Joint Scheduling and Bandwidth Allocation Methods for 10G-EPON and Beyond

Mohammad S. Kiaei * †, Chadi Assi *, Martin Maier †
* ENCS, Concordia University, Montreal, Quebec, Canada
† Optical Zeitgeist Laboratory, INRS, Montreal, Quebec, Canada

Abstract—Dynamic bandwidth allocation and grant scheduling are among the major issues in the design of passive optical networks. In this paper, we investigate the problem of optimal scheduling and bandwidth allocation in next generation 10G-EPON coexisting with 1G WDM-PONs. We first propose a network architecture for providing this coexistence. Then, we derive an ILP model for offline joint scheduling and bandwidth assignment for 10G-TDM and 1G-WDM ONUs. In order to address the scalability of the ILP model, we introduce a Tabu Search based heuristic for obtaining near optimal solutions while the computation time is remarkably reduced. We further explore the tradeoff which exists in terms of delay, scheduling length, and channel utilization, when separate or the same DBA modules are used for 1G- and 10G-ONUs.

I. INTRODUCTION

Ethernet Passive Optical Network (EPON) has been recently extended to 10G-EPON in order to provide ten-fold data rates of 10 Gb/s. 10G-EPON has emerged as a promising candidate for the next-generation high data rate access systems [1]. This new PON has been standardized under the IEEE 802.3av task force with the aim of developing the physical layer specifications and management parameters. The 10G-EPON standard provides symmetric 10 Gb/s downstream and upstream, as well as asymmetric 10 Gb/s downstream and 1 Gb/s upstream data rates. In order to provide backward compatibility with the existing and widely deployed 1G-EPON, the OLT in a 10G-EPON is equipped with dual rate receivers for receiving data from 1G and 10G-ONUs. Furthermore, the downstream transmission channels are separated for sending downstream data and control traffic to 1G- and 10G-ONUs.

Another alternative to cope with the increasing bandwidth demand is to use wavelength division multiplexing (WDM) PONs; however, these systems has not been standardized yet. A WDM-PON can support multiple wavelengths in either or both upstream and downstream directions. WDM-PONs can also be combined with TDM techniques used by the EPON standard in order to further reduce costs and achieve larger bandwidth support. This combination leads to hybrid WDM-TDM PONs, which improves scalability by allowing splitting ratios of up to 1:1000, at the expense of optical amplifiers [2].

The process of dynamic bandwidth allocation (DBA) in a PON system normally consists of two subproblems: grant sizing and grant scheduling. Grant sizing determines the length of the transmission window assigned to an ONU per each polling cycle, while grant scheduling indicates the order of ONU grants during a given cycle [3]. Several grant sizing and scheduling methods have been introduced in the literature for 1G WDM/TDM PON. A straightforward scheduling method for WDM PONs is the online Next Available Supported Channel (NASC) where the OLT schedules the upstream transmission of an ONU on the earliest available wavelength channel, which is supported by the ONU. Alternatively, in offline scheduling, the OLT makes the schedule after receiving bandwidth requests from all ONUs [4].

In this work, we investigate the problem of optimal scheduling and bandwidth allocation in next generation 10G-EPON coexisting with 1G WDM-PONs. We first propose a network architecture for supporting the coexistence. Then, we derive an ILP model for offline joint scheduling and bandwidth assignment for 10G-TDM and 1G-WDM ONUs. Our goal is to develop efficient bandwidth allocation and scheduling algorithms in this system with multi-rate ONUs. Based on the choice of wavelength channels, the OLT may use separate or the same DBA modules for 1G- and 10G-PONs. To address this fact, we study two scheduling scenarios where the 10G TDM channel is either shared between 1G- and 10G-ONUs, or it is dedicated to 10G-ONUs. We exploit the tradeoff which exists in terms of delay, scheduling length, and channel utilization, when separate or the same DBA modules are used for 1G- and 10G-ONUs. To address the scalability of the ILP model, we introduce a Tabu Search based heuristic for obtaining near optimal solutions in remarkably shorter computation time.

