Design of Energy-Efficient EPON: A Novel Protocol Proposal and Its Performance Analysis

Sourav Dutta and Goutam Das

Abstract—Economical and environmental concerns necessitate network engineers to focus on energy-efficient access network design. The optical network units (ONUs) are predominant sources of the energy consumption of Ethernet passive optical network, which motivate us toward designing a novel protocol for saving energy at ONUs. The proposed protocol exploits different low power modes (LPMs) and opts for the suitable one using traffic prediction. This scheme provides a significant improvement of energy-efficiency especially at high load (∼40%) over existing ONU-assisted protocols. A better understanding of the performance and a deeper insight into several design aspects can only be addressed through a detailed mathematical analysis.

The proposed protocol involves traffic prediction, which infringes the Markovian property. However, some pragmatic assumptions along with a proper selection of observation instances and state descriptions allow us to form a discrete time Markov chain (DTMC) of the proposed algorithm. Thus, the primary objective of this paper is to propose a novel scheme for achieving energy-efficiency at ONUs and mathematically analyze the performance of it with the help of a DTMC. The analysis reveals that the energy-efficiency is more sensitive to the power consumption of doze mode as compared to other LPMs. Further, the effect of sleep-to-wake-up time of all LPMs is minor.

Index Terms—Energy-efficiency EPON, ONU-assisted.

I. INTRODUCTION

THE PERSISTENT enhancement of Internet traffic promotes research on diminishing the Internet power consumption. Owing to the primary source (80–90%) of Internet power consumption [1], research on energy-efficient Internet network design is predominantly targeted from the perspective of the access network. An Ethernet Passive Optical Network (EPON), one of the most widely accepted and deployed access technologies, consists of an Optical Line Terminal (OLT), multiple Optical Network Units (ONUs), and Remote Nodes (RNs) [2]. As ONUs are responsible for consuming the substantial fraction (∼70%) of the overall EPON power consumption, energy-efficient ONU design has already emerged to be a well-established research field. Energy-efficiency is achieved at ONUs by switching off some of its active components (low power mode). However, a certain time is required to switch-on those inactive components and this time is also known as sleep-to-wake-up time. It is widely understood that the power consumption figures of ONUs can be reduced by switching off more components whereas it increases the sleep-to-wake-up time leading to a trade-off between the power consumption and sleep-to-wake-up time. This trade-off allows designing different low power modes: doze mode, fast sleep, deep sleep, cyclic sleep [3]. In all those low power modes, it is essential to diminish the waste due to wake-up time in order to maximize energy-efficiency, which requires elongation of the time duration over, which low power mode is employed, termed as sleep duration. However, the limitation on the buffer size of ONUs or the presence of delay-sensitive traffic in an EPON network imposes an upper-bound on sleep durations. Due to the trade-off between the power consumptions and the wake-up time durations of different low power modes, the enhancement of energy-efficiency calls for an efficient protocol design for the proper selection of a low power mode as a function of the calculated upper-bound of the sleep duration. In addition, energy-efficiency at ONUs can be further enhanced by designing advanced circuitries that reduce power consumptions [4], [5] or wake-up time durations of different low power modes [5], [6]. Among them, it is essential to look for the parameters that deserve more attention in order to achieve maximum energy benefit. This requires the knowledge about the dependency of energy-efficiency on all of the concerned parameters, which can solely be obtained by a rigorous mathematical analysis of the underlying protocol. Thus, after designing an efficient protocol for saving energy at ONUs, the performance analysis of it is of utmost importance. In this paper, we focus on both of these aspects.

The protocols for energy-efficient ONU design in EPON are classified as OLT-assisted and ONU-assisted. In case of the OLT-assisted protocols, OLT decides sleep durations of ONUs and informs them through extra information within the GATE message [7]. Several OLT-assisted protocols have been proposed in [8]–[15] and [17]. Dixit et al. [8] have proposed an OLT-assisted protocol, termed as Sleep Mode Aware (SMA) protocol, where OLT send GATE message to every ONU in every polling cycle like traditional Dynamic Bandwidth Allocation (DBA) [2] protocol. However, before sending the GATE message, the OLT calculates the minimum time instant up to which the next GATE message will not be sent to that ONU and this time instant is informed to the ONU through the GATE message. During this time period,
ONUs employ a power saving mode. The effect of employing different power saving modes in this time period for different technologies is demonstrated in [9]. In SMA protocol, the Down-Stream (DS) traffic is queued and it is transmitted to an ONU only during its allocated transmission slot. Different schemes for calculating this transmission slot has been proposed in [10]. In the proposed protocol of [11], ONUs observe the DS traffic of some x number of cycles and if no packets arrive over this period then it sleeps for some y number of cycles. The same authors have extended the protocol for both Down-Stream (DS) and Up-stream traffic (U.S.) in [12]. Zhang et al. [11] and Hosseinabadi and Ansari [12] have mathematically analyzed the energy-efficiency figures of their proposed schemes with the aid of Discrete Time Markov Chain (DTMC). Shagdar et al. [13] have shown that a significant improvement in energy-efficiency can be achieved by selecting between two predefined sleep times (one short and one long) instead of using a single long sleep time. The proposed Green Bandwidth Allocation (GBA) protocol of [14] is analyzed by modeling buffer of each ONU as an M/G/1 queue with vacation [15]. Khotimsky et al. [17] have utilized the advantage of both doze mode and cyclic sleep and proposed a LPM (Watchful Sleep mode). In Watchful Sleep mode, three states have been introduced: Aware state (ONU is fully active), Listen state (ONU is in doze mode), and Sleep state (ONU is in cyclic sleep). An ONU cyclically alters between Aware state and Sleep state and during sleep period it periodically enters into Listen state when the handshaking with the OLT is performed. The performance analysis of this sleep mode has been proposed in [18]. The major drawback of all OLT-assisted protocols is that they require changing the MAC of both OLT and ONUs [16]. Thus, a prodigious amount of expenditure is required to employ an OLT-assisted protocol in a legacy network [16].

The drawbacks of OLT-assisted protocols can be eradicated by ONU-assisted protocols where an ONU itself decides its own sleep duration. Few ONU-assisted protocols have also been proposed in [16], [19], and [20]. However, to the best of our knowledge, mathematical analysis of none of the ONU-assisted protocols is present in the literature. Herreria-Alonso et al. [19] have employed only doze mode for saving energy. However, it is well known that the power consumption figure of doze mode is quite high [3]. In [16] and [20], we have proposed a novel protocol where ONUs dynamically swerve between the active mode and different low power modes where ONUs select the most suitable low power mode based on the sleep duration (time interval over which low power mode is employed) before entering into a low power mode. Once an ONU enters into a low power mode, a certain time interval is required to wake-up from the low power mode and to initiate the U.S. [2] transmission. Traffic arrival over this time period is predicted by Autoregressive-Moving Average (ARMA) model in order to minimize the packet drop probability.

In all existing ONU-assisted protocols, an ONU remains in a low power mode for a certain time interval and then wakes-up from it and transmits U.S. data for multiple cycles before entering into a low power mode again. After waking up from sleep, energy-efficient ONUs behave like traditional ONUs (i.e., both transmitter and receiver are active) till it sleeps once again and no energy saving is achieved over this time period (active period). It is well known that, in Multi-Point-Control-Protocol (MPCP) [7], standardized for realizing statistical multiplexing in EPON, the transmitter of an ONU remains idle during other ONUs U.S. transmissions. This creates an opportunity to employ the doze mode during these periods without losing any possibility of transmission or reception. This process provides a significant improvement in energy-efficiency, especially at high load scenarios when ONUs are needed to be active in all cycles to avoid packet drop. However, it is essential for ONUs to initiate the wake-up process from the doze mode at least $T_{\text{dzsw}}$ (sleep-to-wake-up time from doze mode) duration before the beginning of its allocated U.S. transmission slot and hence, an ONU must know its transmission slot at least $T_{\text{dzsw}}$ beforehand. In traditional MPCP, this provision is absent as OLT informs an ONU about its transmission slot through GATE message and the slot stats immediately after receiving its GATE message. Thus, it appears to be impossible to achieve this mechanism in a complete ONU-assisted manner.

In this paper, we prove that this process can easily be realized in a complete ONU-assisted manner by modifying the traditional ranging process [7]. By adjoining our proposed mechanism of employing the doze mode during active periods with our previous proposal [16], [20], in this paper, we propose a novel ONU-assisted sleep mode protocol for energy-efficient ONU design in EPON (OSMP-EO) where all necessary modifications are provided. It can be noted that modifications are required only for U.S. traffic since receivers remain active during the doze mode. We demonstrate that a significant energy benefit is achieved by employing OSMP-EO especially at high load (~40%) as compared to other existing ONU-assisted protocols.

