

초저지연 실감형 서비스를 위한 5G 네트워크 기술 동향



김효일
UNIST



양현중
UNIST



이경한
UNIST



정방철
충남대학교

1. 서론

최근 ITU-R은 5G 이동통신으로 더욱 알려진 IMT-2020의 서비스 비전을 제시하였다^[1]. 5G 이동통신서비스는 크게 높은 전송속도와 높은 전송 효율을 요구하는 서비스, 높은 연결 밀도를 요구하는 서비스, 고신뢰 및 저지연을 요구하는 서비스로 분류되고 있다. 이러한 비전을 기반으로 3GPP Service and System Aspects (SA)에서는 5G에서 예상되는 서비스 시나리오와 요구사항을 활발히 검토 중이다^[2]. 특히, 저지연 서비스와 관련된 주제는 실시간 통신 및 촉각 인터넷이라 불리는 실감형 서비스로서 차량 및 로봇의 원격 제어, 교통 및 비행체의 실시간 제어, 원격 헬스 케어 등이 실제적인 사용예로 제시되고 있다^[3].

유럽을 중심으로 차세대 저지연 네트워크에 대한 연구가 진행되고 있다. 특히 2014년 발표된 IMT-2020에서는 5G 통신 네트워크의 기술적 목표를 다음과 같이 제안하고 있다^[4].

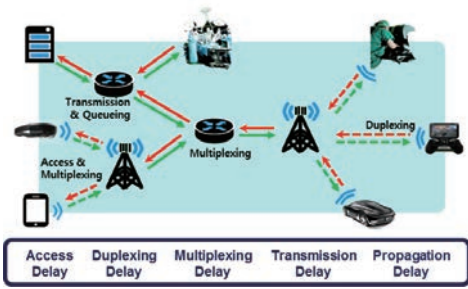
- 1~10Gbps 평균 전송 속도
- 1ms 종단 간 지연시간
- 10~100배 많은 단말기 지원
- 90% 줄어든 네트워크 에너지 소모량

본 기고에서는 1ms 지연시간 요구조건은 기존의 4G 무선 네트워크 구조로는 달성이 불가능하며, 새로운 형태의 무선 접속 네트워크가 필요함을 보인다.

독일의 TU Dresden 대학에서는 촉각 인터넷(Tactile Internet)



주창희
UNIST



〈그림 1〉 5G 네트워크 지연요소

을 목표로 5GNOW (5th Generation Non-Orthogonal Waveforms for Asynchronous Signaling Project) 프로젝트를 주도하여 지연시간 성능 및 주파수 효율성을 극대화를 꾀하고 있다. 이를 위해 비동기식 무선 접속 기술을 5G의 핵심기술로 판단하고 표준화를 위한 연구가 진행 중이다.^[5-8]

또한 Ericsson 및 유럽의 회사들을 중심으로 METIS (Mobile and wireless communications Enablers for Twenty-Twenty (2020) information Society) 프로젝트가 발족되었으며 6개의 연구 주제에 대해 표준화를 위한 선행 연구를 진행하였다. 이들 연구주제 중에서 “Ultra Reliable Communication” 분야에서 지연 시간에 대한 신뢰성 (Reliability) 성능 지표를 정의 하고, 다양한 프로토콜 및 알고리즘에 대해 논의 하고 있다. 최근 이들 프로젝트들은 5GNOW2 및 METIS II^[9] 프로젝트로 확장되어 지속적인 연구가 이루어지고 있다.

그러나 실제 사용자에게 초저지연 실감형 서비스를 종단간으로 제공하기 위해서는 기지국과 단말간의 데이터 전송 지연을 줄이는 기술만으로는 충분하지 않다. 기지국과 유선망의 인터페이스, 그리고 유선망 내에서의 저지연 전송을 통해 데이터를 서비스 사용자까지의 전달할 수 있는 통합 네트워크 구조가 필요하다. 따라서 5GNOW와 METIS 등에서 주도하는 5G 이동통신 기술과 함께, 유무선을 동시에 고려하는 초저지연 네트워크 기술의 개발이 필요하다.

II. 5G 네트워크 지연요소

5G 이동통신 네트워크는 사용자가 직접 단말에 연결하는 무선 접속(access) 단과, 다수의 사용자들을 효과적으로 연결하기 위한 유선 코어 망으로 구분된다. 사용자가 체감하게 되는 종단간의 지연요소는 〈그림 1〉에서와 같이 각 네트워크 요소별로 나눌 수 있다.

- 사용자에서 무선 네트워크에 진입하기 위한 접근 (access) 지연
- 하나의 무선 채널을 통해 양방향 전송을 지원하기 위한 Duplexing 지연
- 자원효율성을 높이기 위해 다수의 플로우들이 자원을 공유함으로써 발생하는 멀티플렉싱 (multiplexing) 및 큐잉(queueing) 지연
- 데이터를 유무선 채널 위에 실기 위한 전송 (transmission) 지연
- 신호를 물리적으로 전달하기 위한 전달 (propagation) 지연

실제 사용자에게 초저지연 실감형 서비스를 종 단간으로 제공하기 위해서는 기지국과 단말간의 데이터 전송 지연을 줄이는 기술만으로는 충분하지 않다. 기지국과 유선망의 인터페이스, 그리고 유선망 내에서의 저지연 전송을 통해 데이터를 서비스 사용자까지의 전달할 수 있는 통합 네트워크 구조가 필요하다.

