Hernández, J. A., Ebrahimzadeh, A., Maier, M., & Larrabeiti, D. (2021). Learning EPON delay models from data: a machine learning approach. Journal of Optical Communications and Networking, 13(12), 322-330.

DOI: https://doi.org/10.1364/jocn.437414

Learning EPON Delay Models From Data: A Machine Learning Approach

JOSÉ ALBERTO HERNÁNDEZ[1,*], AMIN EBRAHIMZADEH[2], MARTIN MAIER[3], AND DAVID LARRABEITI[1]

[1] Universidad Carlos III de Madrid, Spain
[2] Concordia University, Montréal, Canada
[3] INRS, Montréal, Canada
[*] Corresponding author: josealberto.hernandez@uc3m.es

Compiled May 28, 2024

There have been a large number of studies focused on the characterization of the upstream delay in TDM-PONs. However, most of them focus on finding equations for the average delay, and ignore other useful metrics like delay percentiles, which are of paramount interest in dimensioning PONs with delay guarantees. This work shows how to learn delay models from data using supervised Machine Learning (ML) techniques. Essentially, a non-linear regression ML algorithm is trained with PON simulation data, showing that it can provide accurate equations for such metrics of interest. In particular, we obtain R^2-score above 80% under Poisson traffic and above 65% under self-similar traffic, and provide a general equation for any delay percentile in the upstream channel of a PON employing IPACT. We further show its applicability in dimensioning Tactile Internet and 5G transport support scenarios.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Passive Optical Networks (PONs) have become the technology of choice in a large number of countries for providing broadband access to both residential users and business customers as well as for realising mobile backhaul infrastructures. In fact, PON technologies are anticipated to accelerate 5G deployments [1]. To achieve a convergence gain by integrating fixed and mobile infrastructures, an important cornerstone in today's network operator strategy is the use of a common transport platform in a PON-based fiber-to-the-home (FTTH) and small-cell roll-out scenario in support of future 5G low-latency applications [2].

One of the most interesting 5G low-latency applications is the Tactile Internet, which envisions the real-time transmission of not only audiovisual and data traffic but also haptic information (i.e., touch and actuation) for the remote control of physical and/or virtual objects. An example Tactile Internet application is the teleoperation with touch haptics [3, 4]. As illustrated in Fig. 1, the delay tolerance requirements of teleoperation largely depend on the dynamics of a given remote task environment, which may change for time-varying tasks, ranging from quality-of-control (QoC) requirements of as low as 1 ms in highly dynamic environments to less stringent quality-of-experience (QoE) requirements of 10-100 ms or even above for low to medium environment dynamics [3]. In highly dynamic environments with latency requirements of under 1 ms, only control systems can undertake the completion of tasks because humans are not capable of performing such rapid interactions. An example of such QoC systems is the fully automatic control of self-driving vehicles through intersections without any traffic lights and human involvement [3].

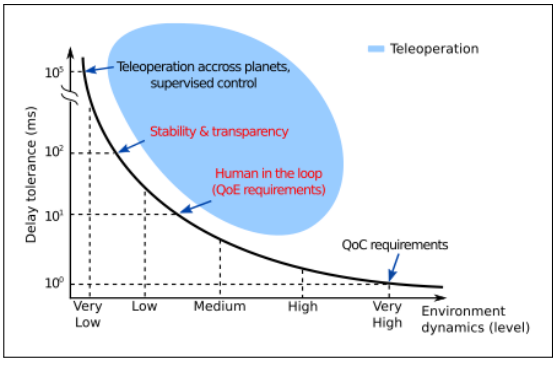

Fig. 1. Delay tolerance requirements of teleoperation as a function of remote environment dynamics [3].

Extensive studies have been carried out on developing analytical models and frameworks for evaluating the delay performance of PON-based network infrastructures under the assumption of different Dynamic Bandwidth Allocation (DBA) algorithms, e.g., the widely studied IPACT algorithm for conventional time division multiplexing (TDM) Ethernet PON

(EPON) [5] as well as the generic DBA algorithm design space for EPONs [6], high-speed multichannel next-generation PONs [7], and TDM-PON based 5G mobile fronthaul [8]. It is important to note, however, that the vast majority of these studies focused on the average delay performance of PONs, but not many studied the delay percentiles and delay guarantees. Providing stochastic guarantees for different delay tolerances is highly demanded to realise the Tactile Internet and other emerging low-latency 5G services, whose delay tolerance requirements may dynamically change due to varying environment dynamics (see Fig. 1). Clearly, dynamically acquiring knowledge of delay percentiles in PON-based infrastructures would be helpful to network operators and service providers for specifying delay guarantees.

