IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 July 2023, accepted 21 July 2023, date of publication 26 July 2023, date of current version 9 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3299037

RESEARCH ARTICLE

A Novel Temporal Dynamic Wavelength Bandwidth Allocation Based on Long-Short-Term-Memory in NG-EPON

ANDREW TANNY LIEM [1], I-SHYAN HWANG [2], ELAIYASURIYAN GANESAN [2], SEMMY WELLEM TAJU [1], AND GREEN ARTHER SANDAG [1]

[1]Department of Computer Science, Universitas Klabat, Airmadidi 95371, Indonesia
[2]Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 32003, Taiwan

Corresponding authors: Andrew Tanny Liem (andrew.heriyana@unklab.ac.id) and I-Shyan Hwang (ishwang@saturn.yzu.edu.tw)

This work was supported in part by the National Science and Technique Council under Grant 112-2221-E-155-009.

ABSTRACT Optical networks have undergone a remarkable transformation with the adoption of Artificial Intelligence (AI) techniques such as Machine Learning (ML) and Deep Learning (DL). The Next Generation (NG)-EPON is one such technology that is essential for supporting high-bandwidth applications like 4K video streaming, ultra-high-definition (UHD) CCTV, and other emerging video-type applications that have strict Quality-of-Service (QoS) requirements. In this paper, we present a ground-breaking Temporal Dynamic Wavelength Bandwidth Allocation (T-DWBA) mechanism based on the Long-Short-Term-Memory (LSTM) architecture. The T-DWBA uses past experiences to learn data as knowledge and predict time series with time lags of unknown size. Our proposed mechanism reduces upstream control message overheads, eliminates idle periods, and significantly improves bandwidth utilization, ensuring superior QoS specifically for video-type applications. The simulation results show that the T-DWBA significantly enhances system performance, reducing packet delay, and jitter, and improving bandwidth utilization. The use of AI techniques like ML and DL coupled with the recent advancements in SDN-Enabled Broadband Access (SEBA), hardware/software, and cloud technologies provide the perfect platform for deploying our proposed T-DWBA mechanism. Overall, our research proposes a promising solution for boosting EPON performance, revolutionizing optical networks, and providing seamless access to high-quality video streaming for a next-generation audience.

INDEX TERMS NG-EPONs, T-DWBA, LSTM, SEBA, system performance.

I. INTRODUCTION

Video is the 5G "killer app" for businesses and consumers. Moreover, due to the COVID-19 pandemic, consumers and enterprises have seen video capabilities as a significant benefit of 5G. In [1], Nokia discovered that while 90 percent of consumers would consider video to be a valuable component of 5G, just 83 percent of enterprises saw it as a compelling 5G use case. Cisco predicts that 5G speeds will be 13 times faster by 2023, while fixed broadband speeds will double. Average speeds for 5G and broadband are expected to reach 575Mbps and 110.4Mbps, respectively, [2]. As these rates increase, there is a growing opportunity for video use cases.

The associate editor coordinating the review of this manuscript and approving it for publication was Petros Nicopolitidis.

However, the introduction of Internet-enabled Ultra-High-Definition (UHD) or 4K video streaming has doubled the HD video bit rate and is nine times higher than the Standard-Definition (SD) video bit rate. Cisco predicts that by 2023, two-thirds (66%) of flat-panel TV sets will be UHD installations, up from 33% in 2018. Furthermore, the increasing use of Machine-to-Machine (M2M) connections in many industries is also contributing to the demand for bandwidth. Although there may be fewer M2M connections compared to end-user devices such as smartphones, TVs, and PCs, the traffic is growing at a faster rate due to the deployment of video applications on M2M connections, telemedicine, and intelligent car navigation systems.

The Cisco Annual Report predicts that video and other applications will continue to be in high demand in

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

average households. However, future application requirements, such as UHD camera security and streaming, VR streaming, cloud gaming, 8K wall TV, and HD/UHD VR streaming, will significantly increase the need for more bandwidth. Additionally, the rise of immersive services like remote learning, telepresence, and interactive graphical applications, live 360 video streaming, volumetric video, immersive gaming, and live immersive production services such as free-viewpoint video (FVV) [3], can potentially strain the network. As a result, the access network may be further impacted by these new immersive video applications and advancements in backbone network technology [2].

Passive Optical Networks (PONs) have emerged as a promising solution for addressing access network gaps. Ethernet Passive Optical Network (EPON) technology, known for its cost-efficiency, high bandwidth efficiency, and efficient QoS support, has been widely recognized as one of the best PON technologies available. The IEEE 802.3ca standard, approved in 2020 as the next-generation EPON (NG-EPON), enables the expansion of EPONs to multiple channels of 25 Gbps, facilitating data transmission at various rates downstream and upstream [4]. The NG-EPONs can distribute massive bandwidth and channel capacity, broaden coverage, and speed up data transfer. Furthermore, it also can deliver reliable and efficient communication compared to the legacy/old EPON system.

EPON can utilize different multiplexing techniques such as Time-Division-Multiplexing (TDM), Wavelength-Division-Multiplexing (WDM), and Time-Wavelength-Division-Multiplexing (TWDM), which combines the strengths of TDM and WDM. The IEEE and FSAN groups chose the TWDM approach for NG-EPON based on technology, performance, and cost-efficiency. TWDM distributes time slot frames to subscribers via wavelength channels, offering TDM's capability and WDM's flexibility [5].

The IEEE 802.3ca working group has defined NG-EPONs [4], which can provide a cost-effective and practical solution for future EPONs, increasing the bandwidth fivefold of a single channel to 25*Gbps*. Furthermore, the channel bonding allows the NG-EPON to achieve higher data rates that can provide aggregated data rates of N x 25*Gbps*. Figure 1 shows the architecture of a 25G NG-EPON with two 25G wavelength channels $\{\lambda 1, \lambda 2\}$ that are bonded to achieve transmission of up to 50G between Optical Network Units (ONUs) and an Optical Line Terminal (OLT). Connecting several channels is recommended for increased data speeds and flexibility, but it should be noted that the cost of an ONU increases when more transceivers are deployed in the ONUs [6]. It is important to consider these factors when designing NG-EPON networks.

The NG-EPON architecture integrates the advantages of both TDM and WDM by transmitting frames to ONUs via multiple wavelengths [7], surpassing all other access technologies. Each wavelength pair can provide a data rate of 25Gbps for downstream and upstream transmission, resulting in a full operational NG-EPON with up to 50Gbps

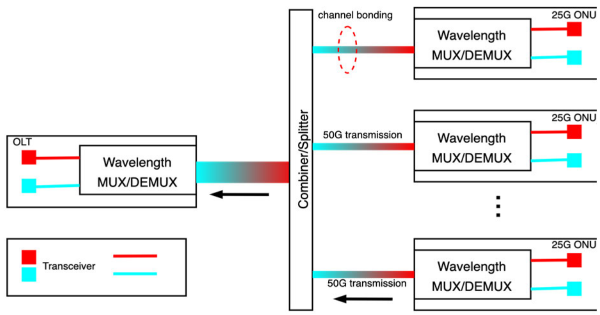


FIGURE 1. Generic NG-EPON architecture.

for downstream and 50Gbps for upstream transmission [8]. Although the IEEE 802.3ca NG-EPON specifies two wavelength channels, the NG-EPON standards are not limited to a single generation. Therefore, subsequent generations beyond 100G will also be developed. Additionally, these multiple generations of NG-EPONs will require different wavelength deployment schemes and should coexist on the same network to achieve cost-effective transmission. Hence, a proper wavelength deployment scheme must be selected when designing future coexisting NG-EPONs [8].

Furthermore, optical amplifiers (OAs) and WDM Mux/Demux are located on the OLT side, creating in a passive optical distribution network (ODN) [9]. The OAs on the OLT side amplify the downstream signal and pre-amplify the upstream signals. In addition, the transceiver in OLTs should be able to adapt to changing traffic conditions. For example, to optimize energy consumption, if traffic loads are below 25%, the OLT's transceiver can be configured to activate only one of the two transceivers available in the ONU [10].

In NG-EPON, several issues must be addressed: critical enabling technologies and device upgrades, security assurance concerns due to the downstream broadcasting nature, eavesdropping, and Dynamic Wavelength and Bandwidth Allocation (DWBA) optimized bandwidth utilization. DWBA is a mechanism used by the OLT to dynamically allocate channel bandwidth between multiple wavelengths, using time slots for upstream data transmission. Each ONU can transmit the data at its specified wavelength during the assigned time. DWBA helps prevent data transmission conflicts between different ONUs [11].

Furthermore, three different DWBA mechanisms have been identified to address the traffic behavior issue: 1) wavelength-agile (WA)-EPON; 2) single-scheduling domain (SSD)-EPON; and 3) multi-scheduling domain (MSD)-PON. The conventional DWBA mechanism can be applied to the SSD-PON or MSD-EPON, but not WA-PON. MSD-EPON enables an ONU to transmit its desired traffic on a single wavelength channel at a time, or multiple ONUs simultaneously transmit on different wavelength channels. In SSD-EPON, data transmission by multiple ONUs is not possible on the same channel simultaneously.