II. IEEE 802.3AV SPECIFICATIONS

One major difference between 1G- and 10G-EPON is that the latter supports both symmetric 10 Gb/s downstream and upstream, and asymmetric 10 Gb/s downstream and 1 Gb/s upstream data rates, while 1G-EPON provides only a 1 Gb/s symmetric data rate. The wavebands utilized for upstream (US) and downstream (DS) transmissions of 1G- and 10G-EPON standards are illustrated in Fig. 1. As can be seen in Fig. 1(a), 1G-EPON allocates a 100 nm waveband centered at 1310 nm for upstream (US) transmission and a 20 nm window centered at 1490 nm for downstream (DS) transmission. As shown in Fig. 1(b), the US waveband of 10G-EPON is allocated in a 20 nm window which is completely covered by a part of the 1G-EPON US waveband. In the DS direction, since the wavebands are distinct, a WDM-overlay is a straightforward way to provide the coexistence with the legacy 1G-EPON. On the contrary, as depicted in Fig. 1, the US waveband of
10G-EPON is in fact a subset of the 1G-EPON waveband; hence a dual-rate burst-mode operation is the only remaining option to retain the coexistence requirement by using dual-speed TDM [1]. For this purpose, the OLT is equipped with a dual-rate receiver. Moreover, the OLT provides three kinds of MAC instances for operating on symmetric and asymmetric data rates; namely, the OLT supports 1/1 Gb/s, 10/1 Gb/s, and dual-rate receiver. Moreover, the OLT provides three kinds of MAC instances; namely, 1/1 Gb/s, 10/1 Gb/s, and 10/10 Gb/s MAC instances.

III. NETWORK ARCHITECTURE

As illustrated in Fig. 2, our PON structure comprises one OLT connecting in a tree topology to multiple 1G- and 10G-ONUs. In order to enjoy the benefits of multi-channel upgraded PON or hybrid WDM-TDM PONs, the US transmission waveband should be split into multiple wavelength channels. We note that the US waveband for 10G-EPON is too narrow to be split into multiple wavelengths, whereas the 100 nm waveband of 1G-EPON can be more easily split into multiple channels for upstream transmissions. Therefore, we consider a dual-rate EPON architecture with 10G-TDM ONUs coexisting with future 1G-WDM ONUs. One of the most cost-effective technologies for realizing WDM ONUs is to utilize so-called “colorless ONUs” which are wavelength-independent, and make use of a reflective semiconductor optical amplifier (RSOA) at the ONU for remote modulation of the upstream data [5]. In this approach, the OLT is equipped with laser diodes to send optical continuous wave (CW) signals to the attached reflective ONUs where the CW signal is modulated and sent back to the OLT; hence, no light source is required at the ONU.

In our architecture, the OLT is equipped with an array of fixed-tuned receivers and fixed-tuned transmitters for receiving from and sending out data to the ONUs. Two types of receivers are deployed at the OLT. One is denoted by \( RX_{1G} \) which is used at one of the US channels \( \lambda_1, \ldots, \lambda_U \) for receiving data from 1G-ONUs. The other is the dual-rate receiver \( RX_{1G,XG} \) tuned at the center of 10G-EPON US waveband (\( \lambda_{XG} = 1270 \) nm) for receiving data from 10G-ONUs and from those 1G-ONUs sharing the US channel with 10G-ONUs. Each of the transmitters at the OLT are either fixed tuned at one of wavelengths \( \lambda_1, \ldots, \lambda_U \) for sending CW signals to the reflective 1G-ONUs, or they are tuned at one of the wavelengths \( \lambda_U+1, \ldots, \lambda_U+D \) for sending DS data and control traffic to 1G-ONUs. Also, there is a 10 Gb/s transmitter at the OLT fixed tuned at \( \lambda_{XG}^{down} \) for transmitting DS data to 10G-ONUs.

