

5G 사물인터넷 응용을 위한 태스크 오프로딩 기반의 처리율 향상 기법

박석우*, 김병웅**, 김형진**, 나인호*

*군산대학교 전자정보공학부

**전북대학교 IT응용시스템공학과

e-mail : p_sw@kunsan.ac.kr, bwkim@lghitachi.co.kr, kim@jbnu.ac.kr,
ihra@kunsan.ac.kr

A Scheme for Throughput Improvement based on Task Offloading for 5G IoT Applications

Seok-Woo Park*, Byung-Woong Kim**, Hyung-Jin Kim**, In-Ho Ra*

*School of Computer, Information and Communication Engineering

Kunsan National University

**Department of IT Applied System Engineering

Jeonbuk National University

요 약

최근 급증하는 모바일 디바이스와 이에 따른 데이터 트래픽 폭증 문제를 해결하기 위하여 네트워크 재설계에 대한 필요성이 요구되고 있으며, 이와 관련하여 클라우드 컴퓨팅, 엣지 컴퓨팅, 소프트웨어 정의 네트워크, 네트워크 기능 가상화, 네트워크 슬라이싱 등의 기술을 5G 및 차세대 네트워크에 적용하기 위해 연구가 활발히 진행되고 있다. 본 논문에서는 5G 사물인터넷 환경에서 다양한 사용자의 전송 요구를 충족하면서 처리율을 향상할 수 있는 사용자 요구사항 별 우선순위 기반의 태스크 오프로딩 기법을 제안한다.

키워드: 태스크 오프로딩, 5G, IoT, 처리율, MEC 컴퓨팅

1. 서 론

정보통신 기술의 혁신과 급속한 성장에 따라 최근에는 5G 기술을 이용한 수십 Gbps 수준의 데이터 전송속도, ms 단위의 초저지연, 100만개 수준의 IoT 디바이스 동시 연결 등 무선통신 기술의 새로운 전환기가 시작되고 있으며, 이와 더불어 사물인터넷(IoT)의 급속한 확산과 다양한 서비스 제공이 가능해지고 있다[1, 2]. 또한, 다양한 센서 및 액츄에이터, 모바일 장치, 차량 및 기타 임베디드 장치와 같은 장치를 이용한 복합적 사물인터넷 환경이 활발히 구축되고 있으며, 이러한 환경 변화에 의해 네트워크에서 실시간에 처리해야 하는 데이터가 폭증하고 네트워크 과부하가 발생하고 있다.

급증하는 모바일 디바이스와 데이터 트래픽으로 인해 발생하는 문제점들을 해결하기 위해서는 네트워크 재설계가 필요하다. 이에 따라 클라우드 컴퓨팅(Cloud Computing), 엣지 컴퓨팅(Edge Computing), 소프트웨어 정의 네트워크(SDN), 네트워크 기능 가상화(NFV), 네트워크 슬라이싱(Network Slicing) 등의 기술을 활용한 네트워크 자원관리 기법에 관한 연구가 다각적으로 진행되고 있으며, 이 중에서 5G 모바일 엣지 컴퓨팅(MEC, Mobile Edge Computing) 기술은 네트워크의 데이터 트래픽 문제를 해결할 수 있는 솔루션으로 간주되고 있다. MEC를 사용하면 클라우드의 컴퓨팅 기능이 클라우드에

만 존재하는 것이 아니라 네트워크 엣지까지 확장된다. 이로 인해 네트워크 트래픽이 감소하고 네트워크 리소스 할당 효율성이 개선된다. 또한, 사용자 장비(UE)와 가까운 엣지에서 서비스를 처리하면 전송 지연을 축소하고 작업 처리시간을 감소시킬 수 있다.

5G MEC에 대한 연구는 아직 초기 단계에 있으며, 5G 응용 서비스는 주로 eMBB (Enhanced Mobile BroadBand), URLLC(Ultra Reliable Low Latency Communication), mMTC(Massive Machine Type Communications)를 대상으로 진행되고 있다. 따라서 본 논문에서는 5G 사물인터넷 환경에서 네트워크 트래픽 제어, 처리시간 감축 및 처리율 향상을 위한 태스크 오프로딩 기법을 제안한다.

2. 관련 연구

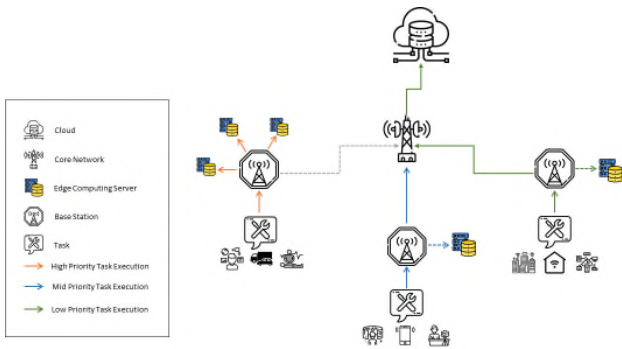
5G 환경에서 다양한 응용 서비스의 성능을 개선하기 위해 다양한 연구가 진행되고 있다. 논문 [3],[4],[5]에서는 5G 네트워크 및 엣지 컴퓨팅 시스템에서 처리량 및 서비스 품질 향상을 위한 방법으로 태스크 오프로딩의 가능성을 연구한 바 있다. 논문 [3]에서는 해당 프로세스의 구현을 위한 전략, 알고리즘 및 응용 사례를 포함하는 MEC 시스템의 태스크 오프로딩에 대한 연구를 수행하였으며, 논문 [4]에서는 태스크 오프로딩을 이용하여 엣지 및 클라

