화되고 있다. 생성되는 데이터는 과학적 발견을 위하여 연구에 활용되고 있다. 이러한 데이터는 그 규모가 커짐에 따라 원거리 전송 및 장비내 생성과정에서 다양한 경로의 전송이 더욱 기술적으로 보완이 필요하게 되었다. ScienceDMZ 기술과 60Ghz 대역의 무선기술을 활용하여 과학데이터 전송에 안전하고 빠르게 잘 활용될 것이다.

#### ACKNOWLEDGMENT

본 논문은 KISTI 연구사업, “과학실험데이터의 근거리 초고속 전송을 위한 sub-Thz 무선기술연구” 지원받아 수행되었습니다.

#### 참 고 문 헌

- [1] The ATLAS Collaboration, “ATLAS Phase-II Upgrade Scoping Document (ver 1.0),” September 2015.
- [2] R.L.Grossman et al., “Data mining middleware for wide-area high-performance networks,” Future Generation Computer Systems, pp. 940~948, May 2006.
- [3] J. H. Moon et al., “A study on the DTN optimization for high performance data transfer over science DMZ,” KNOM Conf., Gwangju Korea, Jun. 2017.