



3-Stage Hierarchical Quality of Service for Multi-tenant Passive Optical Networks

Frank Slyne, Bruno Cornaglia, Marco Boselli, Marco Ruffini

► To cite this version:

Frank Slyne, Bruno Cornaglia, Marco Boselli, Marco Ruffini. 3-Stage Hierarchical Quality of Service for Multi-tenant Passive Optical Networks. 23th International IFIP Conference on Optical Network Design and Modeling (ONDM), May 2019, Athens, Greece. pp.193-203, 10.1007/978-3-030-38085-4_17. hal-03200648

HAL Id: hal-03200648

<https://inria.hal.science/hal-03200648>

Submitted on 16 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

3-stage Hierarchical Quality of Service for Multi-tenant Passive Optical Networks

Frank Slyne¹, Bruno Cornaglia², Marco Boselli² and Marco Ruffini¹

¹ Connect Research Center , Trinity College, Dublin, Ireland

² Vodafone Group Technology, Turin, Italy
fslyne@tcd.ie

Abstract. Passive Optical Networks (PONs) are an economically efficient means of providing high bandwidth services to end users because end-points can share the cost and the benefit of fibre optic connection. There is a growing interest in the use of PONs as a bearer for services such as Long Term Evolution (LTE), however work remains to be done to ensure stable, low latency and low packet loss for full duplex Transmission. In this paper we are mostly concerned with the characteristics of the downstream channel so as to further the case for PON sharing. We review typical downstream schedulers used in the industry and propose a three-stage downstream scheduler that assures quality of service, accurately, across Virtual Network Operators (VNOs) in multi-tenant PONs environment. We benchmark our scheduler against other algorithms, obtaining almost ideal scheduling performance.

Keywords: Quality of Service, Two Rate Three Color Marker, Scheduler, Policier, Passive Optical Networks, Virtualization, Slicing.

1 Introduction

Up until recently, residential broadband has been the dominant service delivered over PONs. However, because PONs are an economically efficient means of providing high capacity bandwidth across wide rural and urban geographical areas [1], they are becoming increasingly attractive for the delivery of non-retail services such as LTE [2], where dedicated fibre capacity would not be economically feasible. PON sharing across multiple VNOs is also becoming increasingly important, in order to foster competition without replicating costly hardware infrastructure. Thus, in a multi-tenant scenario each VNO should have the ability to assign its portion of assured and best effort traffic to its customers, without affecting or being affected by other VNOs. Given that LTE, when implemented as Cloud Radio Access Networks (C-RAN) has a low tolerance to excessive packet jitter and latency, the VNO needs granular slicing and resource control in order to share PON infrastructure with non-critical services such as residential broadband.

Achieving separation between VNOs in a PON environment is complicated by the fact that upstream and downstream transmissions function in different manners. Unfortunately, granular resource control is only defined for upstream traffic in the PON standards, with downstream left to the devices of operators and whatever capabilities

their vendor supplied equipment provide. By its nature, upstream transmission in a PON is a TDM channel where there is an inherent control over how much and when each Traffic Container (TCont) of each Optical Network Terminations (ONT) can transmit, down to the granularity of each 32-bit word [3]. In this context a TCont can be thought of as a dedicated queue at the ONT to which a specific class of service can be assigned. The scheduling of the upstream transmission is regulated by the Dynamic Bandwidth Allocation (DBA) mechanism [4], which allocates capacity to the ONTs located at the end user side [5]. One solution to provide full control to the VNOs over their upstream capacity allocation, by virtualizing and slicing the DBA mechanism, was recently proposed in [6], prototyped in [7] and standardized in [8].