우드 컴퓨팅에서의 처리율 향상 방안을 제시하고 있다. 논문 [5]에서는 5G 네트워크에 대한 태스크 오프로딩 및 자원 할당 기술에 대한 연구를 수행한 바 있다. 논문 [6]에서는 소규모 셀 네트워크에서 대기 시간 및 자원 효율성 문제를 해결하기 위해 태스크 오프로딩을 위한 협업 모델을 제안하였으며, 논문 [7]에서는 5G 무선 액세스 네트워크에서 유연한 기능 분할을 위한 공동 태스크 오프로딩 접근 방식을 제안하였다.

또한, 인공지능 기술을 MEC 및 5G 네트워크에 적용하는 다양한 연구도 진행되고 있다. 이것은 태스크 오프로딩을 이용하여 네트워크 처리율을 향상시키는 효율적 접근 방식으로 간주되고 있다. 논문 [8]에서는 네트워크 상태의 전환과 관계없이 효율적인 자원 제공을 위한 DDPG 알고리즘을 제안하였으며, 논문 [9]에서는 심층강화 학습 기법을 통해 오프로딩의 불확실성으로 인한 처리 지연을 해결하는 방안을 제시한 바 있다.

3. 제안된 기법

3.1 시스템 아키텍처



(그림 1) 시스템 모델

본 논문에서는 (그림 1)과 같이 시스템 모델을 제안한다. 제안된 5G 네트워크는 기지국($B = \{1, 2, \dots, B\}$)과 엣지 컴퓨팅($N = \{1, 2, \dots, N\}$)으로 구성된다. 기지국은 코어 네트워크를 통해 클라우드와 연결되고 장치($d \in D$)는 시간 슬롯(t)에서 BS($b \in B$)에 작업을 요청(r)하고 큐잉(Queueing)에서의 수집 작업 요청($R = \{1, 2, \dots, R\}$)이 추가된다. 이후, 각 작업은 높음(H), 중간(M), 낮음(L)으로 우선순위를 분류되며, 각 작업의 분류에 따라 실행된다. 작업 요청을 분류하는 주된 이유는 요구사항에 따라 네트워크의 자원을 적절하게 분배하여 네트워크 자원의 활용을 최대화하기 위함이다.

실시간 처리 및 응답이 필요한 URLLC는 높은 단계의 우선순위로 분류한다. URLLC의 경우, 태스크 처리를 위해 BS의 EC 서버로 전송되어 처리한다. eMBB는 중간 단계로 분류하고 작업 처리를 위해 CN의 EC 서버 또는 CN에 있는 중앙 서버로 전송하여 처리한다. 마지막으로 mMTC는 낮음 단계로 분류하고 태스크를 BS를 거쳐 클

라우드로 전송하여 처리한다.

3.2 태스크 오프로딩 메커니즘

본 논문에서 제안한 태스크 오프로딩 메커니즘은 네트워크를 통해 전송된 요청의 우선순위를 중심으로 태스크 크기, 네트워크 대역폭, 전송 지연, 노드 처리 용량, 처리 및 실행 시간과 같은 다양한 요소를 기반으로 한 네트워크 자원할당 모델을 사용한다.

발신지 노드에서 목적지 노드로 태스크 오프로딩을 수행할 때의 총 에너지 소비는 킬로와트 (KW), NRG_{cons} 는 기가헤르츠(GHz), 컴퓨팅 용량을 C 로 나타낸다. 전송 지연(D_T)은 발신지 노드에서 태스크 요청 및 응답 지연 시간을 말하며 다음 식 (1)과 같다.

$$D_T = \alpha_i / \beta_{j,k} \quad (1)$$

여기서, α_i 는 태스크 처리를 위해 요청된 전체 태스크의 크기, $\beta_{j,k}$ 는 시작 노드 j 에서 목적지 노드 k 까지의 대역폭을 나타낸다. 지연 시간을 기반으로 시작 노드에서 목적지 노드까지 총 요청 지연 시간은 식 (2)와 같이 계산할 수 있다.

$$D_{Total} = \sum_{n=hops}^{\max} D_{request, response} \quad (2)$$

또한, 시작 노드에서 목적지 노드까지의 총 전파 지연 시간을 계산할 수 있으며, 이때, 두 노드의 거리(meter)와 전송속도(mb/sec)를 고려한다.

$$D_P = x / r_T \quad (3)$$

요청 처리 시간(P)은 계산된 요청 및 전파 지연 시간을 토대로 다음 식 (4)를 이용하여 계산한다.