The OLT provides three kinds of MAC instances; namely, 1/1 Gb/s, 10/1 Gb/s, and 10/10 Gb/s. The 10G-ONUs are TDM ONUs working on \( \lambda_{XG}^{up} \) and \( \lambda_{XG}^{down} \) channels for their US and DS transmissions, respectively. As shown in Fig. 2, a given 10G-ONU generates either a 1 or 10 Gb/s signal, depending on which one of the two specified transmit paths is implemented at the ONU [6]. Conversely, the 1G-WDM ONUs are equipped with RSOA which can be tuned on all the existing upstream channels including \( \lambda_{XG}^{up} \) for transmitting upstream data and control traffic to the OLT. This way, the 1G-ONUs are capable of transmitting on all available channels including the 10G channel; we do not however allow simultaneous transmissions on multiple channels. Also, each 1G-ONUs employs an array of fixed tuned receivers, each tuned at one of the wavelengths \( \lambda_U+1, \ldots, \lambda_U+D \) for receiving downstream data and control traffic from the OLT.

IV. MOTIVATION AND PROBLEM STATEMENT

It is important to note that some network operators may restrict the US waveband of 1G-EPON customers in order to avoid inventory problems [7]. Thus, the upstream coexistence can be achieved using WDM. Whether the US waveband is restricted or not, 10G-ONUs may or may not share their upstream channel with 1G-WDM ONUs. When the 1G-ONUs and 10G-ONUs operate on the same channel and the dual-rate burst-mode is used, all ONUs should be controlled by a single scheduler and DBA module at the OLT. Conversely, if the allocated wavebands for 1G and 10G-ONUs are different, the OLT can deploy separate DBA and scheduling modules for 1G and 10G-ONUs. We illustrate these two scenarios and their effects on the channel utilization and scheduling length. Our instance network consists of 10 1G-WDM ONUs and 2 10G-TDM ONUs. There are four 1G WDM wavelength and one 10G wavelength for 10G ONUs. The round-trip time (RTT) between each ONU and OLT is 100 \( \mu \)sec.

The allocated time slots of each ONU are illustrated in Fig. 3. In Fig. 3(a), the 10G wavelength channel is dedicated to 10G-ONUs, which are arbitrated according to online NASC scheduling method. In this case, the OLT polls the 10G-ONUs every 100 \( \mu \)sec and grants the requested bandwidth. The 1G-WDM ONUs are scheduled using the non-joint offline scheduling method presented in [8]. The initial gap represents the inter-scheduling cycle gap (ISCG) which is mainly determined by the RTT of the first ONU scheduled on each channel [9]. We observe that almost 60% of the 10G channel is wasted, whereas this channel could have been utilized more efficiently.
if it had been shared by 1G-ONUs. In Fig. 3(b), we see that a more efficient schedule with smaller polling cycle length (or makespan) and higher channel utilization can be obtained when the 10G channel is shared with the 1G-ONUs. As the polling cycle increases for 10G-ONUs, they will have larger bandwidth requests compared to the online scheduling in Fig. 3(a). Consequently, the average packet delay will increase for 10G-ONUs. To mitigate this problem, we can further reduce the makespan using joint scheduling and bandwidth allocation for 1G-ONUs (Fig. 3(c)). The window size of highly loaded 1G-ONUs is reduced based on the minimum guaranteed bandwidth of each channel. We observe that in the joint method, the scheduling length and therefore the packet delay for 10G-ONUs are decreased at the expense of a larger delay for the 1G-ONUs. In summary, Fig. 3 illustrates a clear tradeoff between transmission delay, channel utilization and scheduling period when using different scheduling methods.

V. SCHEDULING AND BANDWIDTH ALLOCATION FOR 10G-TDM AND 1G-WDM ONUS

We assume that each 1G-ONU can not transmit on more than one channel per cycle. This way, during each polling cycle, the OLT has to send only one CW signal to the WDM-ONU in order to remotely modulate the upstream data; therefore, the planning cost decreases compared to a scenario where the ONU transmissions per cycle are allowed to be bifurcated into different channels and the OLT has to send multiple CW signals to WDM-ONUs in each polling cycle.