Of most pertinence in this paper, is the downstream channel of the PON which functions in the manner of a statistically multiplexed packetized channel or shared bus. The packetization of data followed by statistical multiplexing through a shared transmission channel is an essential means of increasing the utilization of the channel, when compared to time or frequency division multiplexing means over the same channel. This technological trend is similar to how synchronous SDH and cell-based ATM networks in the core and metro networks are being replaced by multi-gigabit Ethernet [9] networks running over Dense Wave Division Multiplexing (DWDM) fiber [10], at times over an Optical Transmission Network (OTN) sublayer [11]. However, a drawback for real-time services that use a statistical multiplexed channel is the lack of a guarantee over the data information rate, packet loss and latency, which go together to define QoS. As central office infrastructure is moving towards virtualization and slicing [12], the ability to provide assurance over QoS is becoming increasingly important to provide slice independence across multiple tenants, as though they were being served by dedicated infrastructure. This is typically achieved through appropriate queue management, to separate traffic with different QoS requirements and policing the rate of incoming traffic. However, such methodology is not defined in PON standards, where decisions on how to implement downstream QoS are left to vendors and operators. Typically, the options vary between over-dimensioning capacity on each PON, executing vendor specific scheduling and policing, or shaping traffic prior to entering the Optical Line Termination (OLT) [2].

In this paper, we first review the current practices used by operators for the scheduling of downstream traffic in a PON, typically divided into one-stage and the two-stage scheduler architectures. We will show that while these methods can be effective in single-operator scenarios, they fail to provide the necessary support for VNOs sharing the same infrastructure. We thus propose a novel three-stage scheduler that includes a QoS stage for providing additional separation between VNOs. In order to make the case for the scheduler, we build the architecture for the one-, two- and three- stage schedulers in a simulator, which are then exercised using a common set of scenarios. The performance of the three simulators are then compared using typical metrics such as packet loss. Throughout the experiment, we are careful to distinguish between the packet handling capabilities for both High Priority (HP) and Low Priority (LP) traffic. Examples of HP traffic includes Signal data and traffic related to services that deteriorate significantly due to packet loss, such as Voice over IP (VoIP)[13]. Examples of LP traffic

of the trTCM are given below. The mechanism by which these queue management functions can be composed into typical downstream QoS schedulers are shown in Figure 1. Firstly, we see a single-stage QoS mechanism in Figure 1 (a), where QoS is only implemented at the ONT level, with WRR scheduling traffic within the same priority class, and SP operating across the classes to provide, for example, strict higher priority to voice and signaling with respect to video and other data types. Secondly, Figure 1 (b) shows an alternative two stages vendor solution. The first stage operates at the ONT level and is composed of 4 queues, a SP scheduler and trTCM policer. The SP scheduler handles traffic in the decreasing priority of signaling, voice and the weighted fair queuing balancing of video and data. The SP assigns a color to the traffic based on the source. Both signaling and voice are HP and thus marked as green. Video and data are instead LP and marked as yellow. Because the color of the traffic is already configured by SP, trTCM is configured in color-blind mode, according to RFC 4115[14], and passes the colored traffic directly to the egress Committed Information Rate (CIR) and Excess Information Rate (EIR) ports respectively. A weighed fair queuing is applied at the ONT level between video and data, while strict priority is assigned to voice and signaling. trTCM policers enforce proportionality between CIR and EIR by marking the packets according to RFC2698 (trTCM) color definition [15].

Packets that are marked as red are dropped, green packets are passed through, yellow packets are remarked packets that may be dropped at a later stage or passed if there is sufficient available capacity. At the OLT port level, a weighted round robin mechanism is applied to schedule packets across all ONTs, respecting aggregate levels of CIR and EIR. The second stage is the PON port module, composed of two WRR queues operating on the CIR and EIR queues. The weighting of the WRR queues is based on the total service profile, CIR and EIR respectively, that each ONT should carry as a proportion of the total capacity allowed in the PON port. For example, the weighting given to the packets from the i^{th} ONT

$$w_i^c = \frac{CIR_i}{\sum_{j=1}^k CIR_j}, w_i^e = \frac{EIR_i}{\sum_{j=1}^k EIR_j}$$