In this paper, we also analyze the proposed OSMP-EO protocol. The presence of traffic prediction in OSMP-EO infringes the memoryless property making the analysis a hard problem to solve. Some realistic assumptions along with an intelligent selection of discrete observation points and state descriptions aid to eliminate the dependency on the past and allow to formulate a DTMC of OSMP-EO for analyzing it. The analysis reveals that if the doze mode power consumption can be reduced by efficient circuitry design, a drastic enhancement in energy-efficiency figures of OSMP-EO is achieved. On the other hand, the energy-efficiency figures of OSMP-EO are less sensitive to the power consumption of the deep sleep and the fast sleep. Further, the decrement of sleep-to-wake-up times has a minor impact on energy-efficiency.

The rest of the paper is organized as follows. In Section II, the proposed OSMP-EO algorithm is described. The Markov model, used for analyzing the proposed OSMP-EO algorithm, is formulated in Section III. State transition probabilities and state probabilities are calculated in Sections IV and V respectively. The energy-efficiency figure of OSMP-EO is analyzed mathematically in Section VI. Model validation and few design insights are described in Section VII. In Section VIII, we provide concluding statements.
II. PROTOCOL DESCRIPTION

In this section, we first propose a simple mechanism for exploiting the doze mode during active periods in a complete ONU-assisted manner. This is followed by a detail discussion on the proposed OSMP-EO protocol.

A. Mechanism to Employ Doze Mode During Active Period (MDA)

As discussed before (refer Section I), in all existing ONU-assisted protocols, no low power mode is employed during active periods of ONUs. The transmitter of an ONU being idle during U.S. transmission of other ONUs, an opportunity of saving energy during the active periods by employing doze mode is created. However, to facilitate this, an ONU must ensure that OLT transmits its GATE message at least $T_{dzsw}$ duration beforehand so that it gets enough time to wake-up from doze mode and initiate its U.S. transmission exactly at the beginning of its allocated transmission slot. We now explain that this can easily be achieved by an ONU by modifying the round-trip process employed at the OLT to measure the round-trip time ($T_{rtt}$ - round-trip time of ONUi) of ONUs. In general, the following steps are used to measure $T_{rtt}$ at the OLT [7]. At the time of sending the GATE message for ONUi, denoted by $G_i$, the OLT sets the time-stamp field of $G_i$ by its local time (say $T_0$). ONUi resets its clock to $T_0$ immediately after receiving $G_i$. When ONUi sends the REPORT message ($R_i$), the time-tamp field of $R_i$ is set to its local time (say $T_1$). After receiving $R_i$ (say at $T_2$), the OLT calculates $T_{rtt}$ by (refer Fig. 1(a)):

$$T_{rtt} = T^{ds,i} + T^{us,i} = (T_2 - T_0) - (T_1 - T_0) = T_2 - T_1$$  

(1)

Instead of setting the time-stamp field by $T_1$, if ONUi sets it by $T_1 - T_{dz}$ then the OLT measures round-trip time of ONUi as $T_{OLT} = T_{rtt} + T_{dz}$ (refer Fig. 1(b)). While scheduling ONUi, OLT considers round-trip time of ONUi as $T^{OLT}_{OLT}$ whereas its actual value is $T^{OLT}_{OLT} - T_{dz}$. Thus, ONUi receives the GATE message exactly $T_{dz}$ duration before its U.S. transmission slot while the OLT remains oblivious of this fact. We now calculate the average power consumption of an ONU during the active periods when MDA is employed.

Calculation of the Average Power Consumption During Active States: Let us denote the duration of the $k^{th}$ cycle and the U.S. transmission slot of ONUi as $T^{k}_{US}$ and $T^{k}_{i}$ respectively. In our proposed mechanism, an ONU starts waking up from doze mode exactly $T_{dz}$ duration before the initialization of its allocated transmission slot and again enters into doze mode immediately at the end of it. Thus, $ONUi$ remains in active mode in the $k^{th}$ cycle for $T_{avg}^{i,k} = T^{k}_{US} - T^{dz}$ duration, when it consumes $P_{on}$ amount of power. For the rest of the $k^{th}$ cycle (i.e., $T^{k}_{c} - T^{k}_{i} - T^{dz}$), it is in doze mode and the power consumption is $P_{dz}$. Hence, the average power consumption of ONUi during the active mode ($P^{avg,i}_{on}$) is given by eq. (2).

$$P_{avg,i}^{on} = \frac{E_k[(T^{i,k}_{s} + T^{dz}_s)P_{on} + (T^{k}_{c} - T^{i,k}_{s} - T^{dz}_s)P_{dz}]}{E_k[T^{k}_{c}]}$$  

(2)

where $E_k$ denotes expectation over $k$. Let the average arrival rate of ONUi and the bandwidth of the feeder fiber be $\lambda_i$ and $\lambda_d$ respectively. The average duration that is occupied by the U.S. data of ONUi in a cycle is $\lambda_i T^{avg}_{c}/\lambda_d$, where $T^{avg}_{c} = E_k[T^{k}_{c}]$. In MPCP, each transmission slot includes a REPORT message of duration $T_R$ and guard duration $T_G$. Thus, $E_k[T^{i,k}_{s}] = \frac{\lambda_i}{\lambda_d} T^{avg}_{c} + T_R + T_G$. Substituting the values of $E_k[T^{i,k}_{s}]$ in eq. (2), we get:

$$P_{on}^{avg,i} = P_{dz} + \left(\frac{\lambda_i}{\lambda_d} + \frac{T_R + T_G + T^{dz}_s}{T^{avg}_c}\right)(P_{on} - P_{dz})$$  

(3)

Now, we describe our proposed OSMP-EO protocol in detail.

B. OSMP-EO

Here, we describe our proposed OSMP-EO protocol for delay insensitive traffic where the buffers of ONUs have limitations. The protocol can be easily extended for delay sensitive traffic as well. In OSMP-EO, ONUs dynamically alter between three modes: deep sleep ($ds$), fast sleep ($fs$), and active mode ($on$). During active mode, ONUs employ MDA (refer Section II-A). We denote the mode of ONUi at the current and next decision instants (say $t$ and $t_1$ respectively) by $S^{i}_{on}$ and $S^{i}_{ons}$ respectively. Further, the time interval between the current and the next decision instants is denoted by $T_{nd}$ (i.e., $t_1 = t + T_{nd}$). Now, we describe the value of $T_{nd}$ and the rules that are followed by ONUi to decide its mode at time $t$ (i.e., $S^{i}_{ons}$) with the information of the buffer fill-up time, predicted at $t_1$ ($T^{i}_{bf}$), for all values of $S^{i}_{ons}$.
Fig. 2. Illustration of (a) OSMP-EO protocol (b) energy consumption during sleep mode $S_m$.

### TABLE I
**DEFINITION OF NOTATIONS**

| Notation | Description |
|----------|-------------|
| $N_{m}^{i}$ | Maximum granted bandwidth of ONU in one cycle (packets) |
| $P_{i}$ | Power consumption of ONUs when they are in mode $i$ (Watts) |
| $T_{m}$ | Time interval between two consecutive decision instances when ONUs are in sleep mode (seconds) |
| $T_{bw}$ | Sleep-to-wake-up time of sleep mode $S_m$ (seconds) |
| $T_{bf}$ | Predicted buffer fill-up time of ONU at the current observation instant (seconds) |
| $T_{bf,n}$ | Predicted buffer fill-up time of ONU at the next observation instant (seconds) |
| $T_{S_m}$ | Minimum buffer fill-up time to enter sleep mode $S_m$ (seconds) |
| $T_{pc}$ | Time interval between the current and the next observation instant (seconds) |
| $T_{fs}$ | Time interval over which prediction is performed at the current observation instant (seconds) |
| $T_{on}$ | Time interval over which prediction is performed at the next observation instant (seconds) |
| $S_{m}^{i}$ | Mode of ONU at the current observation instant ($S_m^{i}$) |
| $S_{m}^{i+1}$ | Mode of ONU at the previous observation instant |
| $S_{m}^{i+2}$ | Mode of ONU at the next observation instant |
| $S_{m}^{ic}$ | Mode of ONU at the next decision instant |
| $N_{i}^{m}$ | Buffer size of ONU (packets) |
| $b_{i}$ | Buffer threshold of ONU (packets) |
| $b_{i}$ | Current buffer state of ONU (packets) |
| $b_{i}$ | Buffer state of ONU at the next observation instant (packets) |
| $T_{ip}$ | Packet arrivals at ONU over the time interval $T$ (packets) |
| $A_{ip}$ | Predicted packet arrivals at ONU over duration $T$ (packets) |
| $T_{f}$ | Time required to up-stream $N_{th}^{i}$ packets (seconds) |
| $T_{sw}^{m}$ | Time required to up-stream packets that are currently stored in the buffer, denoted by $b_{i}$, which is same as our previous proposal [16]. |

The OSMP-EO concentrates only on the worst case scenario in order to avoid the packet-drop possibility. However, in practical scenarios, the prediction error may cause a non-zero probability of packet-drop. This can easily be managed by setting a threshold ($N_{th}^{i}$), smaller than the actual buffer-size ($N_{i}^{m}$). Hereafter, buffer-fill up time indicates the time taken to fill-up $N_{th}^{i}$ packets.

