Resource Allocation for H2M Applications over Optical Mobile Fronthaul Networks

Elaiyasuriyan Ganesan
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu 41566, South Korea
suryaphd2018@gmail.com

Ho-Shin Cho
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu 41566, South Korea
hscho@ee.knu.ac.kr

***Abstract*— In the current, 5G and future 6G networks era, the development of low-latency tactile Internet (TI) is expected to bring new impact globally. The tactile internet will enable haptic and kinesthetic interactions between humans and machines in real and virtual environments through robots/machines. This necessitates low-latency, high capacity, and reliability. Therefore, in this paper we present an edge intelligence based Dynamic Wavelength Bandwidth Allocation (EI-DWBA) scheme in H2M services in optical mobile fronthaul. Simulation results show that the proposed scheme significantly improves the system throughput and reduces delay in different traffic situations, and also promises the quality-of-service (QoS) performance.**

Keywords—H2M, Optical Mobile Fronthaul, EI-DWBA

I. INTRODUCTION

Currently, the transition from the first generation (1G) to the fifth generation (5G) and subsequently to beyond future 6G networks mobile internet has brought about a significant advancement in information and communication technology. Communication networks have progressed from merely sending conventional traffic (i.e. Voice, Video, and Data) to also deliver machine-oriented traffic, such as the Internet of Things (IoT). The next stage of tactile internet (TI) has achieved a significant milestone in advancing human-to-machine (H2M) communication [1]. The objective is to enable remote control of both real and virtual objects through tactile and haptic communications. This allows humans to feel touch, force in their activities as if they were in a distant environment interacting with them. Fig.1 shows bidirectional haptic communications in the teleoperation system enabling a real-time exchange of control commands from human operators and robot feedback. The H2M enables a wide variety of applications such as healthcare, autonomous vehicles, augmented reality/virtual reality (AR/VR), gaming, etc. [2]. These H2M applications require ultra-low-latency, high capacity, high reliability and high bandwidth.

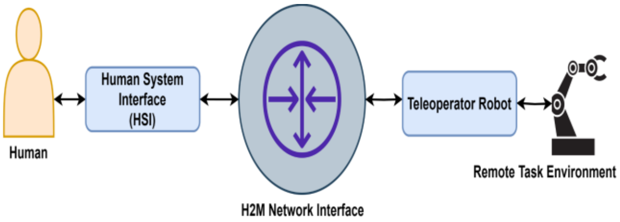

Fig. 1. H2M generic communication architecture.

The implementation of optical access and 5G mobile fronthaul networks in close proximity to end-users and robots presents a viable architectural approach for facilitating low-latency human-to-machine communication. Further, the improved fronthaul connectivity is necessitated by the ultra-low latency need in 5G and the future communication transmission and density of cell sites [3]. In order to achieve this objective, the IEEE 802.3ca task force released the 25G and 50G high capacity Next Generation Passive Optical Networks (NG-EPON). The NG-EPON utilizes multiple wavelength channels with a capacity of 25 Gbps each, enabling the transfer of data at different rates in both the downstream and upstream directions. In addition, channel bonding is a feature that enables NG-EPON to attain elevated data rates, resulting in aggregated data rates of Nx25Gbps [4]. Towards this end, NG-EPON technologies widely used in today's fiber-to-the-home (FTTH), residential, business, IoT and TI constitute promising solutions to 5G mobile fronthaul traffic flows. NG-EPONs should also meet the fronthaul network's requirements for low latency, jitter, and high throughput. Therefore, it is crucial to implement dynamic wavelength bandwidth allocation (DWBA) schemes that meet these demands [5]. NG-EPON DWBA scheme is contains the three-mechanism based on the wavelength manages such as single scheduling (SSD), multi-scheduling (MSD) and wavelength agile (WA).