Where w_i^c and w_i^e are the fractions of configured CIR and EIR for that ONT over the total CIR and EIR respectively for all (k) ONTs. The total weightings for each WRR then sums to 1.

$$\sum_{i=1}^k w_i^c = 1, \sum_{i=1}^k w_i^e = 1$$

It should be noticed that QoS scheduling becomes especially important in multi-tenant scenarios, where different VNOs can have different QoS contracts with their own customers and with the infrastructure provider. However, current scheduling mechanisms such as those described above, do not take into consideration the existence of multiple VNOs, and thus there is no QoS differentiation at the VNO level. In this paper we attempt to tackle this issue by proposing a novel three-stage scheduler that

adds VNO-oriented scheduling layer and by comparing its performance improvement against the baseline mechanisms shown in Figure 1.

Finally, we can define the behavior of the ideal scheduler, against which current and proposed architectures can be compared. To do this, we calculate X_i^{HP} and X_i^{LP} the expected HP and LP traffic egressing from the PON related to the i^{th} ONT, given offered traffic Z_i^{HP} and Z_i^{LP} . The high priority traffic marked as green should not exceed the CIR allocated to that ONT. In the event, that HP traffic exceeds the CIR, the excess is recolored as yellow.

$$G_i^{HP} = \min(Z_i^{HP}, CIR_i), \quad Y_i^{HP} = Z_i^{HP} - G_i^{HP}$$

Any excess CIR may be used to allow LP traffic to be colored as green. The excess LP traffic is marked yellow.

$$G_i^{LP} = CIR_i - G_i^{HP}, \quad Y_i^{LP} = Z_i^{LP} - G_i^{LP}$$

The expected HP traffic egressing from the PON related to the i^{th} ONT is a sum of HP traffic marked as green and HP traffic marked as yellow contending for the total EIR EIR_T . w_i^c and w_i^e are as defined above.

$$X_i^{HP} = G_i^{HP} + w_i^c \cdot EIR_T \cdot \frac{Y_i^{HP}}{Y_i^{HP} + Y_i^{LP}}$$

Similarly, the expected LP traffic egressing from the PON related to the i^{th} ONT is a sum of LP traffic marked as green and LP traffic marked as yellow contending for the total EIR EIR_T .

$$X_i^{LP} = G_i^{LP} + w_i^e \cdot EIR_T \cdot \frac{Y_i^{LP}}{Y_i^{LP} + Y_i^{HP}}$$

In the next section we explain the architecture of our proposed three-stage QoS scheduler and the implementation of the simulation environment. This allows us to compare the results of our scheduler with the one-stage and two-stage schedulers described above, showing the advantages brought by our three-stage design for multi-tenant operations.