Case $S_{m}^{i} = S_m$: If $S_{m}^{i} = S_m$, the next decision is taken after a fixed time interval $T_m$ (i.e., $T_{nd} = T_{m}$) when $ONU_i$ decides whether to retain the same sleep mode $S_m$ (i.e., $S_{nx}^{i} = S_m$) or wake up from it (i.e., $S_{nx}^{i} = on$), same as our previous proposal. Thus, if $ONU_i$ decides $S_{nx}^{i} = S_m$, it must remain in $S_m$ for $T_{m}$ duration and then only the wake-up process can be initiated, which takes $T_{sw}^{m}$ duration ($T_{m} + T_{bw}$). Sleep-to-wake-up time for sleep mode $S_m$. The OSMP-EO being an ONU-assisted protocol, in the worst possible case, an ONU may wake-up from sleep mode immediately after the arrival of its GATE message requiring a complete cycle to receive the next one when the desired bandwidth will be reported. The U.S. transmission can only be initiated after receiving the next GATE message requiring another complete cycle. Thus, $ONU_i$ requires at most $2T_{cm}$ (Maximum cycle time) duration to initiate the U.S. transmission. Therefore, if $T_{sw}^{m}$ denotes the minimum value of the time interval between the time instants when the buffer fills up and the ONU initiates the wake-up process from sleep mode $S_m$, which is required to eliminate the possibility of packet drop, then $T_{sw}^{m} = T_{sw}^{m} + 2T_{cm} + T_m$. The rules for deciding $S_{nx}^{i}$ are the following:

- If $T_{bf,n}^{i} > T_{sw}^{m}$ then $S_{nx}^{i} = S_m$
- Otherwise, it initiates the wake-up process from sleep mode $S_m$ ($S_{nx}^{i} = on$)

| Case $S_{m}^{i} = on$: | If $S_{m}^{i} = on$ then in OSMP-EO, $T_{ip}$ depends on the mode of $ONU_i$ in the previous decision instant, denoted as $S_{ip}^{i}$. If $S_{ip}^{i} = S_m$ then the next decision is taken after up-streaming $N_{th}^{i}$ packets. On the other hand, in case of $S_{ip}^{i} = on$, the next decision is taken after up-streaming all packets that are currently stored in the buffer, denoted by $b_{i}$, which is same as our previous proposal [16]. We show in Section II-C that due to the trade-off between the sleep-to-wake-up time and power consumption [3], a sleep mode provides improvement in energy-efficiency as compared to other modes with higher power consumption figures if $T_{bw}^{i} > T_{bf,n}^{i}$. The following conditions are followed for deciding $S_{nx}^{i}$:

- If $T_{bf,n}^{i} > T_{bw}^{i}$ then select the mode in the next decision point as deep sleep ($S_{nx}^{i} = ds$).
- If $T_{bw}^{i} < T_{bf,n}^{i} < T_{bw}^{i}$ then select the next mode as fast sleep ($S_{nx}^{i} = fs$).
- Otherwise, the active mode is retained ($S_{nx}^{i} = on$).

Now, we illustrate the OSMP-EO protocol with the aid of Fig. 2(a). Let, at a decision time $t$, $ONU_i$ is in sleep mode $S_m$. So, the next decision is taken after $T_m$ duration by following the rules, discussed for the case: $S_{c}^{i} = S_m$. Let, at time $t_1 (= t + T_m)$, the mode is decided as $on$. Since, in the previous decision point (i.e., at $t$) mode was $S_m$, the next decision is taken after up-streaming $N_{th}^{i}$ number of packets, which requires $T_{th}^{i}$ duration. The mode, at $t_2$ (i.e., $t + T_{th}^{i}$), is decided by following the rules, mentioned for the case of $S_{c}^{i} = on$. As, $S_{ip}^{i} = on$, the mode will be decided after up-streaming $b_{i}$ number of packets (requires $T_{bw}^{i}$ duration) by following the rules same as $t_1$. 

![Diagram](image-url)
C. Calculation of $T_{ds}^{fs}$ and $T_{lb}^{fs}$

Let, at the current decision instant, the buffer fill-up time is predicted as $T_{bf}^{fs}$. We now explain the energy consumption over this $T_{bf}^{fs}$ time interval with the help of Fig. 2(b) if ONU$_i$ decides its mode as sleep mode $S_m$. In OSMP-EO, ONU$_i$ wakes up from sleep mode $S_m$, $T_{sw}^{S_m} + 2T_{cm} + T_m = T_{sw}^{S_m}$ duration before the buffer gets filled up. If the buffer fill-up time is correctly predicted then ONU$_i$ remains in sleep mode $S_m$ for $T_{bf}^{fs} - T_{sw}^{S_m}$ duration when it consumes $P_{Sw}$ amount of power. Thereafter, ONU$_i$ requires $T_{sw}^{S_m}$ duration to wake-up when the power consumption is $P_{on}$. During the rest of the period (i.e., $2T_{cm} + T_m$) only one REPORT message will be transmitted. Since, in OSMP-EO, ONUs enter into doze mode immediately after waking up from sleep mode $S_m$, ONU$_i$ needs to wake-up from doze mode before sending the REPORT, which takes $T_{sw}^{S_m}$ duration. Thus, the power consumption over $T_{sw}^{S_m} + T_R + T_G$ time interval is $P_{on}$ and during the remaining period power consumption is $P_{ds}$. Thus, the energy consumption of ONU$_i$ during sleep mode $S_m$ ($E_{sm}$) is:

$$E_{sm} = T_{bf}^{fs} P_{Sw} + T_{sw}^{S_m} (P_{on} - P_{Sw}) + (2T_{cm} + T_m)$$

$$\times \left( P_{dz} - P_{Sw} \right) + \left( T_R + T_G + T_{ds}^{S_m} \right) (P_{on} - P_{dz})$$

The achieved energy-efficiency of $ds$, is better than $fs$ if $E_{ds} > E_{fs}$ and it requires:

$$T_{bf}^{fs} \geq \frac{T_{bf}^{fs} P_{on} + T_{sw}^{S_m} (P_{on} - P_{Sw}) + \left( T_{sw}^{S_m} - T_{sw}^{S_m} \right) P_{on}} {P_{fs} - P_{Sw}}$$

$$+ 2T_{cm} + T_m \triangleq T_{ds}^{fs}$$ (4)

If ONU$_i$ remain active then energy consumption is $E_{on} = T_{bf}^{fs} P_{on}$. The fast sleep provides better energy savings as compared to active mode if $E_{fs} > E_{on}$, which requires:

$$T_{bf}^{fs} \geq \frac{T_{bf}^{fs} (P_{on} - P_{fs}) + \left( T_{cm} + T_m \right) (P_{dz} - P_{fs})} {P_{on}^{avg,i} - P_{fs}}$$

$$+ \left( T_R + T_G + T_{ds}^{S_m} \right) (P_{on} - P_{dz}) \triangleq T_{lb}^{fs}$$ (5)

Next, we prove that energy-efficiency of the OSMP-EO protocol can be analyzed with the help of a Discrete Time Markov Chain (DTMC).

III. MODEL FORMULATION AND DESCRIPTION

In this section, we formulate a Markov model of the proposed OSMP-EO protocol, which is used for analyzing the achieved energy-efficiency figure. The OSMP-EO protocol, being an ONU-assisted protocol, the Markovian analysis can be performed for each ONU separately. Without loss of generality, in this paper, we formulate the Markov model for ONU$_i$. Since, in this paper, we are focused on analyzing the performance of the OSMP-EO protocol, the grant-sizing scheme is taken to be the simplest one: the fixed grant-sizing scheme [2]. We also explain how this analysis can be extended for other grant-sizing schemes (refer Section IV). The following assumptions are taken in this analysis.

- The traffic arrival to ONU$_i$ follows Poisson process with mean $\lambda$, which has already been assumed in all existing queuing analysis in this area (EPON) [11], [12], [15], [21]–[24].
- The predicted buffer fill-up time always replicates the actual values. Thus, our analysis provides the best possible performance and it is independent of traffic prediction mechanism. As the prediction mechanism is improved, the actual performance of the OSMP-EO protocol asymptotically follows our analytical results.