As a result, various early studies have been conducted to improve bandwidth allocation through the implementation of predictive dynamic bandwidth allocation (DBA) or DWBA, employing a variety of methodologies, including predictive and genetic expression-based approaches, as mentioned in [6]. In recent years communication networks like, software-defined networking (SDN), cloud/edge and machine learning technologies have played a major role in optical and wireless network resource allocation. SDN is an essential part of low-latency networks because it decouples the control and data planes and offers centralized management over network devices. Also, the bandwidth allocation DWBA scheme for H2M telesurgery system based on SDN integrated optical access network is proposed in [7,8]. Moreover, Edge computing intelligently allocates computing and storage resources to devices at the edge of the wired and wireless network to reduce latency and jitter. Despite increasing interest in deep learning for optical access networking, the potential of edge intelligence to improve immersive and transparent H2M operation for human operators remain largely unexplored. Therefore, we use

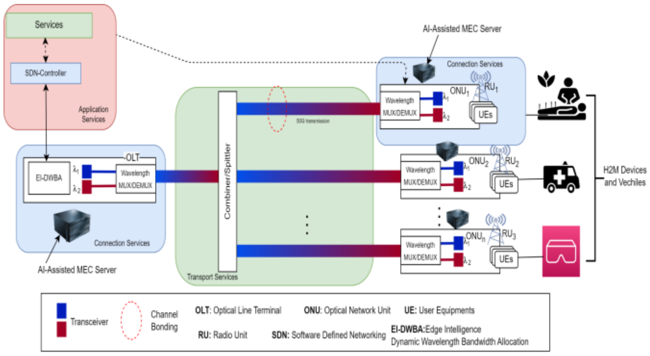

Fig. 2. Edge intelligence supporting H2M optical mobile fronthaul networks.

deep learning at the edge of our communication network to create immersive, seamless H2M experiences. Furthermore, in this study machine learning based DBA schemes are investigated in [9]. In this paper, in order to meet H2M application requirements, we propose a novel edge intelligence DWBA (EI-DWBA) mechanism for NG-EPON based on deep learning. Specifically, we construct a gated recurrent network (GRU)-Recurrent Neural Network (RNN) and predict the bandwidth.

II. LOW-LATENCY OPTICAL MOBILE FRONTHAUL ACCESS NETWORK ARCHITECTURE

Figure. 2 shows the proposed edge intelligence based integrated optical mobile fronthaul access system architecture for H2M applications. The optical fiber backhaul consists of a IEEE 802.3ca 25G NG-EPON with typical fiber range 20 km between the central office (CO) optical line terminal (OLT) and optical network unit (ONU) integrated radio unit (RU). Each stage of NG-EPON is separated by a splitter/combiner or wavelength multiplexer/demultiplexer. The proposed architectural design involves the bonding of two 25G wavelength channels, resulting in a combined transmission capacity of up to 50G. In order to facilitate the allocation of bandwidth in the specified wavelength channels, each optical network unit (ONU) is equipped with a pair of transceivers, denoted as *λ1* and *λ2*. The connection between the OLT and the ONUs is conducted via the multi-point control protocol (MPCP), enabling the ONU to send a REPORT message to the OLT, informing it about its upstream bandwidth demands. In response, the OLT would send a GATE message to the ONU granting it a time interval. Moreover, the OLT and ONU-RUs are enhanced with SDN and AI-assisted MEC server and divided into three different services such as application service, connection service and transport services.

The application service consists of a service module and SDN controller module. The service module is responsible for achieving differentiated services (i.e. H2M applications, cloud data center and machine learning etc.) for the clients. The SDN controller module is responsible for transmitting and receiving control packet data between client-side applications and the SDN controller. The SDN controller interfaces with OLTs and ONUs to gain insight into and manage the entire access network.

The connection services comprise the OLT, ONU-RRUs and AI-assisted MEC server. The OLT is outfitted with a pair of transceivers operating at distinct wavelengths, denoted as λ1 and λ2. The ONU-RU establishes a connection between subscribers and the OLT through the utilization of two transceivers that have the capability to be bonded. Moreover, the OLT has managed the DWBA scheme to allocate bandwidth to ONU-RUs. These AI-MEC servers are equipped with OLT and ONU-RUs based on cost, bandwidth, and latency constraints. Note that the H2M teleoperation services contain the master and slave devices, these devices are connected to an ONU of long-reach PON [10]. Such remote H2M teleoperation can be facilitated by various AI-enhanced edge-computing servers to minimize its latency. Therefore, the system uses AI-enhanced edge-computing servers to predict bandwidth requirements.

Transport services would combine application and connectivity services in a hybrid access network. NG-EPON intelligently supports network slices in different systems, applications, and vendors using SDN. Two bonded transceiver channels allow ONUs to transmit up to 50G to an OLT.