네트워크 지연요소들 중에서 전달지연과 duplexing 지연은, 물리적인 제약으로 인해 더 이상 줄일 수 없거나 양방향 채널할당을 통해 쉽게 해결할 수 있다. 따라서 본 기고에서는 두 요소들을 제외하고, 무선 접근 지연, 멀티플렉싱 및 큐잉 지연, 그리고 전송 지연에 대하여 연구 개발 현황과 앞으로의 발전 방

향을 살펴본다. 5G 초저지연 네트워크의 전반적인 조망을 위해 위의 요소들을 역순으로 살펴보고자 한다.

1. 전송 지연

유선 네트워크의 데이터의 전송은 DWDM (Dense Wavelength Division Multiplexing)을 사용한 광 네트워크를 통해 이루어진다. 따라서 DWDM의 전송지연에 대해 먼저 분석할 필요가 있다. 광네트워크는 크게 광케이블, 광신호를 송수신하는 Transponder, Dispersion



compensation module (DCM), 신호 증폭기 (amplifier) 등으로 구성되어 있다^[10].

DWDM에서 전송지연을 가져오는 요소들 중 가장 중요한 요소는 광케이블의 길이가 있다. 일반적으로 광케이블에서 빛은 공기중에서보다 느리게 전파되며, 광케이블의 refraction 특성에 따라 조금씩 다른 값을 가진다. Single mode fiber에서 일반적으로 1km의 거리를 전송하기 위해 빛이 전달되는 속도는 4.9us이며, 이보다 약 31% 빠른 속도로 전달이 가능한 Hollow Core Fiber (HCF)도 있으나, 신호 감쇄 (attenuation) 특성이 좋지 않음이 알려져 있다.

광케이블내에서 신호의 파장에 따라서 신호의 전달 속도가 다르며, 이로 인하여 Chromatic Dispersion (CD)가 발생한다. CD를 보상하기 위해 Dispersion Compensation Module (DCM)을 사용한다. DCM 방법은 크게 Dispersion Compensating Fiber (DCF) DCM, Non-Zero Dispersion Shifted Fibers (NZ-DSFs) DCM, Fiber Bragg Grating (FBG) DCM이 있으며, DCF DCM은 신호전송에 추가로 15~25%의 지연시간이 추가로 소요되며, NZ-DSF DCM은 약 5%의 지연시간이 추가된다. FBG DCM을 사용할 경우에는 (케이블 길이에 상관없이) 5~50 ns의 지연시간만이 추가로 요구된다.

Optical 신호 증폭기는 Erbium doped fiber amplifier (EDFA)가 널리 사용되며, 내부적으로 약 30m의 Erbium doped fiber를 통해서 신호를 증폭시킨다. 30m의 추가 길이로 인하여 약 147ns의 지연시간이 추가된다. 최근에는 Stimulated Raman Scattering 기법을 활용한 Raman amplifier가 개발되었으며, Raman amplifier를 사용할 경우, 추가적인 지연시간 없이 신호를 증폭할 수 있다.

Transponder는 전기신호와 광신호의 전환을 담당하며, 일반적으로 5~10us의 지연시간이 요구된다. 최근에는 FEC (Forward Error Correction) 이나 in-band management channel 기능이 없는 light-weight 설계

를 통해 4~30ns의 지연시간을 달성하였다^[11].

네트워크 인터페이스에서는 프레임을 전송하기 위한 전송 지연시간이 요구된다. 전기신호와 광신호의 전환은 프레임이 모두 버퍼에 전달된 후에 발생한다. 따라서 프레임의 크기와 전송속도에 의해서, 크기/전송속도의 형태로 정해지며, 작은 크기의 프레임일수록 더욱 작은 지연시간을 가진다.

그밖에 추가적인 지연시간으로 광신호-전기신호-광신호 (OEO) 전환에 약 100us의 지연시간이 필요하며, FEC를 위해서 15~150us, 그 외의 DSP (Digital Signal Processing; e.g., QPSK 모듈레이션)에 약 1us의 추가 지연시간이 요구된다. 따라서 저지연 성능을 위해서는 광 네트워크 내에서 OEO 전환, FEC, DSP의 과정을 거치지 않는 것이 도움이 된다.

이해를 돕기위해 160 km 길이의 10 Gbps 광 메트로 네트워크에서 64 byte의 프레임을 보내는 경우를 고려하자. 송신단과 수신단에서 각각 Transponder를 거치고, 네트워크 내부에서 두 번의 광증폭기와 한번의 DCM을 거친다고 했을 때, DCF DCM과 EDFA를 사용할 경우, 총 1ms 정도의 지연시간이 걸리며, FBG DCM과 Raman amplifier를 사용할 경우 약 0.784 ms의 지연시간이 걸린다.

**우리나라의 서울-부산간 거리 약 325 km
에서 크기가 500~1000 byte인
일반적인 IP 패킷의 왕복 전송을 고려했을
때, 패킷의 최대 전송 지연의 하향
경계값(low bound)은 약 5~7.5ms에
이를 것으로 예상된다.**

우리나라의 서울-부산 직선거리는 약 325 km이며, 광신호의 왕복을 고려했을 때 1000km의 거리가 최대 전송 거리일 것으로 예상된다. 또한 일반적인 IP 패킷이 전송된다고 가정하면, 프레임의 크기는 약 500~1000 byte로 증가하여 네트워크 인터페이스의 전송시간이 10~20배 증가한다. 이들을 고려했을 때, 사용자 중단간 지연시간의 최대값의 하향 경계값(low bound)은 약 5~7.5ms에 이를 것으로 예상된다.