Recently, the rise of Artificial Intelligence (AI) and Machine Learning (ML) techniques and algorithms have appeared in the networking arena, offering applicability in a wide range of scenarios and promising to reduce the complexity of network operations and management. AI/ML together with Software Defined Networking (SDN) open a wide range of possibilities towards the so-called zero-touch network operations [9, 10]. An extensive overview of some applications of ML/AI techniques in the context of computer and communication networks is provided in [11–14]. Examples of applications include network traffic prediction for resource allocation [15], quality of transmission (QoT) prediction of optical lightpaths [16] or solving the classical Routing and Wavelength Assignment (RWA) using a trained ML model from past RWA configuration data [17]. Nevertheless, in all cases, ML models need to learn from data, realistic data if possible, or simulated/synthetic data, on attempts to capture the patterns and provide data-based models of different systems [18, 19].

This work leverages non-linear regression ML techniques for modeling PON delays and provide simple yet accurate models for network designers and practitioners willing to evaluate the suitability of PONs at covering 5G transport scenarios and Tactile Internet applications. The paper shows how to build mathematical models of packet delay in PONs using data samples, either real measurements or synthetic/simulation datasets. This methodology can be applied to packet delay data obtained from different DBAs under multiple traffic models, including Poisson and Self-Similar. Accurate characterisation of high-delay percentiles allow to dimension PONs with delay guarantees, as shown in this article.

The remaining of this work is organised as follows. Section 2 reviews current state of the art regarding TDM PON delay models. Section 3 provides a detailed overview of ML based Non-Linear Regression techniques, its theoretical foundations and the way to validate their models. Section 4 evaluates the ML models with simulation experiments in a number of scenarios. Finally, Section 6 concludes this work with a summary of its main findings.

2. PRELIMINARIES: THEORETICAL DELAY ANALYSIS IN PONS

In TDM-based Passive Optical Networks, the OLT arbitrates access to the shared media (the feeder fiber) to avoid collisions if multiple ONUs transmit simultaneously, especially in the upstream channel. Thus, the OLT decides when and for how long each ONU transmits in the upstream direction by sending grants to the ONUs, typically arranging them in a round-robin fashion following some Dynamic Bandwidth Allocation (DBA) algorithm (see IPACT [5] and gGIANT [20] as typical examples). The DBA employed at the OLT plays a critical role concerning a fair bandwidth sharing among the users, but also comprises an important source of the total delay in the upstream direction.

There are a large number of articles focused on identifying the most important sources of delay in TDM-PONs and their mathematical characterization, see for instance [21–25]. While the propagation delay is a constant value which only depends on the distance between the ONU and the OLT, the upstream delay is highly variable and has been a matter of interest in different contexts. For instance, the authors in [21] derive a general formula for the average delay following the well-known M/G/1 model with vacations. The authors in [24] further extend such characterisation considering the case of traffic bursts and their impact on other ONUs. The upstream delay model is further extended to NG-PON scenarios [23], and applied in Fiber-Wireless networks [22, 26] or PONs for the transport of fronthaul traffic in C-RAN scenarios [27–30].

In summary, the most widely used model for the characterization of the upstream delay in TDM-PONs using a DBA like IPACT follows [22]:

$$E(D) = 2\tau\frac{2-\rho}{1-\rho} + \frac{\rho}{2C(1-\rho)}\left(\frac{\sigma_L^2}{\bar{L}} + \bar{L}\right) + \frac{\bar{L}}{C} \quad \textbf{(1)}$$

where τ is the one-way propagation delay (approximately 5 μs per km of fiber), ρ is the traffic load in the upstream channel, C is the upstream capacity (1 Gb/s for EPON, 10 Gb/s for 10G-EPON) and $\bar{L}$ and σ_L^2 are the average packet length and its variance respectively.

Rearranging terms in eq. 1, we obtain:

$$E(D) = \frac{4\tau + \left(\frac{\frac{\sigma_L^2}{\bar{L}}+\bar{L}}{2C} - 2\tau\right)\rho}{1-\rho} + \frac{\bar{L}}{C} \quad \textbf{(2)}$$