$$P = 1 / (y * \beta) \quad (4)$$

여기서, y 는 대상 노드의 CPU 주파수이고 β 는 대역폭을 나타낸다. 총 요청 처리 시간 P_{total} 은 다음과 같다.

$$P_{total} = \sum_{n=hops}^{\max} P_{(j,k)} \quad (5)$$

네트워크의 각 노드에 대한 태스크 오프로딩 비용은 필요한 모든 매개변수가 계산되면, 벡터 $tc_{j,k} = \{D_{Total}, D_{p,total}, P_{total}, NRG_{cons}, C\}$ 를 이용하여 계산할 수 있다.

$$\min TC_{\alpha} =$$

$$D_{Total} + D_{Ptotal} + P_{total} + NRG_{cons} + C \quad (6)$$

태스크 오프로딩은 5G 네트워크의 mMTC, eMBB, URLLC에 따라 우선순위를 구분할 수 있다.

mMTC는 태스크 오프로딩을 위해 데이터 간헐적으로 전송하는 여러 IoT 장치를 포함하므로 낮은 우선순위를 갖는다. mMTC에 해당하는 IoT 장치들은 대부분 지연에 민감하지 않지만, 네트워크 혼잡을 줄이기 위해서는 충분한 대역폭이 필요하다. 따라서, 낮은 우선순위는 수식 (7)과 같이 계산한다.

$$\min \sum_{n=i}^{\max} NRG_{cons}^{\alpha}, \sum \beta_{j,k} \quad B, N \quad (7)$$

eMBB는 비디오, 스트리밍 등과 같은 서비스이므로 중간 우선순위는 갖는다. eMBB는 높은 데이터 속도가 필요하므로 수식 (8)과 같이 나타낼 수 있다.

$$\min \sum_{n=i}^m r_T, \sum \beta^+ \forall B, N \quad (8)$$

URLLC는 제한된 수의 링크뿐만 아니라 전송 및 전파 지연과 관련되는 공공, 차량 간 통신 및 의료 등의 중요한 서비스를 포함하는 서비스 유형으로 요청 및 전송 지연의 최소화를 위해 높은 우선순위로 분류하며, 다음과 같이 계산된다.

$$\min D_{T,total} + D_{P,total} \quad (9)$$

$$s.t. l_{j,k} \leq 2$$

4. 결론 및 향후 연구

본 논문에서는 다양한 애플리케이션의 서비스 요구사항을 충족하기 위한 시스템 모델을 제안하였다. 또한, 국제 전기통신연합(ITU)에서 정의한 초저지연(URLLC), 초고속(eMBB), 초연결(mMTC) 서비스를 고려한 우선순위 기반의 태스크 오프로딩 메커니즘을 제안하였다.

향후에는 각 노드에 미치는 태스크 오프로딩 비용을 최소화하고 네트워크 노드 정렬 및 매핑을 위한 그래프 이론 개념에 기초한 규칙 기반 알고리즘에 대한 연구를 수행하고 Bellman-ford 알고리즘을 이용하여 네트워크 노드를 정렬하고 네트워크 지연을 감소시킴으로써 태스크 오프로딩 비용을 최적화하는 기법에 대하여 연구하고자 한다.

Acknowledgment

본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1A2C2014333).

참고문헌

- [1] Jyotsna Agrawal, Rakesh Patel, Dr. P.Mor, Dr. P.Dubey and Dr. J.M.keller, "Evolution of Mobile Communication Network: from 1G to 4G", International Journal of Multidisciplinary and Current Research, 2015.
- [2] Amitabha Ghosh, Andreas Maeder, Matthew Baker, Devaki Chandramouli, "5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15", IEEE Access, 2019.
- [3] A. Islam, A. Debnath, M. Ghose, and S. Chakraborty, "A Survey on Task Mechanism in Multi-access Edge Computing," Journal of System Architecture, Vol. 118, No. April, p. 102225, 2021.
- [4] F. Saeik et al., "Task Offloading in Edge and Cloud Computing: A Survey on Mathematical, Artificial Intelligence and Control Theory Solutions", Computer Networks, 2021.

- [5] X. Chen, Z. Liu, Y. Chen, and Z. Li, "Mobile Edge Computing Based Task Offloading and Resource Allocation in 5G Ultra-Dense Networks", IEEE Access, 2019.
- [6] M. D. Hossain et al., "Collaborative Task Offloading for Overloaded Mobile Edge Computing in Small-Cell Networks," International Conference on Information Networking, 2020.
- [7] Z. Cheng, Y. Tang, and H. Wu, "Joint Task Offloading and Flexible Functional Split in 5G Radio Access Network," International Conference on Information Networking, 2019.
- [8] M. Chen, T. Wang, S. Zhang, and A. Liu, "Deep Reinforcement learning for Computation Offloading in Mobile Edge Computing Environment," Computer Communications, 2021.
- [9] M. Tang and V. W. S. Wong, "Deep Reinforcement Learning for Task Offloading in Mobile Edge Computing Systems," IEEE Transactions on Mobile Computing, 2020.