A. Bandwidth Allocation

In order to determine the allocated bandwidth for each ONU, we should first determine the minimum guaranteed bandwidth on each wavelength. Namely, we should determine the minimum guaranteed bandwidth for 10G ONUs on 10G channel denoted by $B_{XG}(\lambda_{XG})$ and that for 1G-ONUs on 1G and 10G channels denoted respectively by $B_{IG}(\lambda_{XG})$ and $B_{IG}(\lambda_{IG})$. Formally, the value of the minimum guaranteed bandwidth is determined by the polling cycle length ($T_c$), which is in turn determined by the scheduling algorithm. Therefore, to obtain a reasonable estimate for minimum guaranteed bandwidth, we have to assume a typical value for the cycle time, e.g., $T_c = 2$ msec. Considering the transmission windows on the 10G channel, we find:

$$T_c = \frac{N_{XG} \cdot B_{XG}(\lambda_{XG}) + n_{IG} \cdot B_{IG}(\lambda_{XG})}{R_{XG}}$$

where $N_{XG}$ is the number of 10G-TDM ONUs, $R_{XG}$ ($R_{XG}$) is the effective data rate of 1G (10G) channels, and $n_{IG}$ is the number of 1G-ONUs which are decided to be scheduled on the 10G channel. We assume that a 10G-ONU can transmit up to 10 times more bytes during a $T_c$ period than a 1G-ONU, that is, $B_{XG}(\lambda_{XG}) \approx 10B_{IG}(\lambda_{XG})$. Then, we obtain:

$$B_{XG}(\lambda_{XG}) = \frac{T_c}{R_{XG}} \cdot \frac{N_{XG} \cdot B_{XG}(\lambda_{XG}) + n_{IG} \cdot B_{IG}(\lambda_{XG})}{R_{XG}}$$

The minimum guaranteed bandwidth for the rest of 1G-ONUs on the 1G wavelength channels can be determined as follows:

$$B_{IG}(\lambda_{IG}) = \frac{M}{N_{IG} - n_{IG}} \times (T_c \cdot R_{IG})$$

where $M$ is the total number of 1G channels, and $N_{IG}$ is the total number of 1G-WDM ONUs. Next, we determine the allocated bandwidth for each ONU. Let $Q_i$ be the requested bandwidth and $P_i$ be the allocated bandwidth for ONU$_j$. If the requested bandwidth is less than the minimum guaranteed bandwidth, then the whole request will be granted. Otherwise, if the requested bandwidth is greater, the grant size of ONUs will be reduced to meet the following inequality:

$$B_{min}(i, j) \leq P_j \leq Q_j \quad \forall j: Q_j > B_{min}(i, j)$$

where $B_{min}(i, j)$ is the guaranteed bandwidth of ONU$_j$ on channel $i$ that can be determined from (2), or (3) (it is obvious that $B_{min}(i, j) = 0$ for 10G-ONUs on 1G channels).

![Fig. 4. Dynamics of bandwidth allocation](image-url)
B. Delay Analysis

We assume that each ONU request is generated according to its instantaneous data rate $R_{ij}^a$ on the US channel; that is, during each polling cycle, the ONU generates a bandwidth request of $Q_j = T_c R_{ij}^a$. The delay analysis is illustrated in Fig. 4. The upper axis illustrates the buffer occupancy of ONU $j$, while the lower axis shows the data transmission on the US channel assigned for ONU $j$. In the first cycle, the ONU transmits the packets stored in its buffer and reports its current buffer occupancy. In the second cycle, the ONU transmits what has been requested (and allocated) in the first cycle; meanwhile, the ONU buffer is being filled with newly generated packets, which will be scheduled for transmission in the third scheduling cycle.