TABLE 1. Comparison of existing studies.

Papers	Technology	DWBA Mechanism	Prediction Model	DiffServ	SDN	Applications
[17]	Hybrid TDM/WLAN PON	DBA	ANN	Yes	No	H2M
[18]	NG-EPON	DBA	LSTM	No	No	Not specified
[19]	PON	DBA	XGBoost Learning Algorithm	No	No	Not specified
[20]	WDM/TDM-PON	DWBA	Error back propagation NN	Yes	Yes	Not specified
[21]	PON	No, only prediction model	LSTM	No	No	Not specified
[22]	GPON	No, only prediction model	RNN	No	No	Not specified
[23]	PON	No, only prediction model	LSTM	No	No	Not specified
[24]	XG-PON	DBA	LSTM	No	No	Not specified
This paper	**NG-EPON**	**DWBA**	**LSTM Multistep**	**Yes**	**Yes**	**Video Applications**

In NG-EPONs, the Multipoint MAC Control (MPMC) sublayer reconciles the 25Gbps or 50Gbps PON into the Ethernet framework. The MPMC consists of two main protocols: the Multipoint Control Protocol (MPCP), responsible for arbitration of TDM-based access to the Point-to-Multi-Point (P2PM) medium. The second protocol is Channel Control Protocol (CCP), which is responsible for querying and controlling multiple channels within Nx25G-EPON PHY. The MPCP is responsible for timing and arbitrating the ONU transmissions. The arbitration is achieved by allocating transmission windows (GATE Message) to ONUs. On the other hand, the ONU will transmit its queued at the total line rate based on the reporting queue occupancy state using a REPORT message. The NG-EPONs do not specify specific bandwidth allocation strategies, Quality of service (QoS) definitions, provisioning, or management.

Therefore, numerous early studies have been conducted to increase bandwidth efficiency by proposing a predictive Dynamic Bandwidth Allocation (DBA) or DWBA with various approaches such as constant predictive, credit-based predictive, genetic expression programming [12], [13], [14], or using multiple statistical predictive techniques [15] (i.e., Bayesian estimation and others) or using multiple combinations of double-phase polling, shortest propagation delay, and excess distribution [16]. While these early approaches can significantly improve bandwidth efficiency and QoS, they rely solely on estimating the number of arriving packets for bursty traffic remains an open challenge. Similarly, [17] argues that statistical predictive approaches suffer from increased latency in predicting instantaneous traffic variations.

In recent times, optical networks have embraced the use of Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), to address complex classification and prediction challenges in Dynamic Bandwidth Allocation (DBA) and enhance Quality of Service (QoS). Notably, a number of cutting-edge studies employing ML/DL in DBA have been conducted (as shown in Table 1) [17], [18], [19], [20], [21], [22], [23], [24]. For example, [21], uses the k-Nearest Neighbors (k-NN) method to adaptively tune neighbors dynamically, which is essential in dynamic environments with frequent changes that affect accuracy.

Moreover, the progress in Software-Defined Networking (SDN), Graphics Processing Unit (GPU) technology, and cloud computing has played a pivotal role in enabling the necessary processing and storage capabilities for training resource-intensive ML models such as DL. SDN, as a set of design principles that structure the development of new abstraction layers, encompasses three key architectural principles: (1) separation of control and data plane functions, (2) logical centralization of control, and (3) network function programmability [25]. These principles greatly contribute to the widespread adoption of ML and DL in optical networks, offering increased flexibility and efficiency.

Several recent studies have explored the use of machine learning techniques such as LSTM models to predict access network traffic, without implementing DBA or considering DiffServ [21], [22], [23], [24]. Another study proposed a DWBA algorithm for WDM/TDM-PONs that leverages neural networks to predict network traffic and reduce RTT delay, while improving bandwidth efficiency [20]. The OLT controller predicts the next queue size by evaluating the previous load condition and median queue size. However, this study did not consider any specific application for their proposed DBA. Table 1 provides a summary of the current state-of-the-art studies.

Nevertheless, in contrast, our work addresses this challenge by developing a robust LSTM Multistep prediction model specifically tailored for immersive video applications. By incorporating sophisticated algorithms and techniques, our model enables accurate predictions, thereby enhancing bandwidth efficiency and ensuring a seamless and immersive video experience. Moreover, while previous studies have explored the use of ML techniques like LSTM models to predict access network traffic, they have not fully implemented DBA or considered DiffServ. In our research, we not only leverage the predictive capabilities of our LSTM model but also integrate it seamlessly into the DWBA mechanism within the NG-EPON architecture. This integration ensures optimal resource allocation, improved network performance, and enhanced user experiences in the context of immersive video applications.

In our previous work [26], we demonstrated that applying ML algorithms such as the k-Nearest Neighbors algorithm (k-NN), Random Forest, Decision Tree, and Naive Bayes achieves higher accuracy in predicting GATE/REPORT message and queue length regardless of cycle time. However, the basic machine learning model requires some guidance or human intervention when it makes an incorrect prediction. In contrast, a DL architecture was introduced, which uses its neural network to determine whether a prediction is correct or not. Although ML and DL are often used interchangeably, ML employs a collection of algorithms to analyze and interpret data, learn from it, and make the best possible decisions based on the learned information. In contrast, DL architectures are divided into multiple layers to create an artificial neural network that can self-learn from complex and high-dimensional data and make intelligent decisions based on the data.

Additionally, a study presented in [18] introduced an LSTM neural network to predict future bandwidth requirements for multiple polling cycles, effectively reducing bandwidth waste and enhancing EPON performance. To the best of our knowledge, there is no current research proposing an LSTM-based DWBA under a limited scheme that includes DiffServ. Our work is similar to the aforementioned LSTM-based DBA, which leverages sequential information instead of the traditional neural network's assumption that inputs and outputs are independent of one another. In a real-time network environment, traffic demand can change at any time, and therefore, retaining temporal and historical user data can aid the network in predicting end-user bandwidth requirements.

The contribution of this paper is a novel Temporal Dynamic Wavelength Bandwidth Allocation based on Long-Short-Term-Memory in NG-EPONs, referred to as T-DWBA. T-DWBA is a bandwidth allocation method in NG-EPONs that utilizes LSTM to enhance bandwidth efficiency (i.e., reduce control message overheads) and utilization, leading to a better QoS, particularly for upcoming video applications. The OLT has incorporated an LSTM that predicts ONUs bandwidth demand for C cycles based on the past R ONUs REPORT messages. Moreover, the proposed integration of AI techniques, SDN-Enabled Broadband Access (SEBA), and cloud technologies architecture provides a promising platform for deploying and revolutionizing optical networks and enabling seamless access to high-quality video streaming for next-generation applications.

The remainder of this paper is organized in the following manner. The LSTM architecture is presented in Section II. T-DWBA's model is described in Section III. Section IV discusses the performance evaluation and simulation. Section V brings our work to a conclusion.

II. LONG-SHORT-TERM MEMORY ARCHITECTURE

LSTM, a boost for conventional RNNs [27], was created to solve the problems of vanishing and exploding gradients [28]. Unlike traditional RNNs, which rely on passing outputs between neural networks for prediction improvement, LSTMs utilize previous data. However, LSTMs face challenges, such as the squishing effect of sigmoid functions that can cause the vanishing gradient problem during backpropagation. Additionally, traditional RNNs lack long-term memory, storing only short-term memory. LSTMs address this by incorporating both short-term and long-term memory, combining them at each stage to generate new memory and predictions. Consequently, LSTMs excel in preserving past information, enabling neural networks to remember historical data.

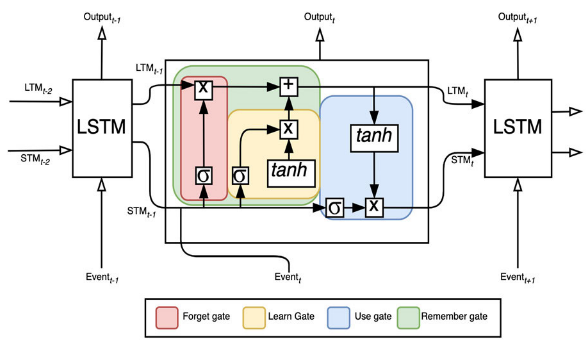


FIGURE 2. Generic architecture of LSTM.