2. 멀리플렉싱 및 큐잉 지연

5G 무선 액세스 망과 유선 옵티컬 네트워크에서의 지연시간 혁신 노력과 함께, 초저지연 인터넷을 실현하기 위해 풀어야할 다양한 숙제들이 남아있다. 예를 들



〈그림 2〉 인터넷 스위치들이 가지는 단순한 큐 구조

면 1) DPI (Deep Packet Inspection), NAT (Network Address Translation), 과금, 방화벽 등에서 추가적으로 발생하는 패킷 연산 지연시간, 2) 서버 및 단말의 운영체제가 발생시키는 NIC (Network Interface Card) 과 네트워킹 어플리케이션 간의 Context switching 및 버퍼링 지연시간, 3) 유선과 무선 네트워크를 연결하는 게이트웨이에서 발생하는 자원 미스매치를 해결하기 위한 큐잉 지연시간, 4) 인터넷의 전송률을 끌어올리기 위해, 모든 스위치에서 기본적으로 발생시키는 패킷 멀티플렉싱을 위한 큐잉 지연시간 등에 대한 해결책이 필요하다.

SDN 기반 컴퓨팅 및 운영체제 구조 혁신을 통해 해결

할 수 있을 것으로 기대되는 1), 2)

와 달리 3), 4) 는 네트워크 분야의 오랜 딜레마인 패킷 큐잉에 의한 “전송률 vs. 지연시간” 문제와 관련이 깊고, 그만큼 해결하기 쉽지 않다. 또한, 그 문제의 심각성도 앞서 다뤘던 유선 네트워크에서의 물리적인 지연요소에 비해 훨씬 심각하

다. 일례로 울산에서 서울까지의 왕복 지연시간은 광케이블 또는 구리선에서의 신호 전파시간인 20만 km/s를 고려할 때 5ms 내외여야 하나, 실제로 관측되는 지연시간은 15~30ms 로, 물리적 지연시간의 수배에서 최대 수십배에 달하는 지연시간이 큐잉에 의해 발생하며, 이는 인터넷 상에서 유무선 네트워크를 막론하고 관찰되고 있다 [12-13]

패킷 큐잉으로 인한 지연시간의 문제는 〈그림 2〉의 예를 통해 손쉽게 이해될 수 있다. 다수의 패킷이 다양한 입력 인터페이스를 통해 유입되는 스위치는 특정한 출력 인터페이스에 대해 아래와 같이 패킷을 누적시키는 큐 구조를 가진다. 큐 구조가 가지는 전송률을 극대화하기 위해서는 해당 큐에 머무는 패킷의 수가 0이 되지 않도록 입력되는 패킷의 양을 적절히 조절해야 한다. 또한, 입력되

는 패킷의 양은 물리적으로 주어진 큐의 크기를 넘어가지 않는 범위내에서 조절되어야 한다. 입력되는 패킷양에 대한 이러한 제어는 인터넷의 말단에서 이루어지며, NUM (Network Utility Maximization) 이라 불리는 최적화 문제를 통해 다양하게 연구되어 왔으며, 혼잡제어라는 개념을 통해 전송계층에서 TCP (Transmission Control Protocol)로써 구현되어 왔다.

중요한 부분은 큐에 쌓여있는 패킷의 숫자를 0가 되지 않게 말단에서 제어하는 것이 쉽지 않다는 것이다. 기존 단말의 트래픽 제어기법들은 효율성만을 고려하여 큐에 많은 수의 패킷을 쌓는 기법을 선택했으며, 이를 통해 손쉽게 최대 전송률을 달성하는 대신, 보장할 수 없는 패킷 처리 시간이라는 반대급부를 얻게 되었다. 이는 패킷의 전송에 있어서 지연시간을 ‘보장’하는 것이 불가능함을 의미한다.

실감형 서비스를 위한 저지연 트래픽과 Best Effort 서비스로 대변되는 지연 허용 트래픽들에 대해 적절히 대응하는 새로운 네트워크 구조를 개발함으로써, 저지연을 보장하면서 높은 전송률을 달성하는 네트워크를 상용화할 수 있을 것으로 예상된다.

그러나 패킷이 지나는 인터넷 경로 상에 존재하는 모든 스위치들의 큐의 크기를 일정값 이하로 제한되도록 제어할 수 있다면, 이 정보를 활용하여 지연시간을 보장하는 것이 가능할 것으로 예상된다. 최근의 연구들은 패킷 왕복 지연시간 정

보가 단말에서 스위치에 큐잉된 패킷을 파악하기 위해 사용될 수 있다는 사실이 데이터센터 네트워크에서 입증되고 있다^[14]. 통제된 데이터센터 네트워크를 확장하여 인터넷에 적용하기 위해서는 훨씬 더 복잡하고 많은 요소들이 고려되어야 할 것으로 예상된다.

스위치의 큐잉 정보와 함께 중요한 것은 어떤 종류의 제어를 단말에서 수행해야 하는 가 이다. 이에 대한 답은 명확하지 않지만, ECN (Explicit Congestion Notification)과 같이 네트워크에서 도움을 받는 기법이 최근 관심을 받고 있다. ECN은 혼잡 상황을 겪는 패킷에 대한, 스위치로부터의 매우 단순한 bit marking 기법으로 인터넷 말단에 존재하는 단말이 인터넷 내부에 존재하는 스위치들과 대화할 수 있는 통로를 제공한다. 최근의 혼잡제어 기법들은 패킷들에 표시된 ECN bit 들을 보다 지능적으로 이해하는 알고리즘을 도입함으로써 말단으로

부터의 제어 효율이 크게 높아질 수 있다는 것을 보이고 있다^[15].