A. *Edge-Intelligence DWBA (EI-DWBA) Mechanism*

The AI-assisted MEC server plays a critical role in the SDN-enhanced NG-EPON system, particularly in the DWBA allocation process. In our proposed EI-DWBA schemes the AI-MEC server will provide the GRU-RNN model to predict and allocate the bandwidth. Gated recurrent unit (GRU) is an enhancement of recurrent neural network (RNN). GRU is able to solve problems involving long-term dependencies because it retains the knowledge acquired during earlier stages of the learning process. There are two main gates in the GRU namely the update gate and reset gate. The update gate U_t is used to control how much information from the previous state is brought into the current state. The reset gate R_t is used to control how much the GRU ignores the previous moment's status information [11].

The NG-EPON bandwidth assignment, the grant/report mechanism is implemented using GATE and REPORT messages exchanged between OLT and ONUs. The communication made between the OLT and ONU based on its cycle time. In this way we applied the GRU algorithm to predict bandwidth demand for next P cycles based on its demands in the past R cycles. The bandwidth demands can be collected in two ways. First, the ONU buffer occupancy report messages R and second OLT GATE message G in every cycle time t. These two-time series messages contain essential information regarding the operational standing of every ONU. Therefore, an EI-DWBA needs the two kind of cycles; the ONU buffer occupancy reporting cycle $(R_1, R_2, \dots\dots, R_r)$ and the prediction cycle $(P_1, P_2, \dots\dots, P_p)$. During cycles, each ONU sends a report message to request bandwidth allocation, which is determined by its queue size, and the OLT calculates and sends the grant for the next cycles. This is similar to conventional dynamic wavelength bandwidth allocation (DWBA) methods. But, the OLT records the above request for predictions. In the last reporting cycle R_r, when the OLT receives the R^{th} REPORT message from the ONU, it utilizes the

R saved requests of this ONU as input to the GRU model in order to predict the request sizes for the future P cycles. When the GRU model generated predictions, and the OLT possesses these predicted request sizes for all ONUs, these predictions are treated as though they were REPORT messages received from the ONUs across the cycles $(P_1, P_2, \ldots\ldots, P_p)$. After prediction, the OLT uses the same DWBA operation used to assign transmission windows to each ONU for cycles $(P_2, \ldots\ldots, P_{p,}\, P_{1,})$ and does not require any more report messages. This prediction mechanism reduces the control message overhead and increases the system throughput.

III. PERFORMANCE EVALUATIONS

We collect training data and run extensive OPNET simulator simulations to test the proposed system. The proposed approach uses 64 ONU and OLT with two transceivers. OLT and ONU downlink/uplink channel rates were dynamically assigned 1-25 Gbps. OLT and ONUs were evenly spaced at 10–20 km, and the ONU buffer was 10 Mb. Network traffic models for HT, AF, and BE traffic were self-similarity and long-range dependence, with maximum transmission cycles of 1.0 ms. The model generated high-burst HT, AF and BE traffic with a hurst parameter of 0.7 and a packet size uniformly distributed between 64 and 1518 bytes. The traffic on the EF was based on a Poisson distribution with a fixed packet size (70 bytes). The network traffic ratios of scenario 1, Expedited Forwarding-EF (10%), H2M Traffic-HT (6%) Assured Forwarding-AF (34%) and Best effort (50%); scenario 2 (EF-10%, HT-7.5%, AF-42.5% and BE-40%); scenario 3 (EF-10%, HT-9%, AF-51% and BE-30%), respectively.

The dataset was collected offline with limited DWBA schemes and the traffic models are generated for EF-Poisson distribution, HT-Pareto, AF and BE are uniform distribution. The data set has 9 features: EF Report, HT Report, AF Report, BE Report, EF Grant, HT Grant, AF Grant, BE Grant, and Cycle time. This means observations are taken every 1 ms. Our dataset includes 2 lakh data point samples from all network loads. The training dataset will comprise 80% of the rows from the original data, while the testing dataset will comprise the remaining 20%.

In addition, we analyses the performance of the proposed EI-DWBA under a offline limited scheme and contrast it to a conventional limited scheme using a different of traffic profiles to evaluate its efficacy. Fig. 3 shows the mean packet delay of EI-DWBA vs. limited DWBA scheme. The packet delay includes polling, grant, and queueing delays. The EI-DWBA may greatly improve compared to the limited DWBA in all traffic profiles in terms of packet delay. Using historical data, the EI-DWBA can reduce control message overheads by predicting the EF, HT, AF, and BE REPORT messages next steps in advance.