Three-Stage Scheduler Description and Architecture

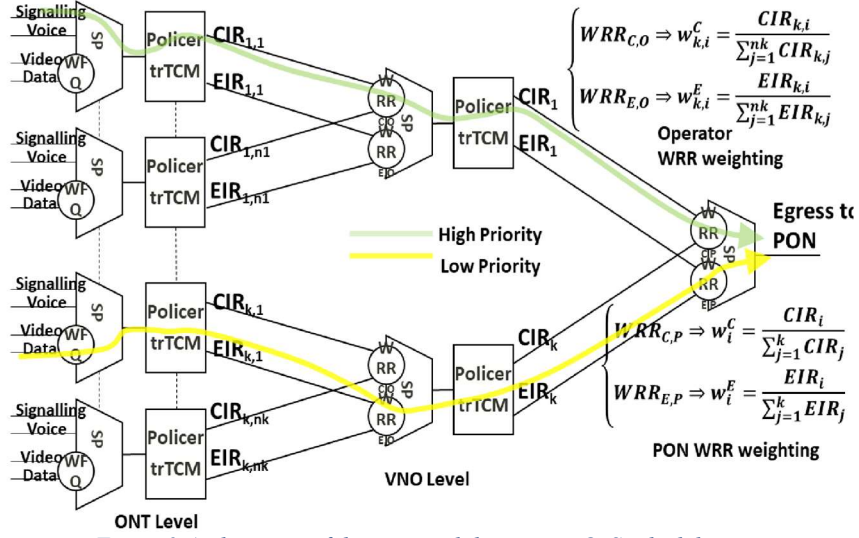


Figure 2 Architecture of the proposed three-stage QoS scheduler

The architecture of our Hierarchical three-stage scheduler is shown in Figure 2, and uses the same array of functions (SP, WRR, WFQ and trTCM) as the one- and two-stage schedulers. We note that in recent years the IETF has developed algorithms and techniques for Queue management and schedulers, such as Controlled Delay (CoDel) and Active Queue Management (AQM) respectively that deal with anomalous treatment of TCP/IP transport flows due to excess buffering of packets. This anomalous behavior is a result of a phenomenon known as Bufferbloat [16], that is, the use of excessively large buffers in routers, switches and modems, that disrupt how normal functioning of TCP/IP flow control mechanism. IETF has combined CoDel and AQM into a single RFC (8290) known as Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm (FQ-CoDel) [17]. With respect to the reference two-stage scheduler shown Figure 1 (b), we have added a further WRR followed by a trTCM function that operate at the individual VNO level. The scope of this additional stage is to further partition the overall capacity, so that any excess capacity from a VNO is preferentially re-distributed across the ONTs belonging to the same VNO. Hierarchical Scheduler is located as a single functional block at the head end of the PON, that is at the OLT. It functions at 3 levels, that is, at the level of the ONT, at the level of the Operator and at the level of the PON, where WRR, SP and WFQ assure proportionality between contracted bandwidths. Downstream from the ONT and Operator schedulers, trTCM policers enforce proportionality between CIR and EIR by marking the packets according to RFC2698 (trTCM) color definition [15]. Packets that are marked as Red are dropped, Green Packets are passed through, Yellow packets are remarked packets that may be dropped at a later stage or passed if there is enough bandwidth. In an ideal

scenario, but not implemented here, a feedback path from the PON functional block, similar to Call Admission Control (CAC), directs specific ONTs to block new flow requests, where the CIR is higher than a specific percentage (typically 50%) of the total capacity of the PON or indeed should the CIR of an operator exceed a specific threshold agreed with the PON infrastructure provider.

Simulator

In order to compare the QoS performance improvement of the three-stage scheduler in PON multi-tenant scenarios we have developed a QoS discrete event simulator based on the Python SimPy [18] package. We use several fundamental QoS modules (trTCM, WRR, WFQ and SP) to construct the downstream PON schedulers operating in the OLT. The trTCM object class marks packets with the three colors: green, yellow and red, depending both on any original marking and on the CIR and the EIR parameters. It should be noticed that the decision on whether incoming data is above or below the CIR and EIR thresholds is computed based on average values across a given burst data size. In the simulation we used burst sizes values for the CIR and the EIR typically configured in Cisco and Juniper routers of 64 Kbytes and 128 Kbytes, respectively. The WRR module accepts packets from several inputs streams and assures fairness based upon a predefined distribution based on number of packets in the input buffer for each stream. WFQ behaves similarly to WRR but uses number of bytes in the input buffer for each stream as a means of determining the fairness. Finally, SP selects packets from different queues in ascending order of priority and being strict it will only select packets from a lower priority queue once the higher priority ones are empty.