With these assumptions, we now try to formulate a Discrete Time Markov Model (DTMC) of OSMP-EO. Let us define the mode of ONU$_i$ in the previous, current and the next observation points as $S_i^1$, $S_i^2$ and $S_i^3$. If the observation points are identical to the decision points then $S_i^3 = S_{ic}$, and $S_i^1 = S_{ic}$; otherwise they are different. Since, in each decision point (refer Section II-B), ONU$_i$ decides its mode, an obvious choice for observation points of the DTMC are the decision instants. With this observation instant, we first try to find a state description that obeys the Memoryless property [25], termed as valid state description. Firstly, we claim that both $S_i^c$ and $b_i^c$ must be included in the state description.

Claim 1: The state description of DTMC must comprise of $S_i^c$ and $b_i^c$.

Proof: We know the energy consumption figures for different modes ({$ds$, $fs$, $on$}) are dissimilar. Thus, evaluation of energy-efficiency requires calculating probabilities of being in all modes, needing $S_i^c$ to be included in the state description. If $S_i^1$ individually hold the memoryless property then mode of ONU$_i$ of the next observation instant, i.e., $S_i^1$, can be decided only by $S_i^c$. However, we now show that the decision about $S_i$ requires $b_i^c$ for the case of $S_i^c = S_m \in \{ds, fs\}$, which proves our claim. In case of $S_i^c = S_m$, ONU$_i$ decide $S_i^c$ by comparing $T_{bf,n}^{fs}$ with $T_{sw}^{S_m}$ (refer Section II-B). $S_i^c$ is decided as $S_m$ if $T_{sw}^{S_m} < T_{bf,n}^{fs}$ or in other words, $\lambda T_{th} > b_i^c + A_p T_{sw}^{S_m}$ where $T_{sw}^{S_m}$ is a random process that denotes the number of packets, predicted to arrive to ONU$_i$ over the time interval $T$. Thus, $S_i^c$ depends on $b_i^c$ and $b_i^c$ can be calculated as: $b_i^c = b_i^c + A_i T_{th}^{S_m}$. Thus, $S_i^c$ depends on $b_i^c$ necessitating $b_i^c$ to be included in the state description.

Claim 1 proves the necessity of including both $S_i^c$ and $b_i^c$ into the state description. Our next step would then be to verify the sufficiency condition that is whether $S_i^c$ and $b_i^c$ jointly ({$S_i^c, b_i^c$}) hold the Markovian property so that ({$S_i^c, b_i^c$}) qualifies as a valid state description for the DTMC. Our assumption of ideal traffic prediction causes the predicted arrival process exactly same as actual one. In case of $S_i^c = S_m$, as $A_i T_{sw}^{S_m}$ follows Poisson process, $T_{bf,n}^{fs}$ is also Poisson distributed with mean $\lambda T_{sw}^{S_m}$. Since $T_{sw}^{S_m}$ depends only on $S_i^c = S_m$, $S_i^c$ depends only on $S_i^c$ and $b_i^c$ (refer Claim 1). Further, in this case, we know that $T_{th} = T_m$ (refer Section II-B) and hence, $b_i^c$ can be calculated only by the information of $b_i^c$. Thus, ({$S_i^c, b_i^c$}) seems to be a valid state description of the DTMC for the case $S_i^c = S_m$. Similar argument can further
be provided for the case of $S_{ic}^i = on$ as well. Thus, at the first glance, $\{S_{ic}^i, b_{ic}^i\}$ appears to be qualified as a valid state description with observation instances exactly same as decision instances. However, a closer look into the protocol proves that $\{S_{ic}^i, b_{ic}^i\}$ violates Memoryless property. Not only that, we now claim that if the observation instances are identical to decision points then there doesn’t exist any set of parameters at the current decision point that summarizes the complete history for the next instant, which proves the non-Markovian nature of the model.

**Claim 2:** The modeling of OSMP-EO by choosing observation points identical to the decision instances is non-Markovian.

**Proof:** At current time $t$, prediction is performed over $T_{pc}$ duration and next decision is taken after $T_{no}$ duration. Thus, there always exists an overlap region between $T_{pc}$ and $T_{no}$. Owing to the assumption of ideal traffic prediction, $A_pT_{pc}$ imposes an additional condition on deciding the next state (i.e., both $b_{ic}^i$ and $S_{ic}^i$). In OSMP-EO, if $S_{ip}^i = S_m$ then $T_{pc} = T_{sw}^{S_m}$ while if $S_{ip}^i = on$ then $T_{pc} = T_{lb}^{S_m}$ or $T_{hf}^{S_m}$, depending on $S_{ic}^i$ (refer Section II-B) and therefore, $T_{pc}$ depends on $S_{ip}^i$. Hence, mode of $ONU_i$ in the next observation point ($S_{ic}^i$) depends not only on $S_{ic}^i$ but also on $S_{ip}^i$. We now prove with the aid of Fig. 3 that $S_{ic}^i$ depends not only on $S_{ip}^i$ but also on the mode of multiple previous decision instants. Let, at time $t$, $ONU_i$ decides to change its state from active mode (on) to sleep mode $S_m$ (refer Fig. 3). This requires satisfying the condition $N_{ih}^i > b_{ic}^i + A_pT_{lb}^{S_m}$ (refer Section II-B). Owing to the assumption of ideal traffic prediction, buffered of $ONU_i$ contains lesser than $N_{ih}^i$ number of packets till $t_0(= t + T_{sb}^{S_m})$. Since, at $t$, mode of $ONU_i$ is $S_m$, the next decision is taken after $T_{no} = T_m$ duration (i.e., at $t_1 = t + T_m$) by comparing $N_{ih}^i$ with $b_{ic}^i + A_pT_{lb}^{S_m}$ (refer Section II-B). As $T_{sb}^{S_m} > T_{sw}^{S_m} + T_m$ (refer eq. (4) and eq. (5)), the condition $N_{ih}^i > b_{ic}^i + A_pT_{sw}^{S_m}$ is surely be satisfied, which ensures $S_{ic}^i = S_m$ at $t_1$. By following the same argument, it can be proved that $S_{ic}^i = S_m$ for the next $N_{po} = \left\lfloor \frac{T_{sb}^{S_m} - T_{sw}^{S_m}}{T_{cm}} \right\rfloor$ number of observation instants (refer Fig. 3). Thus, $S_{ic}^i$ of an observation instant depends on $S_{ic}^i$ of the previous $N_{po}$ number of points, which proves the non-Markovian nature of OSMP-EO.

Claim 2 proves that the proposed OSMP-EO protocol violates the Memoryless property. We now reason that intelligent selection of state description and discrete observation points, different from the actual decision points, allows converting this non-Markovian system to a Markovian one firstly for the case of $S_{ic}^i = S_m \in \{ds, fs\}$ and then for $S_{ic}^i = on$.

**Case S_{ic} = S_m:** From the afore-mentioned discussion (refer Claim 2), we understand that once $ONU_i$ enters into sleep mode $S_m$ from active mode, it retains the same state for $N_{po}$ observation instants, which introduces the dependency on the past. This dependency will be absent if $T_{no} > T_{sb}^{S_m} - T_{sw}^{S_m}$. Thus, we have to choose a value of $T_{no}$ such that $T_{no} > T_{lb}^{S_m} - T_{sw}^{S_m}$ is satisfied without loosing any information. Since, in OSMP-EO, $ONU_i$ decides its mode only at some discrete decision points (separated by $T_m$ in this case), observation instants must coincide with some decision points causing $T_{no} \geq (N_{po} + 1)T_m$. At the time instant when $T_{no} = (N_{po} + 1)T_m$, $ONU_i$ has a non-zero probability of entering into active mode. Thus, this information cannot be captured if $T_{no} > (N_{po} + 1)T_m$. Consequently, $T_{no} = (N_{po} + 1)T_m$ is the only possible option, which is selected in the analysis for the case of $S_{ip}^i = on$ and $S_{ic}^i = S_m$. However, if $S_{ip}^i = S_m$ and $S_{ic}^i = S_m$ then the possibility of choosing $S_{ic}^i = on$ is present in all decision points. Thus, in this case, we choose $T_{no} = T_m$. As a result, $T_{no}$ depends on $S_{ip}^i$ resulting in violation of Markovian property. Nevertheless, the system can be converted to a Markovian process simply by including $S_{ip}^i$ into the state description. It is now clear that $b_{ic}^i$ and $S_{ic}^i$ at the next observation point can be summarized by $b_{ic}^i$, $S_{ic}^i$ and $S_{ip}^i$ of the current observation instant (refer Section IV for further discussion). $S_{ic}^i$ of the next state is identical to $S_{ic}^i$ of the current state. Hence, $S_{ip}^i$, $b_{ic}^i$, and $S_{ic}^i$ jointly holds Markovian property allowing $\{S_{ip}^i, S_{ic}^i, b_{ic}^i\}$ as a valid state of the DTMC.