As a numerical example, consider a 1 Gb/s EPON with 16 ONUs offering a total upstream load of $\rho = 0.6$ (i.e. 60%), with OLT-ONU distance of 5 km (i.e. $\tau = 25\ \mu s$). Packet sizes are assumed to follow the classical three modal distribution observed at the Amsterdam Exchange Point [31, 32], namely 7/12 packets are short (40 bytes), 4/12 are medium size (576 bytes), and 1/12 packets is long (1500 bytes), yielding an average packet size of $\bar{L} = 340.33$ Bytes and standard deviation $\sigma_L = 428$ Bytes. Thus, the average upstream delay follows:

$$E(D) = \frac{4\cdot 25 + (3.5 - 2\cdot 25)\cdot 0.6}{1-0.6} + 2.7 = 182.9\ \mu s \quad \textbf{(3)}$$

The same scenario at 10 Gb/s (10G-EPON) provides an average packet delay in the upstream channel of:

$$E(D) = \frac{4\cdot 25 + (0.35 - 2\cdot 25)\cdot 0.6}{1-0.6} + 0.27 = 175.8\ \mu s \quad \textbf{(4)}$$

which is a very similar result to the average delay at 1 Gb/s. In fact, if the values associated with packet sizes are neglected, the resulting average delay follows:

$$E(D) = \frac{4\cdot 25 - 2\cdot 25\cdot 0.6}{1-0.6} = 175\ \mu s \quad \textbf{(5)}$$

where, as shown, the average delay depends on its vast majority on the propagation delay τ and traffic load ρ; hence it can be approximated as:

$$E(D) \approx 2\tau\frac{2-\rho}{1-\rho} \quad \textbf{(6)}$$

Concerning propagation delays, it is worth noting that, according to [33, 34], 75% of the antenna base stations in the Rennes area of France are located within 4 km of distance to the Central Office (i.e. 20 μs prop. delay), while 95% are within 8 km to the CO (40 μs). Similar distance statistics for home users are expected in dense urban scenarios. These ONU-OLT distance numbers shall be used in the simulation experiments.

The above models for delay in eq. 2 only take into account the average packet delay $E(D)$. However, other important delay-based metrics such as Packet Delay Variation (PDV) or high delay percentile values (90th or 99th percentiles for instance) may be necessary in the analysis of certain networking scenarios, for instance, in dimensioning C-RAN based fronthaul scenarios for 5G [35–38]. In other words, designers may require to dimension the network to guarantee that a large percentage of packets, say 99% of them, satisfy certain delay requirements (i.e. the 99-th delay percentile is below some value). In general, obtaining closed-form equations for delay percentiles under traffic conditions different than Poisson can be a difficult task. For this purpose, this article shows how to find simple model equations from simulated data leveraging non-linear parametric regression techniques, a subset of Machine Learning based methods.

3. NON-LINEAR REGRESSION LEVERAGING MACHINE LEARNING ALGORITHMS

A. Regression models

Regression methods are a type of Machine-Learning technique which deal with the ability to fit a function $f_\beta(\mathbf{x})$ characterized by some parameters β to a set of N observed data points $\{(\mathbf{x_i}, y_i)\}_{i=1}^N$. Here, each sample pair $(\mathbf{x_i}, y_i)$ represents a vector $\mathbf{x}$ of independent variables (or features), $\mathbf{y}$ is the dependent variable (continuously valued) and β are the coefficients or model parameters that characterize the model function.

The goal of the ML technique comprises to find the best set of parameters β which minimize some error function computed between the predicted value $\hat{y}_i = f_\beta(\mathbf{x_i})$ for each input $\mathbf{x_i}$ and the actual true y_i value. In other words, starting with the hypothesis of:

$$y \sim f_\beta(\mathbf{x}) \tag{7}$$

The goal is to find the best set of parameters β that minimizes some error or penalty function of the residuals r_i:

$$r_i = y_i - f_\beta(x_i) \tag{8}$$

The error function aims at measuring and penalizing the difference between the model fit and the real value. Typical error functions include norms L1 or Euclidean L2 distance, Hamming distance, and Manhattan distance.

For instance, in the case of the L2 penalty function, the model parameters β are obtained by minimizing:

$$\text{Find } \beta \text{ such that } \min_\beta \frac{1}{N}\sum_{i=1}^{N}\left(f_{\hat{\beta}}(\mathbf{x_i}) - y_i\right)^2 \tag{9}$$

using Stochastic Gradient Descend (SGD) or any other method.

If the function to be fitted is linear, i.e. $f_\beta(\mathbf{x}) = \beta_0 + \beta_1 x + \ldots$, then there is closed-form formula for coefficients β that minimize L2 distance, namely:

$$\beta = (X^T X)^{-1} X^T y \tag{10}$$

which is the classical linear regression formula where matrix X and vector y collect all vectors $\mathbf{x_i}$ and labels y_i of the dataset [39, 40].