Figure 2 shows the generic architecture of the LSTM model, where the long-term memory (LTM) and short-term memory (STM) from the previous time sequence *t-2*, i.e., LTM$_{t-2}$ and STM$_{t-2}$, are fed into the model [29]. An event and an output are received and produced in the current time sequence *t-1*, i.e., LTM$_{t-1}$ and STM$_{t-1}$, which are then passed to the next node. This process is repeated, allowing the LSTM to keep track of the LTM and STM. The updated output from LTM$_{t-2}$ and STM$_{t-2}$ and the prediction Output$_{t-1}$ are used to generate the following output of LTM$_{t-1}$ and STM$_{t-1}$. The LSTM architecture includes four gates: the Learn Gate, the Forget Gate, the Remember Gate, and The Use Gate. The Learn Gate takes the STM and the new input information (E) and combines them, retaining only the essential parts. The output of the Learn Gate is $N_t i_t$, where

$$\begin{aligned} N_t &= tanh\left(W_n\left[STM_{t-1}, E_t\right] + b_n\right) \\ i_t &= \sigma\left(W_i\left[STM_{t-1}, E_t\right] + b_i\right) \end{aligned} \tag{1}$$

Moreover, the Forget Gate takes an LTM$_{t-1}$, deciding what parts to keep and forget. The output of the Forget Gate is $LTM_{t-1}f_t$, where

$$f_t = \sigma\left(W_f\left[STM_{t-1}, E_t\right] + b_f\right) \tag{2}$$

The Forget gate in the LSTM module determines which information to keep or forget based on the output value. If the value is closer to 0 means to forget, and closer to one means to retain the information.

The Remember Gate combines the LTM from the Forget Gate and the STM from the Learn Gate. Therefore, the output of Remember Gate is

$$LTM_{t-1}f_t + N_t i_t, \tag{3}$$

where N_t, i_t and f_t are calculated in Eq. (1) and (2). Finally, the Use Gate will take what is helpful from the LTM and STM and update the STM_t, thus the output of the Use Gate is $U_t V_t$, where

$$\begin{aligned} U_t &= tanh\left(W_u LTM_{t-1}f_t + b_u\right) \\ \sigma V_t &= \left(STM_{t-1}, E_t + b_v\right) \end{aligned} \tag{4}$$

Table 2 summarizes the symbols that are used in this paper.

TABLE 2. Summary of symbol.

Symbol	Description
LTM	Long-term memory
STM	Short-term memory
E	New input information
Nt	Output of the Learn Gate
i_t	Input gate output
f_t	Output of the Forget Gate
Ut	Output of the Use Gate
Vt	Predicted ONU Gate
Wn	Weight matrix for Nt
Wi	Weight matrix for i_t
Wf	Weight matrix for f_t
Wu	Weight matrix for u_t
B_n	Bias term for Nt
b_i	Bias term for i_t
b_f	Bias term for f_t
b_u	Bias term for u_t
sigma	Sigmoid function
tanh	Hyperbolic tangent function

A. LSTM MODEL FOR DWBA MECHANISM

In this section, we describe how the LSTM architecture is used in the DWBA prediction mechanism. The goal of using the LSTM architecture in DWBA is to predict ONU GATE messages based on the long-term history of ONU REPORT messages (LTM), the short-term or recent ONU REPORT messages (STM), and the current ONU REPORT messages (E). For example, let's assume a new ONU REPORT message E arrives at time t (Et). To predict the GATE message for the ONU (Output$_t$), the LSTM module takes the LTM$_{t-1}$, STM$_{t-1}$, and Et to form a prediction of ONU GATE messages at time sequence t. The LTM$_{t-1}$ and STM$_{t-1}$ provide hints or predictions of EF, AF, and BE bandwidth in time sequence t. Additionally, the LTMt and STMt modules are updated with these three pieces of information (LTM$_{t-1}$, STM$_{t-1}$, and Et).

 # Initialize the LSTM parameters
W_n, W_i, W_f, W_u = initialize_weights 0
b_n, b_i, b_f, b_1, b_v = initialize_biases 0
 # Initialize the LSTM memory and states
LTM$_{t-2}$, STM$_{t-2}$ = initialize_memory_states 0
LTM$_{t-1}$, STM$_{t-1}$ = initialize_memory_states 0
 # Define the LSTM model
function LSTM (Et):
  # Calculate the Learn Gate output
  N_t = tanh (W_n* concatenate (STM$_{t-1}$, Et) + b_n)
  i_t = sigmoid (Wi* concatenate (STM$_{t-1}$, Et) + b_i)
  # Calculate the Forget Gate output
  f_t = sigmoid (Wf* concatenate (STM$_{t-1}$, Et) + b_f)
  # Calculate the Remember Gate output
  LTM$_t$ = LTM$_{t-1}$ *f_t + N_t * i_t
  # Calculate the Use Gate output
  Ut = tanh (Wu * concatenate (LTM$_t$, f_t) + b_u)
  Vt = sigmoid (STM$_{t-1}$ + Et + b_v)
  # Update the memory and state for the next time sequence
  LTM$_{t-2}$ = LTM$_{t-1}$
  STM$_{t-2}$ = STM$_{t-1}$
  LTM$_{t-1}$ = LTM$_t$
  STM$_{t-1}$ = Ut * Vt
return Vt

In this pseudocode, $\boldsymbol{E_t}$ is the current ONU REPORT message, and $\boldsymbol{V_t}$ is the predicted ONU GATE message for the current time sequence $\boldsymbol{t}$. The LSTM model takes in the previous memory states (**LTM**$_{\boldsymbol{t-2}}$ and **STM**$_{\boldsymbol{t-2}}$) and the current memory states (**LTM**$_{\boldsymbol{t-1}}$ and **STM**$_{\boldsymbol{t-1}}$) to generate the next memory states. The **concatenate** function is used to combine the LTM and STM with the input sequence for each gate calculation. The high-level overview of how the proposed LSTM Model for DWBA mechanism is as follows:

1. At the start of each cycle of the SSD frame, the OLT collects information on the traffic demand and other relevant metrics from the ONUs.
2. The OLT feeds this information into an LSTM neural network, which has been trained to predict the optimal allocation of time slots for each ONU based on past and current traffic patterns.
3. The LSTM neural network processes the input data and generates a set of output values, each representing the number of time slots to be allocated to a particular ONU.
4. The OLT uses these output values to generate grant messages, which are sent to the ONUs to inform them of their allocated time slots.
5. The ONUs use the allocated time slots to transmit data to the OLT.
6. As the cycle progresses, the OLT continues to collect data on the traffic demand and other relevant metrics, and feeds this data back into the LSTM neural network to update its predictions for the next cycle.

III. PROPOSED ARCHITECTURE

The proposed NG-EPON (Next-Generation Ethernet Passive Optical Network) architecture, outlined in section A,

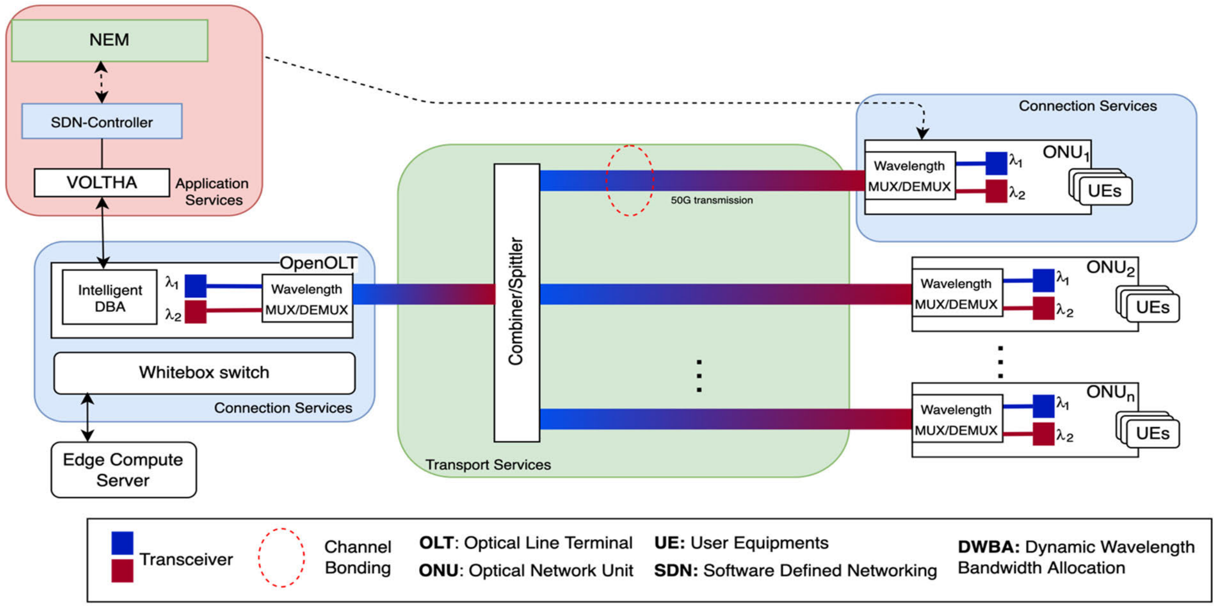


FIGURE 3. **Proposed High-Level SEBA architecture of NG-EPON with cloud-Based Ml engine.**

incorporates a cloud-based machine-learning engine along with two essential components discussed in subsequent sections. Section B introduces the cloud-based ML engine and edge cloud compute server, which leverage cloud computing resources to enable advanced data processing and machine learning capabilities at the network edge. Section C focuses on the utilization of sliding windows within the architecture, allowing for efficient data processing and analysis by dynamically adjusting the window size to capture temporal patterns. Lastly, section D introduces the Temporal-Dynamic Wavelength and Bandwidth Allocation (T-DWBA) mechanisms, which consider the temporal characteristics of network traffic and optimize resource allocation in real-time using the power of the cloud-based ML engine and sliding windows. Together, these components form an integrated framework that enhances the performance and efficiency of the NG-EPON system through cloud-based machine learning, advanced data processing, temporal analysis, and adaptive resource allocation.