Therefore, the data generated at time $s_j + t_j$ will be transmitted at $s_j + 2T_j$, where $s_j$ is the start time and $t_j$ is the length of the window allocated for ONU $j$. If the ONU data request is small enough to be transmitted in one cycle, then the maximum delay is:

$$D_j = 2T_c - t_j \quad \forall j : P_j = Q_j$$

In general, as depicted in Fig. 4, if the allocated bandwidth is less than the ONU request, some packets will be accumulated until the ONU buffer is full. Therefore, the maximum packet delay can be obtained as:

$$D_j = (N_j^C + 2) T_c - t_j - N_j^C \left( \frac{P_j}{R_{ij}^a} \right) \quad \forall j : P_j < Q_j$$

where $N_j^C$ is the number of polling cycles elapsed until the buffer of ONU $j$ is full. That is:

$$N_j^C = \left\lfloor \frac{F_j}{Q_j - P_j} \right\rfloor$$

where $F_j$ is the buffer size of ONU $j$.

C. ILP Model

In [4], the offline scheduling problem in a WDM-PON is formulated as a non-joint optimization problem, where the grant sizing is done in advance using a bandwidth allocation method such as “limited service” or “gated service” [3]. The authors presented an ILP model based on the scheduling theory, in that, each ONU is considered as a job, its grant sizing is done in advance using a bandwidth allocation method such as “limited service” or “gated service” [3]. The authors presented an ILP model based on the scheduling theory, in that, each ONU is considered as a job, its grant sizing is done in advance using a bandwidth allocation method such as “limited service” or “gated service” [3].

Our only channel restriction is that the 10G-ONUs can only transmit on the PON represent machines. Therefore, the problem reduces to a Parallel Machine (PM) scheduling problem, where a set of jobs, with specific processing times, are executed on a set of machines. Using the same concept, we derive an ILP model for joint scheduling and bandwidth allocation in 10G-EPON coexisting with 1G WDM-PON. Our only channel restriction is that the 10G-ONUs can only be granted on the 10G channel. In our model the size of transmission window for ONU $j$ on channel $i$ is a variable which is determined inside the model along with the schedule.

**INPUT PARAMETERS:**

$O_T = \text{set of all existing ONUs} (O_T = O_{1G} \cup O_{XG})$

$\Lambda_T = \text{set of all transmission channels} (\Lambda_T = \Lambda_{1G} \cup \{\lambda_{XG}\})$

$\Delta_j^{max} = \text{maximum affordable delay per ONU } j$

$\delta_{ij} = \begin{cases} 0 & \text{if } i \text{ is a 1G channel and } j \text{ is a 10G-ONU} \\ 1 & \text{otherwise.} \end{cases}$

**OUTPUT VARIABLES:**

$C_{max} = \text{maximum completion time of all channels}$

$t_{ij} = \text{length of transmission window for ONU } j \text{ on channel } i$

$x_{ikj} = \begin{cases} 1 & \text{if position } k \text{ on } \lambda_i \text{ is selected for ONU } j \\ 0 & \text{otherwise.} \end{cases}$

$y_{ikj} = t_{ij} \times x_{ikj}$

$\alpha_{ij} = \begin{cases} 1 & \text{if channel } i \text{ is assigned to ONU } j \\ 0 & \text{otherwise.} \end{cases}$

Our objective is to minimize the maximum completion time:

$$\min C_{max}$$

In order to avoid quadratic terms in our model, we define variables $y_{ikj} = t_{ij} \times x_{ikj}$. Using these variables, we write constraint (9) to assure that the makespan is greater than the completion time of each channel.

$$C_{max} \geq \sum_{j \in O_T} \sum_{k \in O_T} k \delta_{ij} y_{ikj} \quad i \in \Lambda_T \tag{9}$$

Variables $y_{ikj}$ are determined in the following set of linearization constraints where $L$ is a large positive number:

$$y_{ikj} \leq t_{ij} + L(1 - x_{ikj}) \quad i \in \Lambda_T, k, j \in O_T; \delta_{ij} = 1 \tag{10}$$

$$y_{ikj} \geq t_{ij} - L(1 - x_{ikj}) \quad i \in \Lambda_T, k, j \in O_T; \delta_{ij} = 1 \tag{11}$$

$$y_{ikj} \leq L \times x_{ikj} \quad i \in \Lambda_T, k, j \in O_T; \delta_{ij} = 1 \tag{12}$$