이러한 과정을 통해 인터넷을 통과하는 패킷들의 지연 시간이 예측 가능하더라도, 실제 저지연 인터넷을 실현하여 상용화하기 위해서는, 앞서 언급한 전송률로 표시되는 네트워크의 효율성과, 요구되는 지연시간 간의 상관관계에 대한 대응이 중요하다. 즉, 저지연을 요구하는 트래픽에 대해 지연시간을 보장하는 것과 동시에, 저지연이 요구되지 않는 트래픽을 수용함으로써 전송률의 극대화를 달성할 수 있는 방안이 요구된다. 실감형 서비스를 위한 저지연 트래픽과 Best Effort 서비스로 대변되는 지연 허용 트래픽들에 대해 적절히 대응하는 새로운 네트워크 구조를 개발함으로써, 저지연을 보장하면서 높은 전송률을 달성하는 네트워크를 상용화할 수 있을 것으로 예상된다. 이를 위해 저지연 트래픽의 진입허용을 제어하기 위한 단말/스위치간 협업 구조, 각 스위치

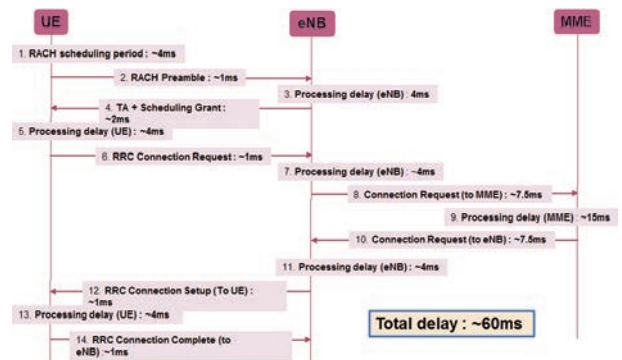
무선접속 지연시간을 단축하기 위하여 GFDM 등의 새로운 Waveform 기술, 서브프레임 설계기술, 상향링크에서 스케줄링 허가 없이 데이터를 전송하는 Grant-Free Access 기술 등이 고려되고 있다.

에서의 패킷 큐잉 구조에 있어서 하드웨어와 소프트웨어의 재설계가 필요하다. 일반적으로 이러한 네트워크 재설계 요구가 인터넷 전체에서 받아들여지기 힘들다는 것은 잘 알려진 사실이나, Tele-presence

3. 무선 접근 지연

무선 접근 (access) 지연은 사용자 단말에서 발생한 데이터를 기지국에 전달하기 위해 소요되는 시간을 의미한다. LTE 프로토콜은 초기 접속 및 자원 할당 등에 있어서 사용자확인, 과금, 사용자등록 등 세션 연결을 담당하는 제어 부분(Control Plane)과 데이터를 전송하는 사용자 부분(User Plane)으로 크게 나눌 수 있다. IMT-A는 4G 이동통신을 위해 제어 부분 100ms, 사용자 부분 10ms의 최대 지연시간을 요구 조건으로 제안하고 있다.

〈그림 3〉은 단말이 초기 Idle 상태에서 네트워크에 연결되는 Connected 상태까지의 프로토콜 진행과 각 과



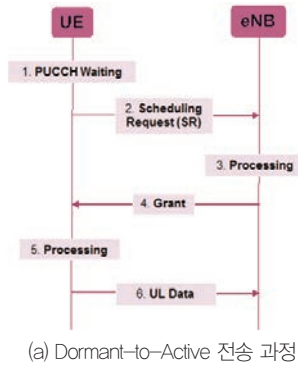
〈그림 3〉 Idle-to-Connect 평균 지연 시간 분석^[16]

정에서의 평균적으로 측정되는 지연시간을 보여준다^[16]. Idle-to-Connect 과정에서는 단말이 Idle 상태에서 경쟁을 통해 RACH (Random Access CHannel)에 접속하여 초기 등록 및 인증을 위한 절차를 거친 후에, 코어 네트워크 MME (Mobility Management Entity)와의 정보 교환이 필요하기 때문에 평균 지연시간이 60ms 정도 걸리는 것으로 알려져 있다.

단말이 Connected 상태 돌입 이후에도 송수신하는 신호가 일정시간동안 발생하지 않으면 휴면 모드인 Dormant로 전환되게 된다. 〈그림 4〉는 단말이 Dormant에 있으면서 네트워크와 동기가 유지되고 있을 때, 데이터를 보내기 위해 자원할당을 받고 초기 상향링크 데이터를 전송하는데 까지 소요되는 평균 지연시간을 분석한 결과이다. RACH의 경쟁이 없음에도 불구하고 자원 할당을 위한 핸드셰이킹과 최소 1ms의 슬롯 단위로 인하여, 10ms내외의 지연시간이 소모되는 것을 알 수 있다.

이러한 결과는 현재의 LTE 슬롯 구조와 프레임 구조를 가지는 무선 접속 방식으로는 5G에서 요구하는 종단 간 1ms를 만족시킬 수 없다는 것을 보여준다. 또한 〈그림 3〉에서 알 수 있듯이, 사용자 인증 및 코어 네트워크와의 정보 교환 절차를 제외하더라도, 프레임 및 슬롯 구조로 인하여 자원을 할당하는 과정에서 10ms 내외의 시간이 소요된다.