A. HIGH-LEVEL OF THE PROPOSED NG-EPON ARCHITECTURE WITH CLOUD-BASED MACHINE-LEARNING ENGINE

The DBA in NG-EPON works cyclically, and the maximum cycle time can be determined by the size of the REPORT message from ONUs or can be limited by the OLT by setting up the entire cycle time (e.g., 1.5 *ms* or 2 *ms*). In other words, time series forecasting can be applied to this DBA mechanism. It is important to note that relying solely on the previous cycle to predict the GATE message for the next cycle is not effective, as users' traffic behavior in access networks varies over time. For example, ONUs serving office buildings may have higher bandwidth demands during the day than ONUs serving residential areas. Additionally, traffic demand can fluctuate due to specific events, such as lockdowns caused by COVID-19, which may result in more employees working from home. Therefore, our proposed DWBA considers not only the past ONUs REPORT messages but also utilizes them to predict the GATE message for multiple future cycles.

Figure 3 illustrates the high level of the proposed NG-EPON with a Cloud-Based ML Engine. The OLT and ONUs are enhanced with Software Defined Networking (SDN) and divided into three different services, namely application services, connection services, and transport services. The application services comprise a cloud-based ML engine, SDN controller, and DWBA module. The connection services consist of the OLT and ONUs. Lastly, the transport services are responsible for transporting services over NG-EPON using Channel bonding technology that can bond several wavelength channels to achieve high peak data rates [6]. As depicted in Fig. 3, two 25G wavelength channels are bonded to provide an aggregate transmission of up to 50G. To support the bandwidth assignment in these two wavelength channels, each ONU has two transceivers, i.e., $\{\lambda 1, \lambda 2\}$. The OLT employs the Channel Control Protocol (CCP) to activate or modify the state of ONU's upstream or downstream channel(s) through the exchange of CC_REQUEST and CC_RESPONSE Channel Control Protocol Data Unit (CCPDU). Afterward, the OLT identifies the upstream channel(s) granted to the ONU in a given GATE Multi-Point Control Protocol Data Unit (MPCPDU). Furthermore, the Single Scheduling Domain (SSD) mechanism is selected in

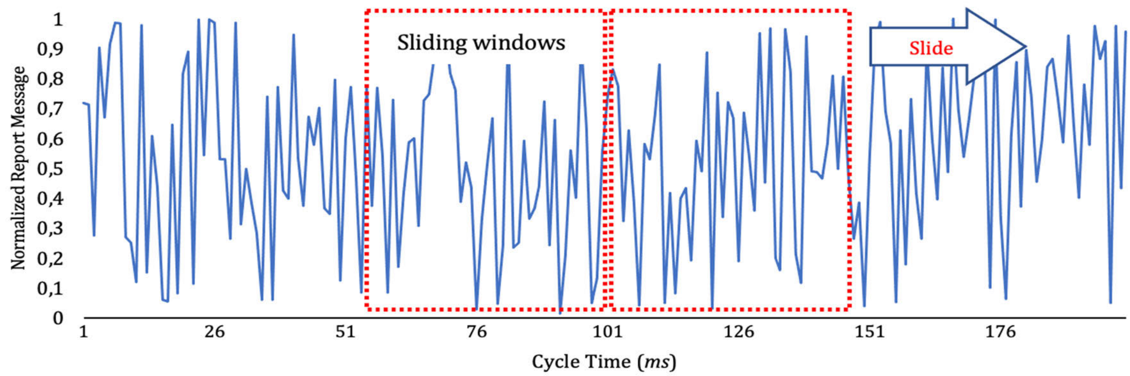


FIGURE 4. Illustration of multistep time series LSTM of ONU report message.

our architecture. Each ONU must transmit on all upstream wavelengths [8].

1) APPLICATION SERVICES

The application services consist of a VOLTHA, SDN controller, and NEM module. The OLT can obtain all the ONUs REPORT messages since the DBA offline approach is utilized. The SDN controller is used to produce an intelligently governed network capable of supporting network slices for different technologies. With SEBA, operators can use open-source software to build a flexible and programmable access network that supports advanced traffic management and QoS capabilities [30]. By integrating LSTM into the SEBA framework, operators can use predictive analytics to optimize the allocation of bandwidth in NG-EPON networks. LSTM can be used to analyze traffic patterns and predict future demand, enabling operators to allocate bandwidth dynamically in real-time. This can help to improve network efficiency, reduce congestion, and improve the user experience.

2) CONNECTION SERVICES

The OLT is equipped with two transceivers with different wavelengths (λ_1 and λ_2). The ONU connects subscribers to the OLT via two transceivers that can be bonded. The LSTM model can identify and learn the deep numerical dependencies of multistep time series, allowing it to predict future network behavior with high accuracy based on historical network granularity. Some studies have shown that the LSTM proved efficient for predicting network traffic granularity [18], [24], [26]. Lastly, the DWBA module is to deliver efficient packets and data. The scheduler based on Machine Learning will be designed.

The offline scheduling is chosen so that the OLT can have all the traffic patterns of each ONUs, store them in the ML-Engine, and use this data to improve the prediction model. All the information, such as traffic patterns and ML-Engine, is informed to the OLT by the SDN controller, which is orchestrated with the SDN app in the application services [31]. In this way, the DWBA is always adapted according to the network conditions. Consequently, the proposed LSTM-based DWBA model is adaptive, meaning it will be acting according to the behavior or characteristics in the optical distribution networks (ODN).

3) TRANSPORT SERVICES

The transport services would integrate all services (application and connection services) in a single hybrid access network. NG-EPON uses the SDN to adaptively support network slices in various systems, applications, and vendors. The ONUs have two bonded transceiver channels to provide aggregate transmission of up to 50G between ONUs and an OLT.

B. CLOUD-BASED ML ENGINE/EDGE CLOUD COMPUTE SERVER

The cloud-based ML engine plays a critical role in the SDN-enabled NG-EPON architecture, particularly in the dynamic wavelength and bandwidth allocation process. The ML engine utilizes the historical data of the network to predict the required bandwidth for each ONU in the next cycle. To achieve this, the ML engine employs a Long Short-Term Memory (LSTM) algorithm. The proposed LSTM algorithm uses sliding windows to process the historical data, where each window contains the bandwidth utilization data of all ONUs for a certain period (see Fig. 4). The ML engine then analyzes the data within the window to predict the required bandwidth for the next cycle. The size of the window determines the accuracy of the prediction, whereas a larger window size may result in more accurate predictions at the cost of increased computational complexity.

To ensure that the ML engine is always up-to-date with the latest network data, the cloud-based ML engine is periodically updated. However, updating the ML model in real-time may introduce delays and impact the performance of the network. To address this issue, the proposed approach involves updating the LSTM model in the cloud, rather than in real-time within the network. This means that the early version of the LSTM model will be used to conduct DWBA, while the cloud ML engine continuously updates the model with new data to improve its accuracy. At predetermined intervals, the cloud ML engine will send the updated LSTM model back to the NG-EPON network, which will then replace the old model with the new one. This approach helps to reduce the delay in the system, as the updates to the LSTM model are done offline in the cloud, rather than in real-time within the network.

Figure 4 employs a multi-step LSTM time-series data prediction model using sliding windows with the concept of SSD. The basic idea behind using sliding windows is to have a fixed-size window that slides over the past ONUs REPORT messages in order to capture the most recent traffic patterns. This window can then be used to feed into an LSTM model to predict the GATE message for the next cycle. In the context of SSD, each ONU must transmit on all upstream wavelengths. This means that the sliding window should also take into account the past REPORT messages from all the ONUs in the same scheduling domain, which can be viewed as a single entity for the purpose of bandwidth allocation. To ensure that the sliding window covers a sufficient amount of data, its size should be set to be at least as large as the maximum cycle time. In addition, since the proposed mechanism uses the SSD mechanism, which requires each ONU to transmit on all upstream wavelengths, the sliding window should also cover all upstream wavelengths for each ONU. Therefore, by combining these two, i.e., the LSTM model can better capture the traffic patterns across all ONUs in the scheduling domain, leading to more accurate predictions of the GATE message. Additionally, this approach can be adapted to handle fluctuations in traffic patterns over time such as those caused by changes in user behavior or external events like COVID-19 lockdowns.