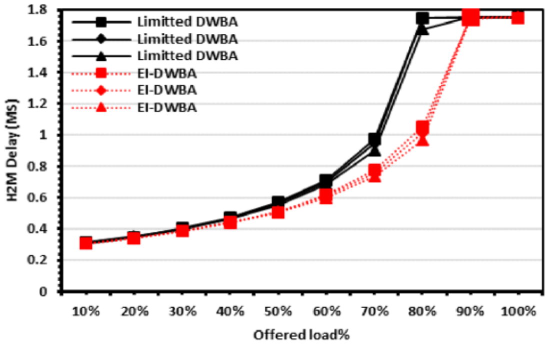

Fig. 3. Mean packet delay for H2M traffic in 1.0 ms.

IV. CONCLUSION

The main objective of this paper is to facilitate H2M applications with an AI-MEC server over optical mobile fronthaul networks. Our contribution is more encouraging for future H2M immersive research on edge-intelligence communications. Moreover, we proposed a edge intelligence DWBA scheme for improving the bandwidth efficiency of NG-EPON by utilizing the GRU-RNN model. These proposed models significantly improve the prediction accuracy, and reduce the loss to improve the system performance.

REFERENCES

[1] L. Ruan, M. P. I. Dias and E. Wong, “Enhancing latency performance through intelligent bandwidth allocation decisions: a survey and comparative study of machine learning techniques,” *Journal of Optical Communications and Networking*, vol. 12, no. 4, pp. B20-B32, April 2020.

[2] E. Ganesan, A. T. Liem and I.-S. Hwang, “QoS-aware multicast for crowdsourced 360° live streaming in SDN aided NG-EPON,” *IEEE Access*, vol. 10, pp. 9935-9949, Jan. 2022.

[3] Y. Nakayama and D. Hisano, “Wavelength and bandwidth allocation for mobile fronthaul in TWDM-PON,” *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7642-7655, Nov. 2019.

[4] M. R. Araujo, B. R. Silva, J. de Santi, G. B. Figueiredo and N. L. S. da Fonseca, “A novel prediction-based dwba algorithm for ng-epon based C-RAN fronthaul,” *IEEE International Conference on Communications (ICC)*, Shanghai, China, 2019, pp. 1-6.

[5] S. B. Hussain, W. Hu, H. Xin and A. M. Mikaeil, “Low-latency dynamic wavelength and bandwidth allocation algorithm for NG-EPON,” in *Journal of Optical Communications and Networking*, vol. 9, no. 12, pp. 1108-1115, Dec. 2017.

[6] J.-Y. Lee, I.-S. Hwang, A. T. Liem, K. R. Lai and A. Nikoukar, “Genetic expression programming: A new approach for QoS traffic prediction in EPONs,” *Photonic Network. Communication*, vol. 25, no. 3, pp. 156-165, Jun. 2013.

[7] E. Ganesan, I.-S. Hwang, A. T. Liem, and M. S. Ab-Rahman, “5G-enabled tactile internet resource provision via software-defined optical access networks (SDOANs),” *Photonics*, vol. 8, no. 5, p. 140, Apr. 2021.

[8] C. Li, W. Guo, W. Wang, W. Hu and M. Xia, “Programmable bandwidth management in software-defined EPON architecture”, *Optical Communication*, vol. 370, pp. 43-48, Jul. 2016.

[9] E. Wong, S. Mondal and L. Ruan, “Machine learning enhanced next-generation optical access networks-challenges and emerging solutions Invited Tutorial],” *Journal of Optical Communications and Networking*, vol. 15, no. 2, pp. A49-A62, Feb.2023.

[10] S. Mondal and E. Wong, “Global-local ai coordinated learning over optical access networks for scalable H2M/R collaborations,” *IEEE Network*, vol. 36, no. 2, pp. 124-130, Mar/Apr. 2022.

[11] E. P. Lens Shiang, W. -C. Chien, C. -F. Lai and H. -C. Chao, “Gated recurrent unit network-based cellular trafile prediction,” *International Conference on Information Networking (ICOIN)*, Barcelona, Spain, 2020, pp. 471-476.