5G 이동통신 시스템에서 제공하기 위하여 고려되고 있



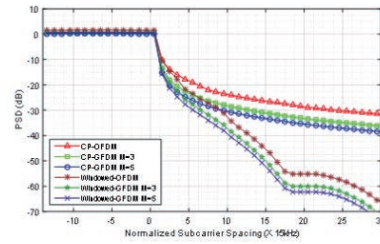
(a) Dormant-to-Active 전송 과정

| FDD | | | TDD | |
|---------------------|--------------------|-----------|-----------------------------|-----------------------------|
| Component | Process | Time (ms) | RACH in 2, 7-subframe (80%) | RACH in 3, 8-subframe (20%) |
| 1 | PUCCH Waiting | 0.5 | 2 | 0.5 |
| 2 | Scheduling Request | 1 | 1 | 1 |
| 3 | eNB processing | 3 | 3 | 5 |
| 4 | Grant | 1 | 1 | 1 |
| 5 | UE Processing | 3 | 5 | 3 |
| 6 | UL Data | 1 | 1 | 1 |
| Total delay | | 9.5 | 13 | 11.5 |
| Average total delay | | 9.5 | 12.7 | |

(b) 평균 지연시간 분석

〈그림 4〉 Dormant-to-Active 상향링크 평균 지연시간 분석^[17]

는 무선 기술들은 크게 웨이브폼 (waveform) 기술, 서브프레임 설계 기술, 상향링크 데이터 전송 기술로 나뉠 수 있다. 먼저 웨이브폼 기술의 경우 3GPP Radio Access Network (RAN)에서는 다양한 주파수 대역과 서비스 시나리오를 하나의 무선 접속기술로 수용하는 것을 지향하고 있어 OFDM 웨이브폼을 기본으로 사용되 서로 다른 주파수 대역과 서비스 시나리오에 따라 numerology를 다양하게 변화시키는 것이 고려되고 있다. 그러나 최근 제안되고 있는 Universal Filtered Multi-Carrier (UFMC), Generalized Frequency Division Multiplexing (GFDM), Filter Bank Multi-Carrier (FBMC) 등의 새로운 웨이브폼이 사용될 가능성도 있다. GFDM은 비직교 무선 접속 방식으로 기본적으로 사용자 간 전송 동기가 없다는 가정 하에 band-pass 필터 등으로 다른 사용자의 스펙트럼을 제거한다^[8]. 따라서 OOB (out-of-band emission)로 인한 간섭이 성능 저하를 결정하는 중요한 요소가 된다. 〈그림 5〉은 기존의 OFDM과 GFDM의 전력 스펙트럼 밀도를 나타낸다. 파라미터 M은



〈그림 5〉 OFDM과 GFDM의 전력스펙트럼밀도 비교^[18]

GFDM 방식에서 subsymbol의 개수를 나타내는 값으로, M이 증가하면 수신기 계산 복잡도가 증가하지만 OOB가 줄어들어 비동기 사용자 간 간섭이 낮아지는 효과가 있다. 기존의 OFDM은 동기 시스템에서는 OOB가 주파수 효율성 성능에 무관하기 때문에 높은 주파수 효율성을 보여주나, 비동기 환경에서는 OOB가 사용자 간 간섭을 결정하기 때문에 GFDM이 주파수 효율성 측면에서 우수한 것을 보여준다.

서브프레임 설계 기술은 저지연 서비스에 적합한 구조로 프레임 구성하는 기술로서 가장 기본적인 설계 철학은 Transmission Time Interval (TTI)라 불리는 데이터 전송의 기본 단위를 줄이는 것이다. 서브프레임의 길이가 줄어들면 스케줄링, 전송, ACK/NACK 등의 절차의 기본 시간 단위가 줄어들어 지연시간을 줄이는 효과가 있다. 추가로 고려되고 있는 저지연 서비스용 서브프레임 설계 기술은 Self-Contained TDD 서브프레임 기술이다. 이 기술은 하나의 서브프레임 안에 하향링크와 상향링크를 모두 포함하여 전송된 데이터에 대한 ACK/NACK를 동일한 서브프레임 내에서 전송할 수 있도록 하여 지연시간을 줄이는 기술이다. 이외에도 다양한 서브프레임 구조가 저지연 서비스를 위하여 제안되고 있다.

마지막으로 상향링크 데이터 전송기술은 기존의 LTE 시스템이 스케줄링에 기반하여 상향링크 데이터를 전송하는 것과 달리 저지연 서비스 트래픽을 상향링크 데이터 전송에 대한 허가 없이 전송하는 이른바 Grant-Free Access가 고려되고 있다. 이와 비슷한 개념으로 경쟁기반 PUSCH (Contention-Based Physical Uplink Shared Channel, CB-PUSCH)도 3GPP 표준에서 고려되고 있다^[19]. 또한 저지연 서비스를 위한 상향링크 데이터 전송기술로서 기존 랜덤액세스 기술을 개선하여 활용



하는 다양한 시도들이 제시되고 있다.

III. 종단간 Quality of Service

종단간 지연시간을 보장하기 위해서는 서로 이질적인 무선 접속단과 유선 코어망을 상호 연결하는 노력이 요구된다. 본 섹션에서는 Quality of Service (QoS) 지원을 목표로 제안된 기존의 지연시간 보장 방법들을 살펴본다.