C. SLIDING WINDOWS

For instance, let's say, we have time series data of bandwidth demand over a period of 24 hours, with 50-cycle time intervals between each data point. The LSTM model is trained to predict future demand based on the previous data points. We want to use this model to conduct DWBA with sliding windows approach. Firstly, we set the window size to 50 cycle data points. This means we will use the previous 50 cycles' data points to predict the next data point, and this window will slide every 50 cycles times to update the prediction. At time t, the LSTM model takes in the previous 50 cycle data points, E_{t-50} to E_t, and produces a predicted demand value Vt. Afterward, we update the LSTM model with the new data point, E_{t+1}, and the window slides to include data points from E_{t-49} to E_{t+1}. We repeat the process for the next time interval, and the window slides include data points from E_{t-48} to E_{t+2}, and so on. By using the sliding window approach, we can continuously update the LSTM model with new data points and adjust the bandwidth allocation in real time based on the predicted demand. This allows for more efficient use of network resources and reduces the likelihood of network congestion.

D. TEMPORAL-DWBA (T-DWBA) MECHANISMS

The overhead associated with the employed DBA scheme can vary. Online schedulers, such as IPACT [32], schedule grants on-the-fly without waiting for all REPORT messages to be received. This eliminates idle time and reduces packet delays. However, ensuring fairness among ONUs and supporting QoS can be difficult because the OLT lacks a holistic view of all ONU demands. This problem is addressed by offline schedulers where the OLT waits for all REPORT messages before performing DBA and scheduling grants. This enables the OLT to support QoS and fosters fairness among ONUs, but at the cost of increased channel utilization due to control overhead between transmission cycles.

In traditional NG-EPON bandwidth assignment, the grant/report mechanism is implemented using GATE and REPORT messages exchanged between OLT and ONUs [33]. Control channel overhead is introduced by the GATE and REPORT messages, and it is affected by the number of scheduled ONUs and the cycle time. The proposed T-DWBA aims to improve QoS and reduce the overhead associated with the control channel in the traditional grant/report mechanism. Other overhead components in NG-EPON include burst mode overhead, Forward-Error-Correction (FEC) encoding overhead, guard-band overhead, etc. Specifically, the T-DWBA can reduce the downstream overhead by predicting the upcoming traffic demands of ONUs and allocating the bandwidth resources accordingly. By using a sliding window technique and LSTM network to analyze the traffic history, the T-DWBA can predict the upcoming demands with high accuracy. In contrast, IPACT DBA uses a simple threshold-based method to allocate bandwidth, which does not take into account the traffic history or the future demand. This often leads to underutilization or overutilization of the bandwidth, resulting in higher overhead. For simplicity, let's assume that the downstream overhead can be calculated by

$$Downstream\ Overhead = (G/T)x100,$$

where G represents the total number of grant messages sent by the OLT to the ONUs. T represents the total number of time slots used in the downstream cycle 100 is a constant used to express the overhead as a percentage. Assuming a cycle time of 1 ms, ONUs generating traffic demand according to a Poisson process with an average rate of 500 Mbps, sliding window size of 10, stride of 5, and an LSTM model with 2 hidden layers of 32 units each trained for 10 epochs, Table 3 shows a comparison of the downstream overhead improvement between T-DWBA and IPACT. To guarantee a delay of 1.5 ms for voice traffic in the access network, the cycle time should be about 1 ms [34].

IV. EXPERIMENT AND EVALUATION

A. DATASET

The proposed T-DWBA mechanism is validated and evaluated in this section using an OPNET simulator with 64 Software-Defined (SD)-ONUs and an SD-OLT with two wavelengths to run a comprehensive simulation and

TABLE 3. **Overheads comparison of T-DWBA vs. IPACT.**

TRAFFIC LOAD	IPACT	LSTM	IMPROVEMENTS
10%	3.05%	2.14%	29.75%
20%	4.68%	3.55%	24.06%
30%	5.91%	4.59%	22.19%
40%	7.38%	5.63%	23.60%
50%	8.27%	6.68%	19.18%
60%	10.07%	7.72%	23.42%
70%	11.27%	8.76%	22.31%
80%	12.64%	9.82%	22.34%
90%	13.91%	10.86%	21.93%
100%	15.39%	11.94%	22.47%

TABLE 4. **Simulation parameters.**

Parameters	Values
Number of ONU	64 ONUs
Number of wavelengths	2
Up/Down link-rate	25Gbps
OLT-ONU distance	Uniform 10-20km
Maximum transmission cycle time	1ms
Guard band	5µs
DWBA computation time	10μs
MPCP transmission time	0.512μs

create the dataset. The SD-OLT and SD-ONU have downstream/upstream channel rates of 25 Gbps, and they were evenly distributed over a range of 10 km to 20 km, with the ONU buffer size set at ten megabits. The maximum transmission cycle was set at 1.0 ms, and the AF and BE traffic network traffic models were set as self-similarity and long-range dependence [35]. To generate high-burst AF and BE traffic, the model used a Hurst parameter of 0.7 and a packet size uniformly distributed between 64 and 1,518 bytes. The EF traffic was modeled using a T1 circuit-emulated line with IP/UDP encapsulation and a constant frame rate of 1 frame/125μs with a fixed packet size of 70 bytes.

This work aims to validate the practicality of the proposed T-DWBA mechanism by generating traffic for the limited offline DWBA, which is the most extensively used legacy in the DBA mechanism. The traffic generator installed at each ONU develops a self-similar and long-range dependent network traffic model for AF and BE traffic, which has the potential to create a significant amount of Internet traffic in a short period. The simulation produced approximately 250,000 data points, including cycle duration, EF, AF, BE REPORT message, GATE message, and the number of wavelengths. To ensure the quality of the dataset, the acquired data was pre-processed to identify and remove any missing, inconsistent, or noisy data.

The pre-processed data was then converted to a matrix to evaluate the proposed T-DWBA mechanism. The matrix was converted to a vector, and the vectors were concatenated to obtain the $N \times N$ traffic-over-time matrix M. The M matrix was divided into two sections: 80% of training M_{train} and 20% of testing $M_{testing}$, which were used to train and validate the proposed T-DWBA based on the LSTM model. Finally, the data was normalized by dividing it by the maximum value. Table 4 summarizes the simulation parameters. More than 250.000 data points were returned from the simulation, including cycle time, EF, AF, BE REPORT message, GATE message, and the number of wavelengths.

B. EVALUATION

The dataset contains seven features: EF Report, AF Report, BE Report, EF Grant, AF Grant, BE Grant, and Cycle time. Each feature was collected every 1 ms. We used 24 hours of observations to train the LSTM regressor for our prediction model. To avoid biased performance estimates, we removed duplicate entries that can ruin the split between train and test sets. From the original data, the training dataset will make up 80% of the rows, while the testing dataset will make up the remaining 20% of approximately 250.000 data points. We chose three critical features, namely AF GATE, BE GATE, and Cycle-Time as the feature index since the T-DWBA must assign bandwidth to all ONUs.

We used the mean and standard deviation to standardize the dataset. The multi-Step model predicts a range of future values, i.e., AF GATE and BE GATE, based on the given past historical data of EF REPORT, AF REPORT, and BE REPORT. In other words, the multi-step model predicts a sequence of the future. In our model, the T-DWBA learns to predict the AF GATE and BE GATE for multistep cycles ahead based on the sliding windows of history cycles. The T-DWBA based on the LSTM model should be designed to be less complex but with higher accuracy to produce speedy learning. Because complex models require more processing and take longer to learn, they may be inefficient.

As the proposed T-DWBA model is designed to forecast multi-steps, we use two LSTM layers along with a dense layer. The Mean Squared Error (MSE) is used to measure the difference between the predicted and actual values of EF Grant, AF Grant, and BE Grant. We use RMSProp, an optimizer that combines gradient descent and AdaGrad, to train the models. Moreover, we adjust the hyperparameters of the LSTM network accordingly. The proposed LSTM algorithm needs to learn from the entire training data, so we train the models for 100 epochs (refer to Table 5 for details). We use TensorFlow and Keras backends along with Scikit-learn (Sklearn) tools for building and training the LSTM models as our regressors.

Figure 5 shows the actual versus predicted features of AF and BE grant messages in the testing dataset over time, which were used by the LSTM regressor to make predictions. The training and validation losses are metrics used to evaluate how well an LSTM model fits the training data, as well as to assess the performance of the validation set by summing the errors for each traffic type in the datasets (the training and validation sets of GATE messages for AF and BE traffic types). The actual AF grant (AF grant real test set) refers to the GATE message for the AF traffic type, while the predicted AF grant (AF grant predicted test) refers to the expected GATE

TABLE 5. Performance comparison on a different number of epochs using the proposed LSTM model.