The ONT module is made up of 4 queues each with a buffer size of 128 Kbytes, a SP scheduler and trTCM policer. The SP scheduler handles traffic in the decreasing priority of signaling, voice and the weighted fair queuing balancing of video and data. The SP assigns a color to the traffic based on the source. Both signaling and voice are HP and thus marked as green. Video and data are instead LP and marked as yellow. Because the color of the traffic is already configured by SP, trTCM is configured in colorblind mode, according to RFC 4115[14], and passes the colored traffic to the egress CIR and EIR ports respectively. This is constructed similarly to the ONT QoS module, however there is a pair (that is for both CIR and EIR) of WRR schedulers for each VNO, whose weightings are calculated based on the total capacity assigned to the specific VNO. Thereafter a trTCM policer colors the aggregate traffic from the SP scheduler as passed (green), remarked (yellow) or dropped (red). The PON port QoS module is composed of two WRR queues whose weighting are based on the aggregate service profiles, CIR and EIR respectively, that each other operator carries. The PON QoS module does not have a policer, but instead must aggregate both CIR and EIR onto the common downstream physical link. To assess the performance of the Hierarchical QoS (H-QoS), we also emulate one and two stage QoS architectures. The one stage architecture is composed of just the PON module from the three-stage architecture, which has been extended to have 4 WRR queues, terminating the Signaling, Voice, Video and Data links. The weighting of the WRR queues is based on the total service

profile CIR and EIR respectively that each other ONT carries as a proportion of the total for the PON.

The two-stage architecture Figure 1(b) is composed of the ONT and PON modules from the three-stage architecture Figure 2, omitting the Operator QoS modules. The ONTs connect directly to the PON. The weighting of the PON WRR queues is based on the total service profile CIR and EIR respectively that each other ONT carries as a proportion of the total for the PON.

Simulation Scenarios and Results

In our simulation, we assume that there are 2 VNOs which share a common PON downstream capacity of 2.488 Gbps [3], of which 1.760 Gigabits per second (Gbps) is allocated to CIR. We define a service profile as a fixed mix of committed and excess information rates that may be assigned by any VNO to one of its ONTs. We denote this mix as a tuple (CIR, EIR). In our simulations, we define two service profiles, Profile-1 (10,100) and Profile2 (100,1000), measured in Megabits per second (Mbps). In order to gauge the ability of the QoS mechanisms described to schedule capacity appropriately across VNOs we introduce an imbalance between their user base. Operator A has 24 ONTs, of which half are Profile-1 and half are Profile-2.

Operator B has 4 ONTs with Profile-1 and 4 with Profile-2. The net result is that Operator A carries three quarters of the total traffic and Operator B carries one quarter. We are especially interested in what happens to bandwidth which is offered to an ONT that is either in excess or in deficit of its service profile's CIR and EIR, which we call the Nominal information rates. For this purpose, traffic offered at each ONT is transmitted at either +20% or -20% of the nominal CIR or EIR for the configured service profile. For instance, ONTs with service profile Profile-1 may be offered (8,80), (8,120), (12,80) or (12,120) Mbps. Finally, in addition to providing a comparison between our three-stage QoS mechanism and the two other baseline algorithms described in section 1, we also provide the output of an ideal scheduler as described previously, under the same conditions, that is, whose output is computed mathematically rather than simulated. This allows use to assess the three QoS architectures against an ideal scheduler.

The results of our simulations are reported in Figure 3 showing the throughput of the different schedulers for HP and LP traffic. The x axis represents the different ONTs, which are identified by the triplet (x:y:z), where x identifies the VNO (A or B), y is the (CIR/EIR) profile (i.e., Profile-1 or Profile-2, as described in the previous section) and z is the ONT identifier. To simplify the plot, we only show the results for a subset of ONTs, which is representative of the performance provided by the three schedulers. In the Figure 4(a) plot, we see that both the proposed three-stage (in orange), the baseline two-stage (in gray) and the baseline one-stage (in yellow) algorithms follow the ideal output (shown in blue) quite closely. This is to be expected since the purpose of CIR is to strictly cap the capacity at the agreed levels, which can be easily enforced already at the ONT level.