1. 유선망의 QoS: IntServ vs. DiffServ

인터넷에서의 QoS provisioning은 크게 IntServ와 DiffServ 두 가지 QoS 기법으로 나뉜다. IntServ는 종단간 세션을 서비스하는 모든 인터넷 라우터에서 리소스를 할당 및 예약함으로써 QoS를 보장하는 기법이다. 이와 달리 DiffServ는 서비스 요구사항 별로 트래픽의 클래스를 정의하고 각 클래스별로 차별화된 QoS를 보장하는 기법이다.

IntServ는 fine-grained QoS를 제공하는데, Resource Reservation Protocol (RSVP) 기반 종단간 시그널링 및 RSPEC 기반 QoS 요구 스펙 명시로 구성된다^[20]. RSVP는 종단간 플로우 path 상에 위치한 라우터들 간에 PATH 및 RESV 메시지를 주고받는 과정을 통해 해당 플로우의 QoS 보장을 위해 필요한 리소스를 선점하게 되며, RSPEC은 QoS 요구 정도에 따라 Best Effort, Controlled Load, Guaranteed 중 하나를 명시하게 된다. 특히, Guaranteed로 명시된 트래픽들은 보장된 지연시간을 넘지 않도록 요구된다.

DiffServ는 시그널링의 복잡도로 인한 Scalability 이슈가 있는 IntServ의 단점을 극복할 대안으로 제시되었으며, IP 헤더의 Differentiated Services (DS) 필드를 통해 최대 64가지의 QoS 클래스를 정의한 후 각 라우터가 패킷 별 클래스에 맞는 차별화된 서비스 보장을 할 수 있도록 하는 기법이다^[21]. QoS 요구 정도에 따라 Default, Class-selector, Assured Forwarding (AF),

Expedited Forwarding (EF) 으로 나뉘며, 이 중 EF가 가장 높은 수준의 지연시간 보장을 제공한다.

종단간 delay 측면의 QoS 보장이라는 관점에서는 상기 두 기법 간에 분명한 장단점이 존재한다. IntServ는 자원 선점을 통한 안정적 서비스가 가능하지만, 각 라우터별로 IntServ 관련 상태들을 별도 관리하여야 하므로 인터넷이라는 대규모 네트워크에서는 Scalability 문제가 발생하고, 리소스 선점으로 인해 다수의 저지연 플로우들이 멀티플렉싱되는 환경에서는 자원 활용률이 현저히 떨어지게 된다. 반면, DiffServ는 기존 IP 패킷 구조를 활용하여 간접적인 QoS 보장이 가능하고 복잡도를 줄일 수 있다는 장점이 있으나, 클래스별 QoS 보장만을 지원하므로 각 플로우 별로 세밀한 지연시간 요구조건은 보장하기에 어렵다는 단점이 있다. 5G 네트워크 규모에서 초저지연 서비스를 실현하기 위해서는, 두 기법의 장점을 결합한 새로운 형태의 저지연 QoS 아키텍처가 요구된다.

2. 무선망의 QoS: WLAN 및 LTE 사례

무선에서의 QoS 보장은 가장 보급률이 높은 IEEE 802.11 WLAN과 3GPP LTE를 중심으로 소개한다.

종단간 지연시간을 보장하기 위해서는 서로 이질적인 무선 접속단과 유선 코어망을 상호 연결하는 노력이 요구된다. 기존의 단말-스위치 간 프로토콜을 확장하여, 유무선 라우터 간에도 자원관리 및 QoS 보장 상황을 유기적으로 연동할 수 있는 새로운 유무선 orchestration 방식을 개발하는 것이 필요하다.

WLAN에서는 IEEE 802.11e 표준을 통하여 QoS를 제공하고 있다^[22]. 해당 표준에서는 QoS의 우선순위에 따라 AC_BK, AC_BE, AC_VI, AC_VO 의 네 가지 Access Category (AC)를 정의하고, 각 AC 별로 Traffic Category

Identification (TCID) 우선순위를 두 가지 씩 할당하였다. 또한, AC 별로 별도의 MAC (Medium Access Control) 파라미터를 설정함으로써 AC간 차별화된 QoS 서비스가 가능하다.

LTE에서는 bearer 단위로 QoS를 제공하는데, bearer는 UE 및 eNB 간 radio bearer, UE 및 P-GW 간 ESPS bearer 등과 같이 LTE 망의 구간 별로 정의하거나, default bearer, dedicated bearer 등과 같이 QoS 보장 특성에 따라 정의된다^[23]. UE가 LTE 망에 처음 접



〈그림 6〉 유무선 orchestration을 위한 피드백 구조

속 시에는 default bearer가 할당되지만, 저지연을 요구하는 서비스를 UE가 시작하는 경우 이에 맞는 dedicated ESPS bearer가 생성된다. 이렇게 생성된 bearer는 8가지 QoS class identifier (QCI) 중 하나를 부여받게 되는데, 각 QCI 별로 지연시간과 Packet Error Rate (PER) 두 가지의 QoS 지표가 차별화되어 정의되어 있다. 예를 들어, Real-time Gaming과 같은 저지연 서비스의 경우 최대 30ms의 지연시간 보장을 명시하였다.