Epochs	MAE	MSE	RMSE	R-Squared (%)
5	0.1	0.02	0.15	75%
10	0.08	0.02	0.14	76%
20	0.05	0.01	0.11	88%
30	0.05	0.01	0.10	89%
40	0.06	0.02	0.13	80%
50	0.02	0.0	0.03	99%
60	0.03	0.0	0.06	96%
70	0.03	0.01	0.07	94%
80	0.04	0.01	0.11	86%
90	0.05	0.02	0.14	78%
100	0.06	0.03	0.16	72%

message for the AF traffic type generated by the proposed LSTM model. The model successfully predicted future steps of the AF and BE grant with a loss of less than 0.02. The loss can be seen as the distance between the valid values of the actual Grant Messages and the values predicted by the model.

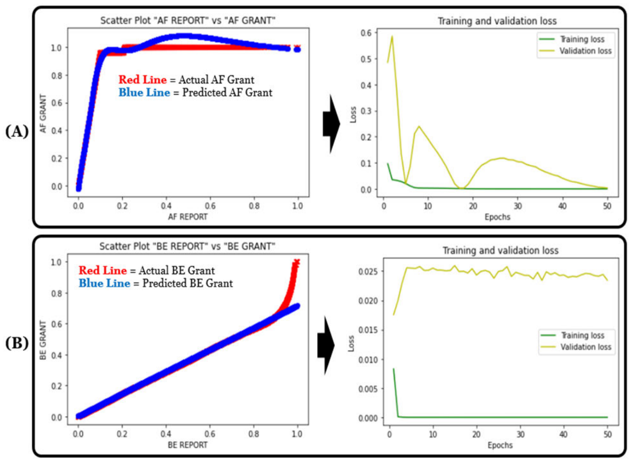


FIGURE 5. Predicted vs. Actual AF and BE grant messages features of testing dataset across time using LSTM regressor.

C. PERFORMANCE EVALUATION

In this section, we analyze the performance of the proposed T-DWBA under a limited scheme and compare it to a typical limited scheme, with various traffic profiles to evaluate its effectiveness. Notably, to the best of our knowledge, there is currently no research that proposes a Machine Learning-based DWBA under a limited scheme that considers DiffServ. Although previous studies such as [17], [18], and [36] have utilized Machine Learning techniques, they do not account for DiffServ, which is a crucial component in providing Quality-of-Services (QoS) in current IP networks [37]. To address this, we conduct an extensive simulation with different scenarios, including Traffic profile 1 (TP1), Traffic profile 2 (TP2), and Traffic profile 3 (TP3), which have varying levels of EF, AF, and BE upstream traffic. Additionally, we incorporate heavy loads in the simulation to provide a comprehensive performance assessment, which can aid in network engineering and user provisioning [18].

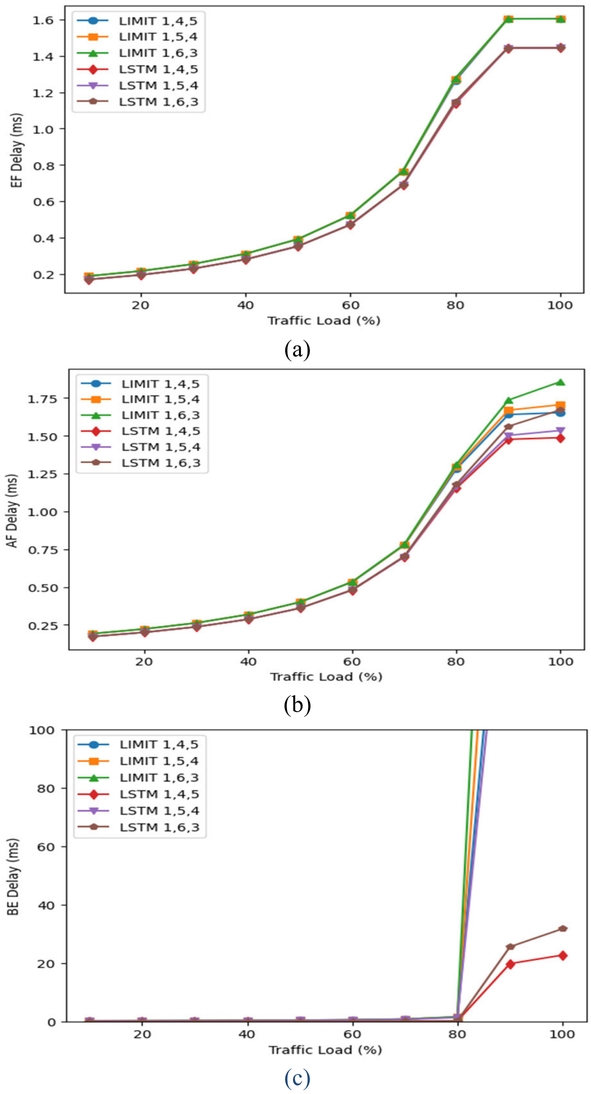


FIGURE 6. Packet delay (a) EF Delay; (b) AF Delay; (c) BE Delay.

The mean packet delay of the T-DWBA vs. limited scheme is depicted in Fig. 6(a)-(c). The packet delay comprises three components: polling delay, grant delay, and queueing delay. As can be seen, the T-DWBA may significantly improve the limited DWBA in all traffic profiles in terms of packet delay, thanks to the LSTM-based prediction of future bandwidth requirements. This helps reduce bandwidth waste due to overheads and idle periods, by adding extra timeslots to each ONU's bandwidth requirements. Additionally, the LSTM model allows for more accurate prediction of bandwidth requirements, leading to reduced queueing delay and improved packet transmission rates. It is worth noting that the simulation depicts the network under extreme load conditions, i.e., greater than 80% load. When traffic loads exceed 80%, the BE packet delay becomes saturated since limited or LSTM-based DWBA prioritizes the EF and AF packets, forcing the BE packet to wait for the next cycle time or until there are available timeslots to transmit.

Apart from packet delay, jitter is critical for the network's temporal performance, as high latency makes interactive applications such as voice and two-way video ineffective. Moreover, jitter also affects all real-time communication, such as video conferencing and VoIP conversations, streaming media, online video games, and desktop-as-a-service employing virtual desktop infrastructure (VDI) [38]. A good jitter performance should be 50*ms* or less for triple-play services.

Figure 7 depicts the jitter performance of EF and AF packets when the traffic loads are highly loaded. As can be seen, the jitter performance of T-DWBA has a minimal delay variation in all scenarios. For instance, a stable network typically experiences a 15% or less jitter percentage. The jitter of TP1, TP2, and TP3 in T-DWBA for EF and AF traffic percentage is below 5%, meaning that the packets of EF and AF are almost transmitted at equal intervals so that the users can experience seamless communications or video streams.

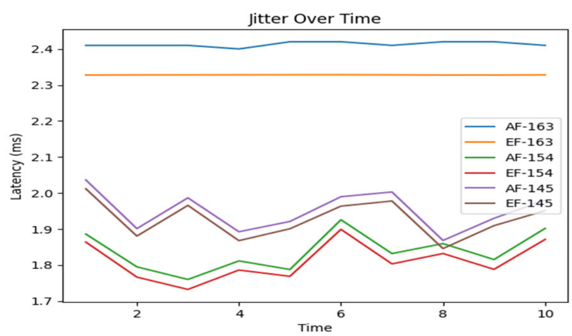


FIGURE 7. **Jitter performance of proposed T-DWMA.**

Figure 8 illustrates the total system throughputs of TP1, TP2, and TP3 as a function of the offered load. The study defines system throughput as the number of packets transmitted by ONUs with two wavelengths. Cycle time, unused residual, and guard time all affect system throughputs. Because the T-DWBA can forecast multiple steps, upstream overheads such as guard time can be eliminated, enhancing upstream bandwidth efficiency. As illustrated in Fig. 8, the T-DWBA system's bandwidth utilizations exceed 89%, whereas the Limit system's bandwidth utilizations are 84%.

In our analysis, we simulated various hurst parameters for video with 4K resolution i.e., 2160p to evaluate their impact on our proposed architecture (see Fig. 9). For pre-recorded video playback or offline video streaming with less bursty traffic patterns, a hurst parameter of 0.3 is suitable. On the other hand, video-sharing platforms like YouTube or Vimeo, where users upload and share videos, benefit from a hurst parameter of 0.5, which accounts for moderate burstiness caused by content and user interactions. When it comes to real-time streaming of events, sports, or live video broadcasts,

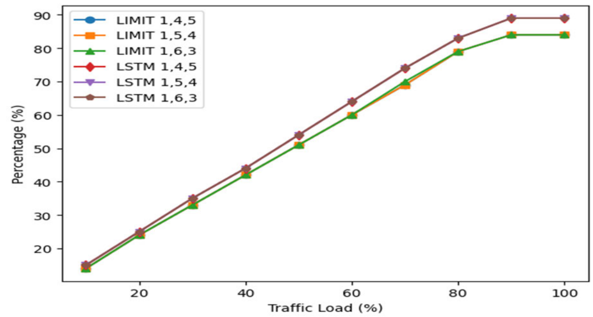


FIGURE 8. **System throughputs.**

such as YouTube Live or Twitch, a hurst parameter of 0.7 is preferred due to the burstiness resulting from dynamic content and viewer engagement. Finally, interactive and real-time experiences like video conferencing or VR/AR require a hurst parameter of 0.8. By considering these different hurst parameters, we can assess their respective effects on our proposed architecture.