Constraints (13) and (14) determine the channel and time slot assignment of upstream bandwidth request of each ONU based on the parallel machine model presented in [4]. Constraint (13) ensures that each ONU is assigned to only one scheduling position and constraint (14) guarantees that each scheduling position is assigned to no more than one ONU.

$$\sum_{i \in \Lambda_T} \sum_{k \in O_T} x_{ikj} = 1, \quad j \in O_T \tag{13}$$

$$\sum_{j \in O_T} x_{ikj} \leq 1, \quad i, k \in \Lambda_T. \tag{14}$$

Constraint (15) indicates that on each scheduling round, only one channel is assigned for each ONU. Also, by involving the $\delta_{ij}$ parameter, this constraint implies that the 10G-ONUs can not be allocated on 1G channels.

$$\alpha_{ij} = \delta_{ij} \sum_{k \in O_T} x_{ikj}, \quad i \in \Lambda_T, j \in O_T \tag{15}$$

Constraints (16)-(18) are required for bandwidth allocation based on the discussion in Section V-A. In these constraints, for all $i \in \Lambda_T$ and $j \in O_T$ we have:

$$R_{ij}^T \times t_{ij} = \alpha_{ij} \times Q_j \quad Q_j \leq B_{min}(i, j) \tag{16}$$

$$R_{ij}^T \times t_{ij} \leq \alpha_{ij} \times Q_j \quad Q_j > B_{min}(i, j) \tag{17}$$

$$R_{ij}^T \times t_{ij} \geq \alpha_{ij} \times B_{min}(i, j) \quad Q_j > B_{min}(i, j) \tag{18}$$

Constraint (20) is for limiting the maximum packet delay per each ONU. Note that expressions (6) and (7) include nonlinear terms which can not be involved in our ILP model.
Hence, to keep the model linear and to obtain a reasonable delay, we approximate \( N_j^C \) as:

\[
N_j^C \approx \left[ \frac{F_j}{Q_j - B_{\min}(i,j)} \right]
\]  

(19)

We replace \( N_j^C \) with \( N_j^C \), and rewrite equation (7) in the form of an ILP constraint as follows:

\[
(N_j^C + 2)T_c - t_{ij} - \left( N_j^C \cdot \frac{R_j^T}{R_j} \right) t_{ij} \leq L(1 - \alpha_{ij}) + \Delta_j^{max}
\]

\[i \in \Lambda_T, j \in \Lambda_T; Q_j > B_{\min}(i,j)\]  

(20)

where \( T_c \) is the total scheduling length considering the initial gap on each cycle, i.e., \( T_c = ISC + C_{\max} \). This constraint implies that the delay for ONU \( j \) is less than a predetermined value \( \Delta_j^{max} \), when ONU \( j \) is scheduled on \( \lambda_i \).

VI. A Tabu Search Heuristic for Solving the Scheduling Problem

The ILP model developed in Section V-C can not be solved for large network instances. In addition, we have to make an approximation for the delay expression in (6) in order to avoid nonlinearity in the model. Thus, it is vital to develop a heuristic in order to involve the nonlinear terms and obtain near-optimal solutions in a reasonable amount of time. To this end, we develop a Tabu search method for solving the joint scheduling and bandwidth allocation problem. Our Tabu search heuristic starts from an initial solution and iterates using two types of moves for achieving the neighbor solution. One is reordering and moving the ONU grants from one wavelength to another supported wavelength. In this move, the neighborhood of the current solution is obtained by moving a transmission window of an ONU from its assigned wavelength to another supported wavelength. Note that this move can not be applied for the grants of 10G-ONUs, since they can only be transmitted on the 10G channel. The other move is reducing the transmission window sizes of different ONUs. This move offers more options for swapping ONU grants between different supported channels. While trying to reduce the grant size, we should take care of the imposed delay given by expression (6) not to exceed a predetermined value.