이러한 WLAN과 LTE의 QoS 보장 기법들은 종단간 초저지연 보장을 위하여 사용되기에는 아직 여러 가지 한계를 노출하고 있다. 먼저, IEEE 802.11e 같은 경우 네 가지의 AC만을 정의하였고 실제 동작은 MAC (Medium Access Control)에 따라서 달라진다. 따라서 임의의 지연시간을 유발할 수 있는 현재의 IEEE 802.11에서는 지연시간을 효과적으로 제한하기가 힘들다. 한편, LTE의 QCI 별 지연시간 요구는 radio bearer에만 국한되며, LTE 코어에서 20ms의 추가적 지연시간이 발생하므로 종단간 초저지연을 지원하기에는 아직 갈 길이 멀다 하겠다.

3. 유무선 간 QoS 매핑 및 orchestration

앞서 살펴본 바와 같이 유선과 무선 네트워크는 각각 다른 방식으로 QoS를 보장하려는 시도를 해왔다. 그러나, 다가올 5G 네트워크는 폭발적인 모바일 트래픽 증가와 더불어 유선 인터넷망과 last-mile 무선 네트워크가 유기적으로 결합된 형태가 될 것이므로, 종단간 10ms와 같은 초저지연 요구조건을 보장하기 위해서는 유무선 간 QoS 보장 기법이 유기적으로 연동되어야 한다.

기존 연구에서는 IntServ의 RSVP 프로토콜을 IEEE 802.11e와 연동하여 유무선 간 정합을 시도하였으며^[24], DiffServ와 IEEE 802.11e를 연동하기 위해 DSCP에

서 정의된 클래스 각각을 TCID 값으로 매핑하고 AC 별 service queue에 할당하는 매핑 기법도 제안하였다^[25]. 이러한 시도에도 불구하고, 여전히 유선 및 무선 각각의 구간에서는 종전의 QoS 보장방식에 의거한 서비스를 제공하게 되므로, 기존의 기술적 한계를 뛰어넘지 못하는 문제를 안고 있다. 게다가, 무선 자원은 주파수, 시간, 안테나 등의 다차원 리소스 블록으로 구분되고 할당되는데 반해 유선 자원은 시간 단위의 스케줄링에만 의존하기 때문에, 유무선 간 자원 관리의 불균형 현상을 해결하기가 어렵다. 예를 들어, 무선 구간은 MIMO, OFDMA 등의 최신 기법을 동원하여 다수의 유저들을 지원할 수 있는 반면, 유선 구간에서는 last-mile 무선 유저들의 폭발적 트래픽 생성량을 제한할 수 있는 효과적인 방법이 부재하다.

이러한 문제를 해결하기 위해서는 기존의 단말-스위치 간 프로토콜을 확장하여, 유무선 라우터 간에도 자원관리 및 QoS 보장 상황을 유기적으로 연동할 수 있는 새로운 유무선 orchestration 방식을 개발하는 것이 필요하다. 이를 실현하기 위해서는 〈그림 6〉과 같이, eNB와 AP와 같은 무선 단 라우터들이 유선 라우터의 스케줄링 기법과 함께 피드백을 교환하며 자원관리 방식을 진화시키는 것이 필요하다.

IV. 맺음말 및 향후 연구방향

데이터와 청각, 시각적 신호의 전달을 넘어 촉각 등 공감각적인 Tele-presence가 가능한 실감형 서비스를 제공하기 위해서, 초저지연 트래픽의 QoS를 만족시킬 네트워크의 개발이 요구된다. 이러한 서비스의 보편적인 활용을 위해서는 자원의 효율성을 담보함으로써, 시장에서의 서비스 경쟁력 또한 반드시 확보해야 한다. 본 기고에서는 5G 네트워크의 다양한 지연요소들을 중심으로 저지연 서비스를 위한 기술 동향 살펴보았으며, 이를 통해 실감형 저지연 서비스를 제공하기 위해서는 현재의 LTE 시스템의 혁신과 함께, 패러다임 전환을 통한 유선 네트워크 구조의 변혁, 그리고 유무선 통합 자원관리를 위한 피드백 시스템의 필요성을 확인하였다.



V. Acknowledgement

This work was supported by the IITP grant funded by the Korea government (MSIP) (No. B0126-16-1064, Research on Near-Zero Latency Network for 5G Immersive Service).