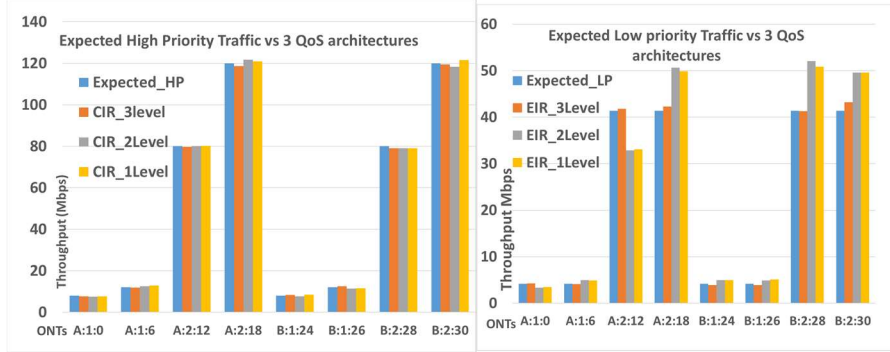


Figure 3 Comparison of (a) High Priority (HP) and (b) Low Priority (LP) Traffic

As we see in Figure 3, where the three-stage mechanism provides considerable improvements over the two-stage and the one-stage is in the handling of LP traffic. Since the EIR traffic is only scheduled if there is spare capacity, our additional VNO-level QoS can reassign such capacity appropriately within each VNO.

The three-stage algorithm provides a much closer output to the ideal scheduler both across ONTs and across VNOs, showing a maximum deviation of $\pm 5\%$, compared to the two-stage scheme that deviates between $\pm 25\%$. This accuracy is highly desirable in a multi-tenant Fixed Access Network Sharing (FANS) scenario [19], where each VNO pays the infrastructure provider for assuring a given level of aggregate CIR and EIR. The one-stage shows a $\pm 26\%$ deviation, however this on the basis that video, and data share the same WRR queue. If this were the not case, then the video queue would always be served at the expense of the data (least priority) queue in this scenario. Figure 4 shows the capacity variance for HP traffic (top graph) and for LP traffic (bottom graph) with respect to the ideal case for the three different scheduling algorithms (labeled three-stage, two-stage and one-stage).

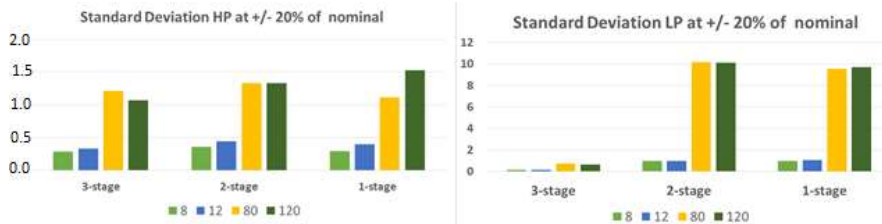


Figure 4 Variance of (a) HP data and (b) LP data against expected ideal behavior

The different bars group the results for all ONTs that offer, respectively, 8, 12, 80 and 120 Mb/s of HP traffic (the first two rates are for ONTs in Profile-1 with a 10Mb/s CIR and the second two rates for ONTs in Profile-2 with 100 Mb/s of CIR). We see that the proposed three-stage architecture provides the lowest variance for all traffic levels, especially for the allocation of capacity to LP traffic. Finally, the one-stage architecture performs poorest with a deviation from the desired output between $+80\%$ and $+200\%$.

Conclusions

In this paper, we have presented a three-stage scheduling architecture that can be implemented at the head-end of the PON or the OLT. This has several benefits. Firstly, the OLT has full knowledge of the traffic that is being offered to the PON and provides real-time insight to network planners as to how the network is complying with contracted QoS metrics. Indeed, the three-stage scheduler is in the best position to react to changing network demands. Secondly, while the data-path of the three-stage scheduler may be implemented in high-speed Application Specific Integrated Circuit (ASIC), the parameters of the WRR, WFQ and trTCM can be dynamically tuned at the discretion of a centralized Software Defined Network (SDN) control plane or orchestrator. Thirdly, the centralization of the scheduler simplifies the functionality of the ONT, allowing for easier functional upgrades, and virtualization of the PON.