Using these two moves, our Tabu search algorithm performs a local search to explore new feasible solutions. In each iteration of the procedure, both moves are assessed, and the one that yields the best result is chosen as the final move to be performed. Our Tabu method also makes use of a short term memory (Tabu list) that stores information associated with recently explored solutions in order to avoid cycling. For example, the Tabu list contains the positions of the swapped grants, and any move that schedules an ONU grant back to its old position is considered Tabu (i.e., forbidden). The Tabu search algorithm needs some stopping criteria. One stopping criterion is to iterate for a certain number of iterations depending on the number of ONUs. Furthermore, we note that the percentage of utilization of different channels varies as the grants of ONUs are resized and reordered among different channels. Therefore, we consider another stopping criterion such that the algorithm runs until the last scheduled ONUs on all wavelength channels have the same finishing time. This is equivalent to maximizing the average bandwidth utilization.

VII. Numerical Results

We evaluate our ILP and heuristic models based on the network architecture in Fig. 2. First, we conduct three experiments assuming that the bandwidth requirement of each 1G-ONU is randomly uniformly distributed over the interval of [0.5B_{1G}(\lambda_{1G}), 0.9B_{1G}(\lambda_{1G})], [0.9B_{1G}(\lambda_{1G}), 1.3B_{1G}(\lambda_{1G})], and [1.3B_{1G}(\lambda_{1G}), 1.7B_{1G}(\lambda_{1G})] for 1G1, 1G2, and 1G3 respectively. In all experiments, the required bandwidth of 10G-ONUs is taken from the interval [0.2B_{XG}(\lambda_{XG}), 0.7B_{XG}(\lambda_{XG})]. Furthermore, we assume a 100 ns round trip delay between each ONU and the OLT, which corresponds to a 10 km distance. The initial gap of each channel (ISC) is approximately set to 110 ns and the guard time between adjacent time slots is 1.5 ns.

We consider a network with 2 10G-ONUs, 8 1G-ONUs, and two 1G WDM channels. We compare our joint ILP model and Tabu heuristic with the non-joint ILP model presented in [8] which is a modified version of the model presented in [4]. For the non-joint model, we assume that the bandwidth allocation is carried out in advance using the “gated” grant sizing technique, where the grant size for an ONU is simply the queue size reported by that ONU [3]. Each scheduling method is employed for two cases: i.e., \( \lambda_{XG} \) is either shared with 1G-ONUs or it is dedicated to 10G-ONUs. As expected, a shared DBA module yields higher channel utilization and shorter scheduling periods, as \( \lambda_{XG} \) is more efficiently utilized by some of the 1G-ONUs.

Next, we investigate the performance benefits obtained by the users when they undergo rate upgrade from 1G to 10G. To this end, we evaluate the joint Tabu method on a network instance when the 1G- and 10G-ONUs are using the same DBA module. We consider a network consisting of six 1G WDM channels and a total of 40 ONUs. We conduct 5 experiments where the 1G-ONUs are gradually being upgraded to 10G. As can be seen in Table II, increasing the number of 10G-ONUs yields a smaller makespan and shorter maximum expected delays (measured as the ratio of the sum of maximum delay per ONU obtained from expressions (5) and (6) to the total number of ONUs). The reason is that when an ONU is upgraded to 10G, its requested bandwidth becomes much less than the minimum guaranteed bandwidth of 10G-ONUs given by expression (2). Therefore, the upgraded ONU would be allocated its whole request on the 10G channel with a 10 times smaller transmission window. However, the average channel utilization decreases, since the request of new upgraded 10G-ONUs can only be transmitted on 10G channel and therefore
the utilization of 1G channels decreases. This means that the network becomes under-utilized and its potential bandwidth becomes wasted if the offered load is not increased to catch up with the line upgrade to 10Gbps. In other words, to exploit the new capabilities of the upgraded system, more subscribers should be granted, resulting in higher traffic loads.