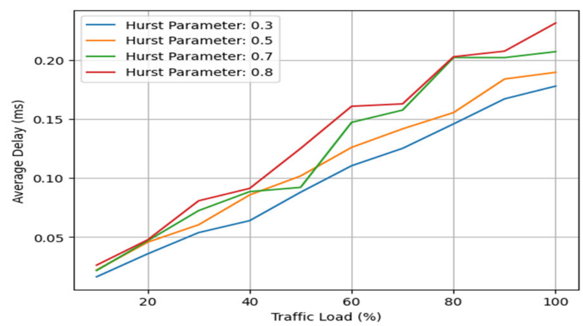


FIGURE 9. **Average delay for different hurst parameters.**

The simulation results reveal that the average delay increases with higher traffic load and Hurst parameter values, indicating increased congestion and queuing delays. Moreover, the impact of bursty traffic, represented by the Hurst parameter, becomes more significant at higher traffic loads. Bursty traffic patterns with higher Hurst parameters exhibit stronger long-range dependence, resulting in longer delays. However, even in the face of higher traffic loads, the proposed LSTM DWBA (Dynamic Wavelength and Bandwidth Allocation) method demonstrates its effectiveness in maintaining the Quality of Service (QoS) requirements for video traffic. By leveraging LSTM models, the method intelligently allocates wavelengths and bandwidth resources to mitigate delays and ensure satisfactory video streaming experiences. This highlights the potential of advanced techniques in managing bursty and high-load traffic scenarios while meeting the specific QoS demands of video applications, paving the way for improved network performance and reliable video services in demanding environments.

Finally, these simulation results demonstrate that the proposed T-DWBA architecture based on LSTM is adaptable to various network environments and settings. As shown in Table 6, the proposed LSTM model in deep learning was compared with other regression algorithms such as Decision Tree Algorithm, k-NN Algorithm, Linear Regression (linear function for regression), Random Forest, AdaBoost Algorithm, Gradient Boosting, and Multi-layer Perceptron (MLP). The proposed T-DWBA architecture based on LSTM model can solve the problems of non-linear data from previous AF REPORTs and BE REPORTs to better predict the AF GATE and BE GATE for multistep cycles in this regression analysis. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used to measure the error rate of regression models. However, we compared those models with errors measured in the same units. During the experiments, the proposed LSTM model outperforms other regressors for predicting AF GATE, achieving the minimum MAE of 0.02 and RMSE of 0.03, and for predicting BE GATE, achieving the minimum MAE of 0.03 and RMSE of 0.06. The cloud-based deep learning engine using LSTM that performs the function of building the model can store data created by real-world traffic or simulations and improve the performance predictions of traditional regression models.

TABLE 6. Performance evaluation using different types of regression algorithms.

Regressor Type	AF Grant		BE Grant	
	MAE	RMSE	MAE	RMSE
Decision Tree Algorithm	0.40	0.69	1.23	2.13
k-NN Algorithm	0.29	0.57	1.16	3.01
Linear Regression	2578.12	3108.67	1066.45	1491.59
Random Forest	554.6	756.20	1613.65	1882.15
AdaBoost Algorithm	236.08	257.01	429.27	516.13
Gradient Boosting	9.66	19.98	40.65	53.70
Multi-layer Perceptron	2714.75	3157.57	6118.39	7260.00
LSTM	0.02	0.03	0.03	0.06

V. CONCLUSION

In conclusion, the proposed T-DWBA scheme based on LSTM for NG-EPON, integrated with the SDN-based SEBA architecture, shows great promise in addressing the challenges of QoS provisioning in modern communication networks. The LSTM model demonstrates high accuracy with minimal MSE loss, enabling accurate forecasting of future bandwidth requirements using historical data. The T-DWBA protocol, in combination with the SEBA architecture, provides reliable QoS in terms of packet delay, jitter, and system throughput even under high network loads. However, it is important to acknowledge that this study does not explore the computational and resource costs associated with implementing the proposed scheme. Future work should investigate the potential computational overhead and resource requirements, particularly considering the increasing sophistication of DWBA algorithms and the emergence of cloud-based machine learning engines. This will contribute to a more comprehensive understanding of the practical feasibility and scalability of the T-DWBA scheme integrated with the SEBA architecture in NG-EPON and other communication networks, providing valuable insights for further improvements in the field of DWBA.

REFERENCES

[1] S. Condon. *Video is the 5G 'Killer App' for Both Enterprise and Consumers, Survey Shows*. ZDNet. Accessed: Aug. 8, 2021. [Online]. Available: https://www.zdnet.com/article/video-is-the-5g-killer-app-for-both-enterprise-and-consumers-survey-shows/

[2] *Cisco Annual Internet Report—Cisco Annual Internet Report (2018–2023) White Paper*. Cisco. Accessed: Aug. 8, 2021. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[3] P. Pérez, D. Corregidor, E. Garrido, I. Benito, E. González-Sosa, J. Cabrera, D. Berjón, C. Díaz, F. Morán, N. García, J. Igual, and J. Ruiz, "Live free-viewpoint video in immersive media production over 5G networks," *IEEE Trans. Broadcast.*, vol. 68, no. 2, pp. 439–450, Jun. 2022, doi: 10.1109/TBC.2022.3154612.

[4] *IEEE Standard for Ethernet Amendment 9: Physical Layer Specifications and Management Parameters for 25 Gb/s and 50 Gb/s Passive Optical Networks*, Standard 802.3ca-2020, 2020, doi: 10.1109/IEEESTD.2020.9135000.

[5] A. Rafiq and M. Hayat, "QoS-based DWBA algorithm for NG-EPON," *Electronics*, vol. 8, no. 2, p. 230, Feb. 2019, doi: 10.3390/electronics8020230.

[6] C. Zhang, M. Yang, W. Zheng, Y. Zheng, Y. Wu, and Y. Zhang, "Analysis of wavelength deployment schemes in terms of optical network unit cost and upstream transmission performance in NG-EPONs," *J. Opt. Commun. Netw.*, vol. 13, no. 9, p. 214, Sep. 2021, doi: 10.1364/JOCN.425722.

[7] H. S. Abbas and M. A. Gregory, "The next generation of passive optical networks: A review," *J. Netw. Comput. Appl.*, vol. 67, pp. 53–74, May 2016, doi: 10.1016/j.jnca.2016.02.015.

[8] L. Wang, X. Wang, M. Tornatore, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee, "Dynamic bandwidth and wavelength allocation scheme for next-generation wavelength-agile EPON," *J. Opt. Commun. Netw.*, vol. 9, no. 3, pp. B33–B42, Mar. 2017, doi: 10.1364/JOCN.9.000B33.

[9] Y. Luo, X. Zhou, F. Effenberger, X. Yan, G. Peng, Y. Qian, and Y. Ma, "Time- and wavelength-division multiplexed passive optical network (TWDM-PON) for next-generation PON stage 2 (NG-PON2)," *J. Lightw. Technol.*, vol. 31, no. 4, pp. 587–593, Feb. 15, 2013, doi: 10.1109/JLT.2012.2215841.

[10] A. F. Pakpahan, I.-S. Hwang, and A. A. Nikoukar, "OLT energy savings via software-defined dynamic resource provisioning in TWDM-PONs," *J. Opt. Commun. Netw.*, vol. 9, no. 11, pp. 1019–1029, Nov. 2017, doi: 10.1364/JOCN.9.001019.

[11] A. Rafiq, M. F. Hayat, and M. U. Younus, "Efficient bandwidth management (EBM) algorithm for NG-EPON," in *Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2018, pp. 663–667, doi: 10.1109/CompComm.2018.8780835.

[12] C.-H. Chen, H.-T. Wu, and K.-W. Ke, "Predictive credit based dynamic bandwidth allocation mechanisms in Ethernet passive optical network," in *Proc. IEEE Region 10 Conf.*, Mar. 2006, pp. 1–4, doi: 10.1109/TENCON.2006.343756.

[13] 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," *Photon. Netw. Commun.*, vol. 25, no. 3, pp. 156–165, Jun. 2013, doi: 10.1007/s11107-013-0399-x.

[14] J.-R. Lai and W.-P. Chen, "High utilization dynamic bandwidth allocation algorithm based on sorting report messages with additive-polling thresholds in EPONs," *Opt. Switching Netw.*, vol. 18, pp. 81–95, Nov. 2015, doi: 10.1016/j.osn.2015.04.003.