참고문헌

- [1] Rec. ITU-R M.2083-0, "IMT vision – Framework and overall objectives of the future development of IMT for 2020 and beyond," Sept. 2015.
- [2] E. Guttman, "Progress and content of SA1 5G study item," 3GPP RAN 5G Workshop, RWS-150032, Sept. 2015.
- [3] 3GPP TR, "Feasibility study on new services and markets technology enablers," 22.891v1.1.0, Nov. 2015.
- [4] IMT-2020 (5G) Promotion Group. "IMT Vision towards 2020 and Beyond", 2014.
- [5] 5GNOW, "5G cellular communications scenarios and system requirements," D2.1, 2013.
- [6] 5GNOW, "5G waveform candidate selection," D3.1, 2015.
- [7] 5GNOW, "Consistent 5G radio access architecture concepts," D2.2, 2015.
- [8] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. ten Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, F. Wiedmann, "5GNOW: Non-orthogonal asynchronous waveforms for future mobile applications," IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, Feb. 2014.
- [9] 5GPPP and METIS II, "5G RAN architecture and functional design," Whitepaper, Mar. 2016.
- [10] V. Bobrovs, S. Spoliitis, and G. Ivanovs, "Latency Causes and Reduction in Optical Metro Networks", SPIE 9008, Dec. 2013.
- [11] "BTI Systems claims 20-ns transit time for low latency transponder", Lightwave online magazine, June 2010.
- [12] J. Gettys and K. Nichols, "Bufferbloat: Dark buffers in the Internet," Communications of the ACM 9(11):57-65, 2011.
- [13] H. Jiang, Y. Wang, K. Lee, and I. Rhee, "Tackling Bufferbloat in 3G/4G Networks," in ACM SIGCOMM IMC, 2012.
- [14] M. Alizadeh, A. Greenberg, D. A. Maltz, J. Padhye, P. Patel, B. Prabhakar, S. Sengupta, and M. Sridharan, "Data center TCP (DCTCP)," In ACM SIGCOMM, 2010.
- [15] R. Mittal, V. T. Lam, N. Dukkipati, E. Blem, H. Wassel, M. Ghobadi, A. Vahdat, Y. Wang, D. Wetherall, D. Zats, "TIMELY: RTT-based Congestion Control for the Datacenter," ACM SIGCOMM, 2015.
- [16] Z. Savic, "LTE design and deployment strategies," Cisco 2016.
- [17] D. Singhal, M. Kunapareddy, and V. Chetlapalli, "LTE-Advanced: latency analysis for IMT-A evaluation," Tech Mahindra Whitepaper, 2010.
- [18] W. Park and H. J. Yang, "On spectral efficiency of asynchronous GFDMA and SC-FDMA in frequency selective channels," in Proc. IEEE Vehicular Technology Conference, May 2016.
- [19] Ericsson et al., "3GPP TR 36.881 V0.5.0," 3GPP RAN2 #92, R2-157181, Nov. 2015.
- [20] Cisco, "Resource Reservation Protocol," DocWiki, http://docwiki.cisco.com/wiki/Resource_Reservation_Protocol
- [21] Cisco, "DiffServ – The Scalable End-to-End Quality of Service Model," White Paper, August 2005.
- [22] IEEE Std 802.11-2012, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," March 2012.
- [23] N.A. Ali, A.M. Taha, and H.S. Hassanein, "Quality of Service in 3GPP R12 LTE-Advanced," IEEE Commun. Mag., vol.51, no.8, pp.103-109, August 2013.
- [24] L. Zhang and S. Zeadally, "Enabling End-to-End QoS over Hybrid Wired-Wireless Networks," Wireless Personal Communications, vol.38, no.2, pp.167-185, July 2006.
- [25] S. Park, et al., Collaborative QoS Architecture between DiffServ and 802.11e Wireless LAN, in Proc. IEEE VTC-Spring, 2003.



김효일

- 1999년 2월 서울대학교 전기공학부 학사
- 2005년 4월 Univ. of Michigan, EE: Systems 석사
- 2010년 12월 Univ. of Michigan, EE: Systems 박사
- 2010년 12월~2011년 7월 IBM T.J. Watson Research Center 박사후연구원
- 2011년 8월~2015년 8월 울산과학기술원 조교수
- 2015년 9월~현재 울산과학기술원 부교수

〈관심분야〉

무선 및 모바일 네트워크, 5G 네트워크, 무선인지통신, LTE-U



이경한

- 2002년 2월 KAIST EE 학사
- 2004년 2월 KAIST EE 석사 (지도교수: 정송)
- 2009년 8월 KAIST EE 박사 (지도교수: 정송)
- 2010년 4월~2011년 8월 North Carolina State University, Post Doc. (CS)
- 2011년 9월~2012년 7월 North Carolina State University, Senior Research Scholar (CS)
- 2012년 8월~현재 UNIST, School of ECE (CS Track), 조교수

〈관심분야〉

모바일 컴퓨팅, 저지연 인터넷, 데이터센터 네트워크, VR/AR 네트워크



양현종

- 2000년 3월 KAIST 전기및전자공학과 학사
- 2004년 3월 KAIST 전기및전자공학과 석사
- 2010년 8월 KAIST 전기및전자공학과 박사
- 2010년 8월 2011년 8월 한국해양연구원 연수연구원
- 2011년 10월~2012년 10월 Stanford University 박사후과정
- 2012년 10월~2013년 8월 Broadcom Staff II Systems Engineer
- 2013년 9월~현재 울산과학기술원 조교수

〈관심분야〉

무선 액세스, 무선 통신 PHY 알고리즘 및 이론



정방철

- 2002년 2월 아주대학교 전자공학부 학사
- 2004년 8월 KAIST 전기및전자공학과 석사
- 2008년 2월 KAIST 전기및전자공학과 박사
- 2008년 1월~2009년 8월 KAIST IT융합연구소 연수연구원
- 2009년 9월~2010년 2월 KAIST IT융합연구소 연구교수
- 2010년 3월~2014년 2월 국립경상대학교 조교수
- 2014년 3월~2015년 8월 국립경상대학교 부교수
- 2015년 9월~현재 국립충남대학교 부교수

〈관심분야〉

5G 이동통신, 통계적 신호처리, 정보이론



주창희

- 1998년 3월 서울대학교 전기공학부 학사
- 2000년 3월 서울대학교 전기공학부 석사
- 2005년 3월 서울대학교 전기공학부 박사
- 2005년 9월~2007년 6월 퍼듀대학교 박사후과정
- 2007년 7월~2010년 2월 오하이오주립대학교 Research Scientist
- 2010년 3월~2011년 8월 한국기술교육대학교 조교수
- 2011년 9월~현재 울산과학기술원 부교수

<관심분야>

유무선 네트워크, 5G 네트워크, Internet of Things