**Case S_{ic} = on:** Our next step is to prove that $\{S_{ip}^i, S_{ic}^i, b_{ic}^i\}$ is a valid state description even for $S_{ic}^i = on$. If $S_{ip}^i = S_m \in \{ds, fs\}$ and $S_{ic}^i = on$, $ONU_i$ requires $T_{sw}^{S_m}$ duration to wake-up from sleep mode $S_m$. The OSMP-EO being an ONU-assisted protocol, $ONU_i$ takes the decision of waking up from sleep mode $S_m$ without any prior information of the GATE message arrival time. Consequently, it is possible for $ONU_i$ to wake-up anywhere in between two consecutive GATE messages. The fixed grant sizing scheme ensures a constant time interval between two consecutive GATE messages: $T_{cm}$. Thus, reception of next GATE message after waking up from sleep mode $S_m$ requires $T_{cm}/2$ duration on an average. Thereafter, the U.S. transmission is initiated and it continues until $N_{ih}^i$ number of packets get up-streamed, which requires $\left\lceil \frac{N_{ih}^i}{N_{po}^i} \right\rceil T_{cm}$ time interval where $N_{ih}^i$ is the maximum granted bandwidth of $ONU_i$ in one cycle. As
a result, the next decision about its state is taken after $T_{sw} + \frac{N_{th}}{N_{pu}} T_{cm} + 0.5 T_{cm} = T_{sm}^{on}$ duration. On the other hand, if $S^i_p = S^i_c = on$ then ONU$_i$ up-stream $b^c_j$ number of packets before taking further decision about $S^c_i$. Since, in this case, no time gets wasted for waking up, up-streaming $b^c_j$ number of packets requires $\frac{b^c_j}{N_{pu}} T_{cm}$ duration which is the next decision point. In this case, if observation instants of the DTMC are considered to be same as decision points of the OSMP-EO protocol, following the similar argument as provided for the case of $S^i_c = S^m_{on}$ (discussed above), it is evident that $\{S^i_p, S^c_i, b^c_j\}$ is a valid state description even for $S^i_c = on$.

From the above discussion it is now clear that the OSMP-EO protocol can be modeled by selecting observation instances for different cases as discussed above, and the states as $\{S^i_p, S^c_i, b^c_j\}$. The proposed DTMC of the OSMP-EO algorithm is presented in Fig. 4(a) where all state transitions from states with $b^c_j = k$ is shown. In OSMP-E0, transition from one sleep mode to other is not possible, and therefore, the states $\{ds, fs, b^c_j (\forall b^c_j)\}$ and $\{fs, ds, b^c_j (\forall b^c_j)\}$ are unfeasible. $N_{sz}$ being the buffer size of ONU$_i$, states with $b^c_j > N_{sz}$ are also invalid. ONU$_i$ decides $S^c_i = S_m \in \{ds, fs\}$ only if the buffer of it contains lesser than $N_{th}$ number of packets till the next $T_{sw}^{on}$ duration and the next observation instant is after $T_m < T_{sw}^{on}$ duration. Thus, at the current as well as next observation instants, the buffer occupancy of ONU$_i$ is lesser than $N_{th}$. Consequently, the states $\{S_m, S^c_i, b^c_j \geq N_{th}\}$ and $(on, S_m, b^c_j \geq N_{th})$ are invalid. As $S^c_i$ of the current state becomes $S^m_{on}$ of the next state, transition from states with $S^c_i 
eq S^m_{on}$ is impossible. If in the current observation instant $S^c_i = S_m \in \{ds, fs\}$, the up-stream transmission of ONU$_i$ is terminated making state transition from states with $b^c_j = k$ to states with $b^c_j < k$ is impractical (refer Fig. 4(a)). From the above discussion it is now clear that the OSMP-EO protocol can be modeled by observing the system in some discrete time instants and the state space is finite. Hence, the proposed Markov model is a DTMC. State transition probabilities for all other cases are calculated next.

IV. CALCULATION OF STATE TRANSITION PROBABILITIES

Here, state transition probabilities are calculated for all cases segregated based on all feasible combinations of $S^i_p$ and $S^c_i$.

A. Transition From States With $S^i_c = S_m \in \{ds, fs\}$

Let, at the current observation time $t$, the state of ONU$_i$ be $\{S^i_p = S_m \text{ or on}, S^c_i = S_m, b^c_i = k\}$. At time $t$, for deciding $S^i_c = S_m$, ONU$_i$ must satisfy the condition $A^i_p T_{pc} \leq N_{th} - k - 1$ where $T_{pc} = T_{sw}^{on}$ if $S^i_p = S_m$ and otherwise, $T_{pc} = T_{bw}^{on}$ (refer Section II-B). The next observation instant is after $T_{no} = T_m$ duration if $S^i_p = S_m$ while if $S^i_p = on$, $T_{no} = T_{no} = (\frac{T_{sw}^{on} - T_{sw}^{on}}{T_m} + 1) T_m$ (refer Section III). As, during sleep mode, no U.S. transmission is possible, at the next observation instant (i.e., at $t_1 = t + T_{no}$), queue length will be $j(k)$ $(b^c_j = j)$ if during $T_{no}$ duration $j - k$ number of packets arrive (i.e., $A^i_p T_{no} = j - k$). The rules, followed by ONU$_i$ for deciding $S^i_c$ at $t_1$, are: $S^i_c = S_m$ if $A^i_p T_{pn} \leq N_{th} - j - 1$ and otherwise, $S^i_c = on$ where $T_{pn} = T_{bw}^{on}$. Thus, at the next observation point the state of ONU$_i$ will be $\{S^i_p = S_m, S^c_i = S_m, b^c_i = j\}$ if the events $A^i_p T_{no} = j - k$ and $A^i_p T_{pn} \leq N_{th} - j - 1$ occurs simultaneously whereas, the state will be $\{S^i_p = S_m, S^c_i = on, b^c_i = j\}$ if both $A^i_p T_{no} = j - k$ and $A^i_p T_{pn} \geq N_{th} - j - k$ happens. However, due to the assumption of ideal prediction mechanism, the condition that is satisfied at time $t$ (i.e., $A^i_p T_{pc} \leq N_{th} - k - 1$) must hold true. Let the events $A^i_p T_{no} = j - k$, $A^i_p T_{pn} \leq N_{th} - j - 1$, $A^i_p T_{pc} \geq N_{th} - j - 1$, and $A^i_p T_{pn} \leq N_{th} - k - 1$ is denoted by $Y_{j-k}$, $E_1$, $E_1$, and $E_2$ respectively. Probability of state transition from $\{S^i_p, S^c_i, k\}$ to $\{S_m, S_m, j\}$ and $\{S_m, on, j\}$, denoted by $p_{S^i_p, S^c_i, k}^{S_m, S^c_i, j}$ and $p_{S^i_p, S^c_i, k}^{S_m, on, j}$ respectively, are given by:

\[
p_{S^i_p, S^c_i, k}^{S_m, S^c_i, j} = Pr\{Y_{j-k}, E_1|E_2\}
\]

\[
p_{S^i_p, S^c_i, k}^{S_m, on, j} = Pr\{Y_{j-k}, E_1|E_2\}
\]
where \( Pr[E] \) denotes probability of occurrence of event \( E \). By using Bayes’ theorem, we get:

\[
p_{s_m,s_m,j}^{sm} = \frac{Pr[E_1, E_2] Y_j \kappa}{Pr[E_2]} \Pr\{Y_j \kappa \}
\]

and

\[
p_{s_m, on, j}^{sm} = \frac{Pr[E_1, E_2] Y_j \kappa}{Pr[E_2]} \Pr\{Y_j \kappa \}
\]  

(6)

As the arrival process to ONU follows the Poisson distribution with mean \( \lambda \), \( Pr\{Y_j \kappa \} \) is evaluated as the probability of arrival of \( j \kappa \) packets from a Poisson traffic source with mean \( \lambda T_{no} \), denoted by \( \alpha \lambda T_{no} \). Thus, \( Pr\{Y_j \kappa \} \) is given by eq. (7).

\[
Pr\{Y_j \kappa \} = \alpha \lambda T_{no} \omega \mu = \frac{e^{-\mu T_{j \kappa}}}{(\mu T_{j \kappa}!)}
\]  

(7)

We now calculate all other probability terms of eq. (6) in Lemma 1, Lemma 2 and Lemma 3.

Lemma 1: \( Pr[E_2] = \sum_{w=0}^{N_{th}-j-1} \lambda \omega T_{pc} \).

Proof: Due to the assumption of ideal traffic prediction, the random process \( A_p^i T_{pc} \) becomes identical to \( A^i T_{pc} \) which follows Poisson distribution with mean \( \lambda T_{pc} \). Thus, the event \( E_2 \) occurs if this Poisson process generates lesser than \( N_{th}-j \) packets which proves this lemma.

Lemma 2: If \( T_1 = T_{pc} - T_{no} \) and \( T_2 = T_{no} + T_{pn} - T_{pc} \) then \( Pr[E_1, E_2] Y_j \kappa \) = \( \sum_{i=0}^{N_{th}-j-1} \sum_{m=0}^{N_{th}-j-l-1} \alpha_i \lambda T_{1} \alpha_m T_{2} \).