However, in many problems a linear function does not well capture the dynamics and patterns in the observed data (i.e. it "underfits" the data), thus needing the use of nonlinear functions. In this case, the non-linear least squares (NLS) technique offers a means to fit non-linear functions to a dataset [39, 40].

NLS behaves similarly to Linear Least Squares since it first approximates the model to a linear function and then it successively refines the parameters by iterating over previous model estimates (see [41] for further details).

In more detail, function $f(\mathbf{x}, \beta)$ is approximated by a linear function using a first-order Taylor series:

$$f(\mathbf{x_i}, \beta) \approx f(\mathbf{x_i}, 0) + \sum_j J_{ij}\beta_j \tag{11}$$

where:

$$J_{ij} = \frac{\partial f(\mathbf{x_i}, \beta)}{\partial \beta_j} \tag{12}$$

Using linear least squares estimators, the new closed-formula for the coefficients are:

$$\beta \approx (J^T J)^{-1} J^T y \tag{13}$$

that is, the same formula as eq. 10, but using matrix J instead of matrix X.

B. NLS for TDM-PON delays

Two following two non-linear functions will used as a starting hypothesis for TDM-PON delays:

$$D = \beta \cdot 2\tau \frac{2-\rho}{1-\rho} \quad \text{(Simple model)} \tag{14}$$

$$D = 2\tau \frac{\beta_1 - \beta_2 \rho}{\beta_3 - \rho} \quad \text{(Full model)} \tag{15}$$

which follow the same structure as Eq. 6, as a function of the network load ρ and propagation delay τ.

Clearly, the first formula (simple model) only requires one parameter to be estimated by the NLS method, while the second model (full model) requires three parameters. In the simple model, the value of β should be equal to one for the average delay. In the full model case, these values should be $\beta_1 = 2$, $\beta_2 = -1$ and $\beta_3 = 1$ for the average delays.

In section 4, the model parameters (β values) for both average delay values and delay percentiles are fit using NLS. As it will be shown, the full model is more versatile at fitting delay curves than the simple one, showing better accuracy values.

C. Model Asessment and Validation

Regression methods are often validated using both the Root Mean Square Error (RMSE) and the coefficient of determination R^2, which follow:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - f_\beta(x_i))^2} \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - f_{\hat{\beta}}(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{17}$$

where $\bar{y} = \frac{1}{N}\sum_{i_1}^{N} y_i$. R^2 generally ranges between 0 and 1 and is often given as a percentage, where 100% represents a perfect fit. RMSE takes the same units as the datapoints y_i (in our case, seconds).

4. SIMULATION FRAMEWORK

In the simulations, we have considered 1- and 10-Gbps EPON comprising eight ONUs. We have examined EPONs with different end-to-end ONU-OLT distances being set to 4, 8, and 20 km. The guard band is set to 128 ns [42]. To arbitrate the channel access in the upstream, we have implemented the well-known interleaved polling with adaptive cycle time (IPACT) protocol.

A. Experiment 1: Poisson traffic

In the first experiment, the incoming traffic follows a Poisson arrival model with various traffic intensities (load varying between 0.2 and 0.99), while packet sizes follow the classical trimodal packet distribution observed at the Amsterdam Internet Exchange (AMS-IX) [27, 32]; namely packet sizes of 40, 576, and 1500 Bytes with percentages of 7/12, 4/12, and 1/12, respectively.

In total, the experiments include six datasets, three at 1 Gb/s (4km, 8km and 20 km), and another three at 10 Gb/s (again 4km, 8km and 20 km). The former three datasets include 10,000 packet delay measurements per dataset, while the later three contain 50,000 packet delay measurements per dataset. We note that these numbers were deemed sufficient to reach the desired statistical stability.

With such individual packet delay values, we construct the ML models both for the average delay and also for different delay percentiles, as explained in the next sections. Delay percentiles are useful when the goal is to dimension PON scenarios where certain worse-case delay guarantees are provided. For instance, the 90th delay percentile represents the delay value experienced by the top-highest 10% delay.

Both datasets and R code are available at Github for the reader to replicate the experiments[1].

A.1. Average Delay: Theoretical vs NLS

Figs. 2 shows the two ML model fit for the average queueing delay under different values of load ρ, both for 1G and 10G EPONs. The figures show both simulation delay observations, along with the fitted model and the theoretical average delay of Eq. 6. As shown, both ML models mimic the average delay behaviour as a function of the traffic load ρ.

The parameters of the two ML models have been fit from the simulation dataset using library *nls* from open-source R programming language, and following the methodology of [43]. The values of the fit are shown in table 1 along with the coefficient of determination R^2 that evaluates the goodness of fit for each model, where the whole dataset has been used to construct the ML models. For completeness, also the RMSE values between the dataset and model's prediction is shown.