[15] M. P. I. Dias, B. S. Karunaratne, and E. Wong, "Bayesian estimation and prediction-based dynamic bandwidth allocation algorithm for sleep/doze-mode passive optical networks," *J. Lightw. Technol.*, vol. 32, no. 14, pp. 2560–2568, Jul. 15, 2014, doi: 10.1109/JLT.2014.2327629.

[16] M. P. McGarry and M. Reisslein, "Investigation of the DBA algorithm design space for EPONs," *J. Lightw. Technol.*, vol. 30, no. 14, pp. 2271–2280, Jul. 12, 2012, doi: 10.1109/JLT.2012.2196023.

[17] L. Ruan, M. P. I. Dias, and E. Wong, "Machine learning-based bandwidth prediction for low-latency H2M applications," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3743–3752, Apr. 2019, doi: 10.1109/JIOT.2018.2890563.

[18] J. A. Hatem, A. R. Dhaini, and S. Elbassuoni, "Deep learning-based dynamic bandwidth allocation for future optical access networks," *IEEE Access*, vol. 7, pp. 97307–97318, 2019, doi: 10.1109/ACCESS.2019.2929480.

[19] B. Cao, X. Zheng, K. Yuan, D. Qin, and Y. Hong, "Dynamic bandwidth allocation based on adaptive predictive for low latency communications in changing passive optical networks environment," *Opt. Fiber Technol.*, vol. 64, Jul. 2021, Art. no. 102556, doi: 10.1016/j.yofte.2021.102556.

[20] B. Liu, L. Zhang, F. Wang, M. Liu, Y. Mao, L. Zhao, T. Sun, and X. Xin, "Adaptive dynamic wavelength and bandwidth allocation algorithm based on error-back-propagation neural network prediction," *Opt. Commun.*, vol. 437, pp. 276–284, Apr. 2019, doi: 10.1016/j.optcom.2018.12.064.

[21] P. Sarigiannidis, D. Pliatsios, T. Zygiridis, and N. Kantartzis, "DAMA: A data mining forecasting DBA scheme for XG-PONs," in *Proc. 5th Int. Conf. Mod. Circuits Syst. Technol. (MOCAST)*, Thessaloniki, Greece, May 2016, pp. 1–4, doi: 10.1109/MOCAST.2016.7495169.

[22] A. Tomasov, M. Holik, V. Oujezsky, T. Horvath, and P. Munster, "GPON PLOAMd message analysis using supervised neural networks," *Appl. Sci.*, vol. 10, no. 22, p. 8139, Nov. 2020, doi: 10.3390/app10228139.

[23] Z. Yang, D. Zhang, and J. Tang, "Predicting PON networking traffic flow based on LSTM neural network with periodic characteristic data," in *Proc. IEEE 5th Optoelectron. Global Conf. (OGC)*, Shenzhen, China, Sep. 2020, pp. 39–42, doi: 10.1109/OGC50007.2020.9260441.

[24] M. Zhang, B. Xu, X. Li, Y. Cai, B. Wu, and K. Qiu, "Traffic estimation based on long short-term memory neural network for mobile front-haul with XG-PON," *Chin. Opt. Lett.*, vol. 17, no. 7, 2019, Art. no. 070603.

[25] A. S. Thyagaturu, A. Mercian, M. P. McGarry, M. Reisslein, and W. Kellerer, "Software defined optical networks (SDONs): A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2738–2786, 4th Quart., 2016, doi: 10.1109/COMST.2016.2586999.

[26] A. T. Liem, G. A. Sandag, I.-S. Hwang, and A. Nikoukar, "Delay analysis of dynamic bandwidth allocation for triple-play-services in EPON," in *Proc. 5th Int. Conf. Cyber IT Service Manage. (CITSM)*, Aug. 2017, pp. 1–6, doi: 10.1109/CITSM.2017.8089260.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[28] A. Azzouni and G. Pujolle, "A long short-term memory recurrent neural network framework for network traffic matrix prediction," 2017, *arXiv:1705.05690*.

[29] E. Ganesan, A. T. Liem, I.-S. Hwang, M. S. Ab-Rahman, S. W. Taju, and M. N. A. Sheikh, "LSTM-based DWBA prediction for tactile applications in optical access network," *Photonics*, vol. 10, no. 1, p. 37, Dec. 2022, doi: 10.3390/photonics10010037.

[30] S. Das, "From CORD to SDN enabled broadband access (SEBA)," *J. Opt. Commun. Netw.*, vol. 13, no. 1, pp. A88–A99, Jan. 2021, doi: 10.1364/JOCN.402153.

[31] M. Robinson, M. Milosavljevic, P. Kourtessis, G. P. Stafford, M. J. Burrell, and J. M. Senior, "Software defined networking for heterogeneous access networks," in *Proc. 18th Int. Conf. Transparent Opt. Netw. (ICTON)*, Trento, Italy, Jul. 2016, pp. 1–4, doi: 10.1109/ICTON.2016.7550412.

[32] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT a dynamic protocol for an Ethernet PON (EPON)," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 74–80, Feb. 2002, doi: 10.1109/35.983911.

[33] R. Roy, G. Kramer, M. Hajduczenia, and H. J. Silva, "Performance of 10G-EPON," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 78–85, Nov. 2011, doi: 10.1109/MCOM.2011.6069713.

[34] G. Kramer, "How efficient is EPON?" IEEE802.org, Tech. Rep., 2003. [Online]. Available: https://www.ieee802.org/3/efm/public/p2mp_email/pdf00001.pdf

[35] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994, doi: 10.1109/90.282603.

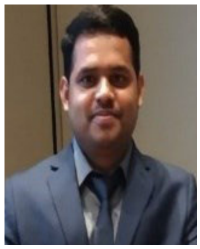
[36] J. A. Hernandez, R. Sanchez, I. Martin, and D. Larrabeiti, "Meeting the traffic requirements of residential users in the next decade with current FTTH standards: How much? How long?" *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 120–125, Jun. 2019, doi: 10.1109/MCOM.2018.1800173.

[37] *QoS: DiffServ for Quality of Service Overview Configuration Guide*, Cisco, San Jose, CA, USA, 2017.

[38] J. A. Hernandez, A. Ebrahimzadeh, M. Maier, and D. Larrabeiti, "Learning EPON delay models from data: A machine learning approach," *J. Opt. Commun. Netw.*, vol. 13, no. 12, pp. 322–330, Dec. 2021, doi: 10.1364/JOCN.437414.

ANDREW TANNY LIEM received the B.S. degree from the Department of Computer Science, Adventist University of Indonesia, Bandung, Indonesia, in 2003, the M.S. degree, in 2006, and the Ph.D. degree in computer science and engineering from Yuan-Ze University, Taiwan, in 2014. He is currently with the Department of Computer Science, Universitas Klabat, Manado, Indonesia. His recent work focuses on NGN, SDN, and P2P over EPON, intelligent networks, and fault tolerance.

I-SHYAN HWANG received the B.S. and M.S. degrees in electrical engineering and electronic engineering from Chung-Yuan Christian University, Chung-Li, Taiwan, in 1982 and 1984, respectively, and the M.S. and Ph.D. degrees in electrical and computer engineering from The State University of New York at Buffalo, NY, USA, in 1991 and 1994, respectively. Since February 2007, he has been a Full Professor with the Department of Computer Science and Engineering, Yuan-Ze University, Chung-Li. His current research interests include high-speed networks, fixed mobile convergence, and heterogeneous multimedia services over fiber optic networks, optical green computing, NGN, and optical networks-based infrastructure over cloud computing.



ELAIYASURIYAN GANESAN received the B.Sc. degree in computer science from the Tamilavel Umamakeshwaranar Karanthai Arts College, Thanjavur, Tamil Nadu, India, in 2010, and the M.C.A. degree from the Dhanalakshmi Srinivasan Engineering College, Perambalur, Tamil Nadu, in 2013. He is currently pursuing the Ph.D. degree with Yuan Ze University, Chung-Li, Taiwan. His research interests include high speed networks, FiWi, SDN, the IoT, tactile internet, and machine learning.

SEMMY WELLEM TAJU received the B.S. degree from the Faculty of Computer Science, Klabat University, Indonesia, in 2013, and the M.S. and Ph.D. degrees from the Department of Computer Science and Engineering, Yuan Ze University, Taiwan, in 2016 and 2020, respectively. He is currently with the Faculty of Computer Science, Klabat University, as a Lecturer. His research interests include machine learning, deep learning, and bioinformatics.

GREEN ARTHER SANDAG received the B.S. degree in computer science from Universitas Klabat, Airmadidi, Indonesia, in 2012, and the M.S. degree in computer science and engineering from Yuan Ze University, Taoyuan, Taiwan, in 2016. Since August 2016, he has been a Lecturer with Universitas Klabat. His research interests include computer vision and natural language processing, with a particular interest in vision and language topics, such as sentiment analysis, emotion classification, and image classification.