Proof: Let \( l \) and \( m \) number of packets arrive over the time interval \( T_1 \) and \( T_2 \) respectively as shown in A of Fig 4(b).

Since, at \( t \), \( b_i^l = k \) the event \( Y_j \kappa \) results in \( b_i^j = j \) at \( t_1 \) and hence, \( A_p^i T_{pc} = j+k+l \) and \( A^i T_{pc} = l+m \). The event \( E_2 \) and \( E_1 \) restricts \( l \) and \( m \) by \( N_{th} - j - 1 \) and \( N_{th} - j - l - 1 \) respectively. Hence, \( Pr[E_1, E_2] Y_j \kappa \) is evaluated as the probability of arrival of \( l \) and \( m \) number of packets over the time interval \( T_1 \) and \( T_2 \) by a Poisson source with mean arrival rate \( \lambda \) where \( l \leq N_{th} - j - 1 \) and \( m \leq N_{th} - j - l - 1 \). This proves Lemma 3.

Lemma 3: If \( T_1 = T_{pc} - T_{no} \) and \( T_2 = T_{no} + T_{pn} - T_{pc} \) then \( Pr[E_1, E_2] Y_j \kappa \) = \( \sum_{i=0}^{N_{th}-j-1} \sum_{m=0}^{\infty} \alpha_i \lambda T_{1} \alpha_m T_{2} \).

Proof: Lemma 3 can be easily proved by following similar argument as Lemma 2.

B. Transition From States With \( (S_p^j = S_m) \) and \( S_c^j = on \)

Let, at current observation instant \( t \), the state of ONU is \( (S_m, on, k) \). For changing the mode from \( S_m \) to \( on \), at \( t \), ONU \( j \kappa \) must ensures \( A_p^i T_{pc} \geq N_{th} - k \) where \( T_{pc} = T_{Sm}^{on} + T_{cm} + T_{sa} = T_{Sm}^{on \ to \ m} \) duration (i.e., \( t_1 = t + T_{no} \)) when \( N_{th} \) number of packets depart. Thus, at \( t_1, b_i^l = j \) requires \( A^i T_{w} = j + k + N_{th} \). An approximate analysis of OSMP-EO protocol for grant sizing schemes different from fixed one can be carried out in the similar manner only by replacing \( T_{no} = T_{sw} + \frac{N_{th}}{b_i^l} \) \( T_{cm} \) duration (i.e., \( t_1 = t + T_{no} \)) when \( N_{th} \) number of packets depart. This average cycle time of the OLT, calculated by analyzing the underlying Dynamic Bandwidth Allocation (DBA) scheme.
Lemma 6: \( \Pr \{ Y_{j-k+N_{th}^i} | \bar{E}_2 \} = \sum_{l=0}^{N_{th}^i-k+j} \alpha_l \lambda T_{pc} \alpha_{j-k-l+N_{th}^i} \) where \( T_{pc} = T_{cm}^{\text{min}} \).

Proof: Let \( l \) and \( m \) number of packets arrive over \( T_{pc} \) and next \( T_{no} - T_{pc} \) duration respectively (refer B of Fig. 4(b)) resulting in \( A^i, T_{no} = l + m \). The event \( \bar{E}_2 \) imposes a lower limit of \( l \geq N_{th}^i - k \). Further, the event \( Y_{j-k+N_{th}^i} \) ensures \( m = j - k - l + N_{th}^i \) \( m \geq 0 \) restricts the value of \( l \) by; \( l \leq N_{th}^i - k + j \). This proves Lemma 6.

Lemma 7: \( \Pr \{ \bar{E}_d | E_F, E_d^i \} = \sum_{l=0}^{N_{th}^i-j-l} \sum_{m=0}^{\infty} \alpha_l \lambda T_{fs}^i \times \alpha_{j-l} \) \( \frac{\lambda T_{fs}^i}{\lambda T_{fs}^i - \lambda T_{ib}^i} \).

Proof: Let \( l_1 \) and \( m_1 \) number of packets arrive over time interval \( T_{fs}^i \) and next \( T_{ib}^i \) (refer B of Fig. 4(b)) causing \( A_p^i, T_{ib}^i = l_1 + m_1 \). The events \( E_F \) and \( E_d^i \) restricts \( l_1 \) and \( m_1 \) by; \( l_1 \leq N_{th}^i - j - 1 \) and \( m_1 \geq N_{th}^i - j - l_1 \) respectively which proves Lemma 7.

C. Transition From States With \( S_c^i = S_c^k = \text{on} \)

Let, at current time \( t \), ONU_i retain the active state which requires satisfying the condition \( A_p^i, T_{ib}^i > N_{th}^i - j \) (i.e., \( \bar{E}_2 \)) if \( b_c^i < N_{th}^i \) whereas if \( b_c^i \geq N_{th}^i \) then no condition is needed to be satisfied. In this case, \( T_{no} = [b_c^i/N_{th}^i] T_{cm} \) (refer Section III) when all \( b_c^i = k \) number of packets get upstreamed. Thus, at \( t_1 = t + T_{no} \), \( b_c^i \) \( j \) requires \( A^i, T_{no} = j \).

Extension of this analysis for other grant issuing schemes can be performed by modifying \( T_{no} \) in the similar manner as discussed above. Further, at \( t_1 \), \( S_c^k \) is decided by verifying the following conditions: \( S_c^k = \text{on} \) occurs, while if both \( E_d^i \) and \( E_F^i \) happens then \( S_c^k = \text{fs} \) and otherwise, \( S_c^k = \text{on} \).

Now, we calculate the state transition probability for three different cases; \( T_{no} < T_{fs}^i, b_c^i < N_{th}^i; T_{no} > T_{fs}^i, b_c^i < N_{th}^i; b_c^i \geq N_{th}^i \).

Case 1 \( (T_{no} < T_{fs}^i, b_c^i < N_{th}^i) \): By following similar steps as used for calculating eq. (6), \( p_{on,ds;j}, p_{on,fs;j}, p_{on,ds;k}, p_{on,fs;k}, \) and \( p_{on,on;j} \) is given by eq. (12).

\[
\begin{align*}
& p_{on,ds;j} = \frac{\Pr \{ \bar{E}_2, E_{ds} | Y_j \} \Pr \{ \bar{E}_2 \}}{\Pr \{ \bar{E}_2 \} \Pr \{ \bar{E}_2 \}} \alpha_l \lambda T_{no}, \\
& p_{on,fs;j} = \frac{\Pr \{ \bar{E}_2, E_{fs} | Y_j \} \Pr \{ \bar{E}_2 \}}{\Pr \{ \bar{E}_2 \} \Pr \{ \bar{E}_2 \}} \alpha_j \lambda T_{no}, \\
& p_{on,ds;k} = \frac{\Pr \{ \bar{E}_2, E_{ds} | \bar{E}_2 \} \Pr \{ \bar{E}_2 \}}{\Pr \{ \bar{E}_2 \} \Pr \{ \bar{E}_2 \}} \alpha_l \lambda T_{no}, \\
& p_{on,fs;k} = \frac{\Pr \{ \bar{E}_2, E_{fs} | \bar{E}_2 \} \Pr \{ \bar{E}_2 \}}{\Pr \{ \bar{E}_2 \} \Pr \{ \bar{E}_2 \}} \alpha_j \lambda T_{no}.
\end{align*}
\]

Pr \( \{ \bar{E}_2 \} \) is calculated by using Lemma 4. All other probability terms of eq. (12) are calculated by using Lemma 8–Lemma 10. For proving Lemma 8–Lemma 10, we assume that \( l, m_1 \) and \( m - m_1 \) number of packets arrives over three consecutive time intervals \( T_1 = T_{pc} - T_{no} \), \( T_2 = T_{no} + T_{fs}^i - T_{pc} \) and \( T_3 = T_{ib}^i - T_{fs}^i \) respectively as shown in C of Fig. 4(b).

The event \( Y_j \) ensures \( A^i, T_{pc} = j + l, A_p^i, T_{ib}^i = l + m_1 \) and \( A_p^i, T_{ib}^i = l + m \).

Lemma 8: \( \Pr \{ \bar{E}_2, E_{ds} | \bar{E}_2 \} = \sum_{l=0}^{N_{th}^i-j-1} \sum_{m=0}^{\infty} \alpha_l \lambda T_{fs}^i \alpha_{m} \lambda T_{ib}^i T_{ib}^i \).

Proof: The events \( \bar{E}_2 \) and \( E_{ds} \) requires \( l \geq N_{th}^i - k - j \) and \( m \leq N_{th}^i - j - l - 1 \). We know both \( l \) and \( m \) are positive, \( m > 0 \) upper-bounds the values of \( l \) by \( N_{th}^i - j - 1 \). Thus, the event \( \bar{E}_2, E_{ds} \) occurs if \( l \) and \( m \) number of packets over the time interval \( T_1 \) and \( T_2 + T_3 \) with all restrictions on \( l \) and \( m \) as discussed above, which proves Lemma 8.