We have assumed that the CIR for each operator is equal to the sum of the CIR of all its customers. However, in our future work, we will analyze where the CIR is less than the sum of the CIR of all its customers (over-subscription) as well as greater than the sum of the CIR of all its customers (under-subscription).

Acknowledgement

Financial support from Science Foundation Ireland (SFI) grants 14/IA/2527 (OSHAIRE) and 13/RC/2077 (CONNECT) is gratefully acknowledged

References

1. Ruffini M, Payne DB, Doyle L. Protection strategies for long-reach PON. In 36th European Conference and Exhibition on Optical Communication 2010 Sep 19 (pp. 1-3). IEEE.
2. Alvarez P, Slyne F, Bluemm C, Marquez-Barja JM, DaSilva LA, Ruffini M. Experimental Demonstration of SDN-controlled Variable-rate Fronthaul for Converged LTE-over-PON. In 2018 Optical Fiber Communications Conference and Exposition (OFC) 2018 Mar 11 (pp. 1-3). IEEE.
3. I. Rec, "G. 987. 6-2012," 10-Gigabit-capable passive optical network (XG-PON) systems, 2012.
4. ———, "G. 983. 1-1998," Broadband optical access systems based on Passive Optical Networks (PON), 1998.
5. Skubic B, Chen J, Ahmed J, Wosinska L, Mukherjee B. A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON. *IEEE Communications Magazine*. 2009 Mar;47(3):S40-8.
6. Elrasad A, Afraz N, Ruffini M. Virtual dynamic bandwidth allocation enabling true PON multi-tenancy. In Optical Fiber Communication Conference 2017 Mar 19 (pp. M31-3). Optical Society of America.
7. Slyne F, Giller R, Singh J, Ruffini M. Experimental Demonstration of DPDK Optimized VNF Implementation of Virtual DBA in a Multi-Tenant PON. In 2018 European Conference on Optical Communication (ECOC) 2018 Sep 23 (pp. 1-3). IEEE.

8. Broadband Forum, "BBF technical report tr-402, functional model for PON abstraction interface," in Optical Fiber Communication Conference. Broadband Forum, 2018.
9. Winzer PJ. Beyond 100G ethernet. *IEEE Communications Magazine*. 2010 Jul;48(7):26-30.
10. Anderson J, Traverso M. Optical transceivers for 100 gigabit Ethernet and its transport [100 gigabit ethernet transport]. *IEEE Communications Magazine*. 2010 Mar;48(3):S35-40..
11. Roese J, Braun RP, Tomizawa M, Ishida O. Optical transport network evolving with 100 gigabit Ethernet evolving with 100 gigabit ethernet [100 gigabit ethernet transport]. *IEEE Communications Magazine*. 2010 Mar;48(3):S28-34.
12. Ruffini M, Slyne F. Moving the Network to the Cloud: the Cloud Central Office Revolution and its Implications for the Optical Layer. *Journal of Lightwave Technology*. 2019 Jan 10.
13. Collins D. Carrier grade voice over IP. New York: McGraw-Hill; 2003..
14. Aboul-Magd O, Rabie S. A differentiated service two-rate, three-color marker with efficient handling of in-profile traffic.
15. Heinanen J, Guérin R. A single rate three color marker. 1999.
16. Gettys J. Bufferbloat: Dark buffers in the internet. *IEEE Internet Computing*. 2011 Apr 19(3):96.
17. Hoeiland-Joergensen T, McKenney P, Taht D, Gettys J, Dumazet E. The Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm. 2018..
18. Matloff N. Introduction to discrete-event simulation and the simpy language. Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August. 2008 Aug;2(2009):1-33.
19. Cornaglia B, Young G, Marchetta A. Fixed access network sharing. *Optical Fiber Technology*. 2015 Dec 1;26:2-11.