Table I

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Non-joint ILP</th>
<th>joint Tabu</th>
<th>Non-joint ILP</th>
<th>joint Tabu</th>
<th>joint ILP</th>
</tr>
</thead>
<tbody>
<tr>
<td>E₁</td>
<td>1696</td>
<td>1695</td>
<td>1698</td>
<td>1355</td>
<td>1342</td>
</tr>
<tr>
<td>E₂</td>
<td>2498</td>
<td>2003</td>
<td>2475</td>
<td>1881</td>
<td>1715</td>
</tr>
<tr>
<td>E₃</td>
<td>3375</td>
<td>2003</td>
<td>3189</td>
<td>2506</td>
<td>2402</td>
</tr>
</tbody>
</table>

Average Makespan (µsec)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Non-joint ILP</th>
<th>joint Tabu</th>
<th>Non-joint ILP</th>
<th>joint Tabu</th>
<th>joint ILP</th>
</tr>
</thead>
<tbody>
<tr>
<td>E₁</td>
<td>73.47</td>
<td>73.46</td>
<td>73.47</td>
<td>90.67</td>
<td>91.32</td>
</tr>
<tr>
<td>E₂</td>
<td>71.04</td>
<td>72.33</td>
<td>71.07</td>
<td>93.05</td>
<td>93.46</td>
</tr>
<tr>
<td>E₃</td>
<td>90.39</td>
<td>73.66</td>
<td>90.04</td>
<td>93.77</td>
<td>95.33</td>
</tr>
</tbody>
</table>

Expected Maximum Delay (µsec)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Non-joint ILP</th>
<th>joint Tabu</th>
<th>Non-joint ILP</th>
<th>joint Tabu</th>
<th>joint ILP</th>
</tr>
</thead>
<tbody>
<tr>
<td>E₁</td>
<td>3782</td>
<td>2655</td>
<td>3236</td>
<td>2655</td>
<td>2630</td>
</tr>
<tr>
<td>E₂</td>
<td>19984</td>
<td>4736</td>
<td>6317</td>
<td>3447</td>
<td>4024</td>
</tr>
<tr>
<td>E₃</td>
<td>21048</td>
<td>15152</td>
<td>19473</td>
<td>4267</td>
<td>17741</td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>NₙG</th>
<th>N₁G</th>
<th>Makespan (µsec)</th>
<th>Expected Maximum Delay (µsec)</th>
<th>Average Channel Utilization (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>32</td>
<td>2698</td>
<td>23269</td>
<td>92.03</td>
</tr>
<tr>
<td>12</td>
<td>28</td>
<td>2614</td>
<td>20997</td>
<td>90.32</td>
</tr>
<tr>
<td>16</td>
<td>24</td>
<td>2562</td>
<td>17404</td>
<td>83.11</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>2300</td>
<td>6022</td>
<td>80.97</td>
</tr>
<tr>
<td>24</td>
<td>16</td>
<td>2316</td>
<td>4783</td>
<td>73.37</td>
</tr>
</tbody>
</table>

VIII. CONCLUSION

We studied the problem of optimal scheduling and bandwidth allocation in next generation 10G-EPON coexisting with 1G WDM-PONs. We derived an ILP model for offline joint scheduling and bandwidth assignment for providing this coexistence. For large network instances, the size of the ILP model becomes prohibitively large. Therefore, we introduced a Tabu Search heuristic to achieve near optimal solutions in notably shorter computation times. We explored the tradeoff which exists in terms of delay, scheduling length, and channel utilization, when separate or the same DBA and scheduling modules are used for 1G- and 10G-ONUs. We also indicated the influence of gradually upgrading the 1G-ONUs to 10G-ONUs in a network with fixed number of ONUs. We conclude that upgrading WDM 1G-ONUs to TDM 10G-ONUs improves the quality of service experienced by end users, yet it decreases the channel utilization on the existing 1G channels.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Lehan Meng for his help in implementing the Tabu heuristic method.

REFERENCES