Lemma 9:

\[
\begin{align*}
& \Pr \{ \bar{E}_2, E_{fs} | \bar{E}_2 \} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha_j \lambda T_{fs}^i \alpha_{m} \lambda T_{ib}^i T_{ib}^i, \\
& \Pr \{ E_{ds} | \bar{E}_2 \} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha_j \lambda T_{fs}^i \alpha_{m} \lambda T_{ib}^i T_{ib}^i.
\end{align*}
\]

Proof: The events \( \bar{E}_2 \) and \( E_{fs} \) ensures \( l \geq N_{th}^i - k - j \) and \( m \leq N_{th}^i - j - l - 1 \) and \( m \geq N_{th}^i - j - l \) respectively. Thus, for all \( m \) and \( m_1 \), \( m > 0 \) are positive. By using this, Lemma 10 can be easily proved.

Case 2 \( (T_{no} < T_{ib}^i, b_c^i < N_{th}^i) \): This case is exactly same to the case \( S_c^k = S_m \) \( \{ \text{ds, fs} \}, S_c^i = \text{on} \) (refer Section IV-B).

Case 3 \( (b_c^i = k \geq N_{th}^i) \): In this case, ONU_i enters into the ON state without any prediction. Further, it can be clearly noticed that in this case, the arrival process over \( T_{no} \) duration is independent of the arrival process over \( T_{fs}^i \) or \( T_{ib}^i \). Hence, in this case, state transition probabilities \( p_{on,ds}, p_{on,fs}, p_{on,on}, p_{on,N_s^i} \) is given by eq. (13).

\[
\begin{align*}
& p_{on,ds} = \alpha_j \lambda T_{no} T_{ib}^i \sum_{l=0}^{N_{th}^i-j-1} \alpha_l, \\
& p_{on,fs} = \alpha_j \lambda T_{no} T_{ib}^i \sum_{l=0}^{N_{th}^i-j-1} \alpha_l, \\
& p_{on,on} = \alpha_j \lambda T_{no} T_{fs}^i \sum_{l=0}^{\infty} \alpha_l, \\
& p_{on,N_s^i} = \alpha_j \lambda T_{no} T_{fs}^i \sum_{l=0}^{\infty} \alpha_l.
\end{align*}
\]
V. Calculation of State Probabilities

Here, we calculate all state probabilities. In OSMP-EO, transition between two sleep modes is not possible. Further, $S^i_p$ of the next state is exactly same as $S^i_c$ of the current state (refer Section III). If $S^i_p = S^i_c \in \{ds, fs\}$ then the up-stream transmission gets terminated, eliminating the possibility of state transition between states with $b^i = k$ to states with $b^i < k$. By considering these restrictions on state transitions, the probabilities of being in states $\{S_m, S^i_c, j\}$ ($\pi_{S_m, S^i_c, j}$) and $\{on, S^i_c, j\}$ ($\pi_{on, S^i_c, j}$) are given by eq. (14) and eq. (15) respectively (refer Fig. 4(a)).

\[
\pi_{S_m, S^i_c, j} = \sum_{k=0}^{j} \pi_{S_m, k} \pi_{S^i_c, k} + \sum_{k=0}^{j} \pi_{on, S_m, k} \pi_{on, S^i_c, k}
\]

(14)

\[
\pi_{on, S^i_c, j} = \sum_{k=0}^{N_s^i - 1} \pi_{on, ds, j} \pi_{ds, j} + \sum_{k=0}^{N_s^i - 1} \pi_{on, fs, j} \pi_{fs, j} + \sum_{k=0}^{N_s^i - 1} \pi_{on, S^i_c, j}
\]

(15)

VI. Calculation of Average Energy-Efficiency

In this section, we analyze the average energy savings figure of the OSMP-EO protocol. $ONU_i$ retains the mode of the current observation instant ($S^i_c$) till the next observation instant, i.e., for a duration of $T_{no}$. We denote $T_{no}$ and the energy consumption of $ONU_i$ over $T_{no}$ for state $\{S^i_p, S^i_c, b^i\}$ as $E_{S^i_p, S^i_c, b^i}$ and $E_{S^i_p, S^i_c, b^i}$ respectively. If $S^i_c = S^i_m \in \{ds, fs\}$ then $T_{no} = T_m$ and power consumption is $P_{S^i_m}$ (refer Section III) and hence, $E_{S^i_p, S^i_c, b^i} = P_{S^i_m} T_m$. In case of $S^i_p = S^i_m$ and $S^i_c = on$, $T_{no} = T_{S^i_m, on, b^i} = T_{sw} + \frac{N_s^i}{N_m} T_{cm} + \frac{T_m}{N_m}$. Among this period $ONU_i$ takes $T_{sw}$ duration to wake-up from sleep mode when power consumption is $P_{sw}$. After waking up, $ONU_i$ requires 1.5 $T_{cm}$ duration to start the up-stream data transmission when only one REPORT message is up-streamed. The rest of the period ($t_{rp} = [N_y^i/N_m] - 1$) is used for up-stream data transmission when the power consumption is $P_{on, b^i}$ (refer eq. (3)). Thus, in this case, $E_{S^i_p, on, b^i} = T_{sw} P_{sw} + (1.5 T_{cm} - T_{sw}) P_{on} + (T_{sw} - T_G) P_{on} + t_{rp} P_{on, b^i}$. In case of $S^i_p = S^i_c = on$, over entire $T_{no} = T_{on, b^i} = \frac{b^i}{N_m} T_{cm}$ duration power consumption is $P_{on, b^i}^{avg, i}$ and hence, $E_{on, on, b^i}^{avg, i} = T_{no} P_{on, b^i}^{avg, i}$. Therefore, the average power consumption ($P_{avg}$) of the OSMP-EO algorithm is given by eq. (16).

\[
P_{avg} = \frac{\sum_{i=0}^{N_s^i} \sum_{k=0}^{N_s^i} \sum_{j=0}^{N_s^i} P_{S^i_p, S^i_c, j} \pi_{S^i_p, S^i_c, j}}{\sum_{i=0}^{N_s^i} \sum_{k=0}^{N_s^i} \sum_{j=0}^{N_s^i} P_{S^i_p, S^i_c, j} \pi_{S^i_p, S^i_c, j}}
\]

(16)

However, if no energy-efficient mode is employed then power consumption is always $P_{on}$. Thus, the average energy-efficiency of OSMP-EO is given by: $\eta_{avg} = 1 - \frac{P_{avg}}{P_{on}}$.

Note that the average delay figure can also be analyzed in the similar manner. However, in this case, an extra variable, $N_y^i$, is needed to be included in the state definition for all states with $S^i_p = S^i_c = S^i_m$ where $N_y^i$ denotes the number of times the ONU decides to retain $S^i_m$ after entering into it.

VII. Result and Discussion

In this Section, firstly, the average energy-efficiency figures of the OSMP-EO protocol are compared with the same of other existing schemes. We then validate the analytical results with the results obtained from simulations. We next compare the results obtained by using our analytical model with that generated by simulations for self-similar traffic. All simulations are performed in OMNET++ for a network runtime of 50s and the results are plotted with 95% confidence interval. The link rate of the feeder fiber, the maximum traffic arrival rate at each ONU, the packet size and $T_m$ are assumed to be 1Gbps, 100Mbps, 1500Bytes and 0.5ms respectively [16]. Sleep-to-wake-up time of ds, fs, and dz are considered to be 5.125ms, 125μs and 1μs respectively [16]. Whereas power consumption of ds, fs, dz and on are 0.75W, 1.28W, 2.39W and 3.984W respectively [16].

A. Comparison With Existing Literature

We have already demonstrated in [16] that our previous proposal outperforms all existing protocols. Here, we compare the OSMP-EO protocol with our previous proposal where no sleep mode was employed during active periods. In OSMP-Eo, we introduce a mechanism of employing doze mode during the active periods, which provides additional energy savings. In order to quantify this improvement, in Fig. 5(a) and Fig. 5(b), the average energy-efficiency figures are plotted as a function of traffic load for $N_y^i = N_z^i = 1Mb$ and 10Mb for two different scenarios: (i) only one ONU (say $ONU_i$) follows OSMP-EO
protocol while all other ONUs are traditional ONUs and (ii) all ONUs follow OSMP-EO protocol respectively. The average delay of OSMP-EO protocol for $N_{th}^i=N_{sz}^i=1$ Mb is plotted if Fig. 5(c). Here, traffic load of an ONU is defined with respect to the maximum arrival rate to that ONU. The traffic arrivals are considered to be self-similar with Hurst parameter $(H)$ 0.8 and it is generated by aggregating 16 ON-OFF Pareto sources. Traffic prediction is performed by using ARMA$(2, 2)$ model in a similar manner as discussed in our previous proposal [16]. Fig. 5(a) and Fig. 5(b) depict that at low load, especially for the case of $N_{th}^i = 10$ Mb, the improvement in energy efficiency, as compared to our previous proposal [16] is very small whereas at high load a significant energy benefit ($\sim 40\%$ for $N_{th}^i = 1$ Mb) can be achieved. This is due to the following fact. Both the reduction of traffic load and the increment of $N_{th}^i$ enhances the buffer fill-up time. As a result, the sleep duration increases and hence, ONUs wake-up from sleep mode less frequently. In OSMP-EO, a certain time interval is wasted when an ONU wakes-up from sleep mode and this wastage can be reduced by increasing of $N_{th}^i$ or by reducing traffic load. Further, the reduction of traffic load decreases the time interval over which an ONU up-streams. As a result of these two facts, the total active period gets reduced with the reduction of traffic load and increment of $N_{th}^i$. Since OSMP-EO enhances energy efficiency by introducing doze mode during active periods; at a load low, for a higher value of $N_{th}^i$ (when the active period is very small), the increment is negligible. Whereas, at a high load with low value of $N_{th}^i$, the improvement is significant. Further, since the increment of traffic load reduces the sleep duration, an increase of traffic load results in a decrement of average delay as seen in Fig. 5(c).

B. Validation of Analytical Model With Simulations

Here, the proposed Markov model of the OSMP-EO protocol is validated with simulation results for Poisson traffic. It is well known that in case of Poisson traffic, the current arrival is independent of the previous arrivals, and hence, traffic prediction with the knowledge of previous arrivals is impossible. In simulations, we consider that $ONU_i$ predicts the number of packet arrivals over $T_{od}$ duration by $\lambda_iT_{od}$ where $\lambda_i$ is average packet arrival rate to $ONU_i$. We assume the buffer size $N_{sz}^i$ as $1.2$ Mb while all other parameters are same as they are taken in Section VII-A. The achieved energy-efficiency figures as a function of traffic load along with the same obtained from our analytical model for different values of $N$, $N_{im}^i$ and $N_{th}^i$ are plotted in Fig. 6(a), Fig. 6(b) and Fig. 6(c) respectively. In Fig. 6(a), we plot the energy-efficiency figures for $N = 16$, 32 by considering $N_{im}^i = 60$ Kb and $N_{th}^i = 0.48$ Mb. Whereas in Fig. 6(b) and Fig. 6(c), the same is plotted for $N_{im}^i = 0.12$ Mb, 60 Kb with $N = 16$, $N_{th}^i = 0.48$ Mb and for $N_{im}^i = 0.48$ Mb, 0.24 Mb with $N = 16$, $N_{th}^i = 0.12$ Mb. It can be observed from the figures that the analytical results match closely with simulation results for all considered scenarios, which validate our analytical model. However, at higher load, the simulation results deviate from the analytical results, which are due to prediction error. It is interesting to note that this deviation increases with an increase in traffic load. This is due to the following facts. In simulations, Poisson traffic arrivals are predicted by its mean and hence, the prediction error is proportional to the variance of the Poisson distribution. An increment of traffic load enhances the variance of the Poisson distribution and, hence the prediction error.

1) Effect of $N$: It can be noticed from Fig. 6(a) that the achieved energy-efficiency figures reduce drastically with an increase in $N$ from $N = 16$ to $N = 32$. The reason can be explained as follows. In OSMP-EO, if $ONU_i$ enters into the active mode from a sleep mode, it remains active until it up-streams $N_{th}^i$ number of packets. Since in each cycle maximum of $N_{im}^i$ number of packets get cleared, $\left\lceil \frac{N_{im}^i}{N_{th}^i} \right\rceil$ number of cycles are required to up-stream $N_{th}^i$ number of packets and over this entire period, $ONU_i$ remains active. If the value of $N$ is doubled then the duration of each cycle almost become double for fixed allocation scheme and hence, the duration of active periods of $ONU_i$ almost become double resulting in a drastic reduction of energy-efficiency.

2) Effect of $N_{im}^i$: It can be observed from Fig. 6(b) that the energy-efficiency figures remain almost same with the variation $N_{im}^i$ at low load. However, at higher load, a slight decrement in energy-efficiency occurs if $N_{im}^i$ is incremented. The reason can be explained as follows. If the value of $N_{im}^i$ is doubled, then the number of cycles over which $ONU_i$ remains in active mode before taking the next decision become half. Whereas, the duration of each cycle become almost double. Thus, the duration of active periods over which the U.S. data is transmitted remains unchanged. However, in OSMP-EO, once an ONU decides to be active from sleep mode, on an
average half of a cycle is wasted to receive the next GATE message (refer Section II-B). An increment of \( N_{im} \) causes an increase in cycle time and hence, this waste of half of a cycle, which results in a reduction of energy-efficiency. Since this waste is present in every occurrence of state transitions from sleep mode to active mode, at low load, when the probability of waking up is very small, the effect of variation of \( N_{im} \) is negligible.

3) Effect of \( N_{ih} \): A decrement of \( N_{ih} \) results in a reduction of buffer fill-up time and hence, the probability of state transition from sleep mode to active mode increases. In every occurrence of waking up from \( S_m \), \( T_{su} + T_{cm}/2 \) duration is wasted (refer Section IV-B). Thus, the decrement of \( N_{ih} \) causes a reduction in energy-efficiency.

C. Applicability to Self Similar Traffic

Here, we compare our analytical results with the results, obtained from simulations with self-similar traffic. In order to do so, in Fig. 7(a), we plot the energy-efficiency figures that are obtained from simulations with self-similar traffic for \( H = 0.55, 0.8 \) and from the analytical model with Poisson traffic as a function of traffic load. Self-similar traffic is generated and predicted in the same way as it is discussed in Section VII-A. We consider \( N = 16, N_{im} = 60Kb \) and \( N_{ih} = 0.48Mb \). It can be observed that the analytical results match closely with simulation results even for self-similar traffic. However, at high load, the energy-efficiency figures for self-similar traffic is little higher than the analytical results. This is due to the fact that the bursty nature of self-similar traffic increases the packet drop probability and hence, reduces the effective traffic load. The decrement of \( H \) reduces the burstiness and wherefore, the packets drop probability. Thus, the results for self-similar traffic approach to the analytical result with the reduction of \( H \), which can also be observed from Fig. 7.

D. Design Insights

Here, we observe the effect of reducing the sleep-to-wake-up times and the power consumption figures of all sleep modes on the energy-efficiency, which can be achieved by designing advance circuitries. For this purpose, energy-efficiency figures that are obtained from the analytical model if the sleep-to-wake-up time and power consumption figures of all sleep modes are reduced individually by 25% are plotted in Fig. 7(b) and Fig. 7(c) respectively. It can be noticed from Fig. 7(b) that the effect of diminishing the sleep-to-wake-up time of all modes is insignificant and hence, makes no sense for a circuit designer to focus on it. Reduction in power consumption of deep sleep and fast sleep increase energy-efficiency at low and moderate load respectively (refer Fig. 7(c)). This is because, \( ONU_i \) enter into the deep sleep if the buffer fill-up time is more than \( T_{ds}^{sl} \), which has a significantly high value (refer eq. (4)) and hence, exists only at low load. Similarly, fast sleep exists at moderate load where the effect of reducing \( P_{fs} \) is present. However, in OSMP-EO, doze mode reduces the energy consumption of active periods and therefore, the effect of reducing \( P_{ds} \) exists in all loads. An increment of traffic load increases the duration of active periods and hence, the enhances the improvement of energy-efficiency. Thus, from the point of view of a design engineer, he should mostly focus on reducing the power consumption figures of the doze mode. Fig. 7(a).

VIII. Conclusion

In this paper, we propose an ONU-assisted sleep mode protocol (OSMP-EO) for energy-efficient ONU design in EPON where ONUs are allowed to save energy even during the active cycles. This provides a significant improvement in energy savings especially at high load (~40%) when the energy-efficiency approaches zero in all existing ONU-assisted protocols. A mathematical analysis of the OSMP-EO protocol is also performed. We first prove that this protocol infringes the memoryless property. We then explain that the entire history can be easily captured by an intelligent selection of discrete observation instants and associated state descriptions, which allow formulating a DTMC for analyzing energy-efficiency figures. The analytical model is verified with simulations for both Poisson and self-similar traffic. The analysis reveals that the protocol is almost insensitive to sleep-to-wake-up times of all sleep modes. An improvement in energy-efficiency can be achieved at low and moderate load if the power consumption figures can be reduced by designing better circuitries. Further, the energy-efficiency figure of OSMP-EO is highly sensitive to doze mode power consumption figure and hence, circuit designer should focus on reducing it.