Both ML models provide good generalization capabilities (as visually observed in the figures) since they capture the curve pattern and dynamics for the 12 different load values depicted. A Leave One Out Cross-Validation (LOOCV) strategy to verify this would require to train the model with 11 load values and testing on the remaining one, repeating the whole process for the 12 datapoints and averaging the 12 R^2 and $RMSE$ results. The values shown in the table have been obtained using the whole dataset in the curve fitting process.

As shown, the so-called Full ML model is more sophisticated; it comprises three model parameters and achieves better performance in terms of RMSE and R^2, in-line with the visual fit of Fig. 2. On the other hand, the Simple ML model, with only one

[1] see https://github.com/josetilos/nls_for_EPONs/, last access 25th May 2020.

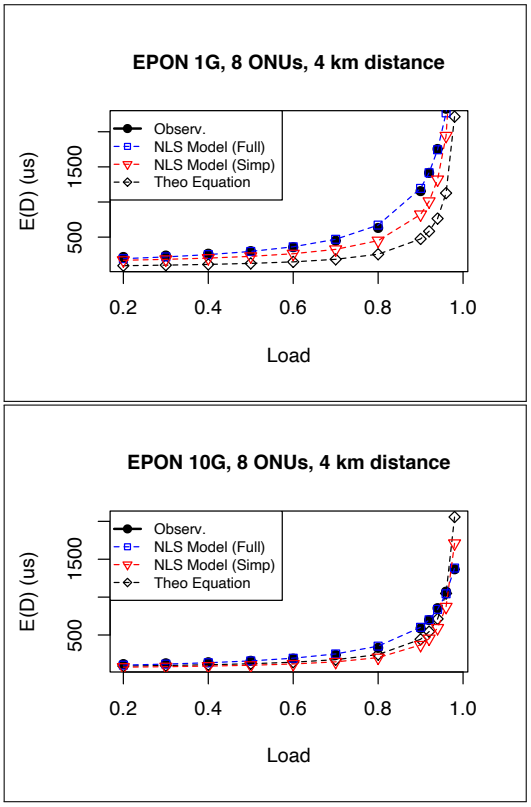

Fig. 2. Average delay fit: Simulation, theoretical equation and ML regression models (full and simplified) for 1G (top) and 10G (bottom) EPONs.

β parameter, is also highly accurate $R^2 \approx 95\%$ and provides a simpler way to model delays and delay percentiles, as shown in the next section.

A.2. Models for Delay Percentiles

Concerning delay percentiles, it is worth remarking that there is no theoretical baseline equation to compare with. Figs. 3 show the simulation results and the ML models in four cases: 25th, 50th, 75th and 90th delay percentiles for a 1G EPON. As shown, the Full ML model is more flexible and achieves better accuracy to the delay curve as a function of the network load than the Simple ML model.

In addition, Table 2 collects the ML model parameters and goodness of fit for most useful delay percentiles at designing PON scenarios with delay guarantees, namely 25, 50, 75, 90, 95 and 99th percentiles. The table includes the ML models for both 1G and 10G EPONs. These models will be put into practise in section 5.

B. Experiment 2: Self-similar traffic

In the second experiment, we have simulated self-similar traffic using the Chaotic-map model described in Section II.A of [44] for the 1G PON. Following this, let the chaotic map be defined by the function $f = f(m_1, m_2, d)$ in the unit interval $I = (0, 1)$ as $x_{n+1} = f(x_n)$:

$$x_{n+1} = \begin{cases} x_n + (1-d)\left(\frac{x_n}{d}\right)^{m_1}, & 0 < x_n \leq d \\ x_n - d\left(\frac{1-x_n}{1-d}\right)^{m_1}, & d < x_n < 1 \end{cases} \quad \textbf{(18)}$$

Table 1. ML model fit (simple and full) along with theoretical equation for a 1 Gb/s EPON with 32 ONUs and 4 km distance between OLT and ONU.

Model	Equation $E(D)$	R^2	$RMSE$
Theo. Eq. 2	$2\tau\frac{4-2\rho}{1-\rho} + 2.7\ \mu s$	0.952	151.5 μs
ML Model (Simple)	$1.07 \cdot 2\tau\frac{2-\rho}{1-\rho} + 2.7\ \mu s$	0.952	148.0 μs
ML Model (FULL)	$2\tau\frac{4.45-0.51\rho}{1.02-\rho} + 2.7\ \mu s$	0.999	13.6 μs

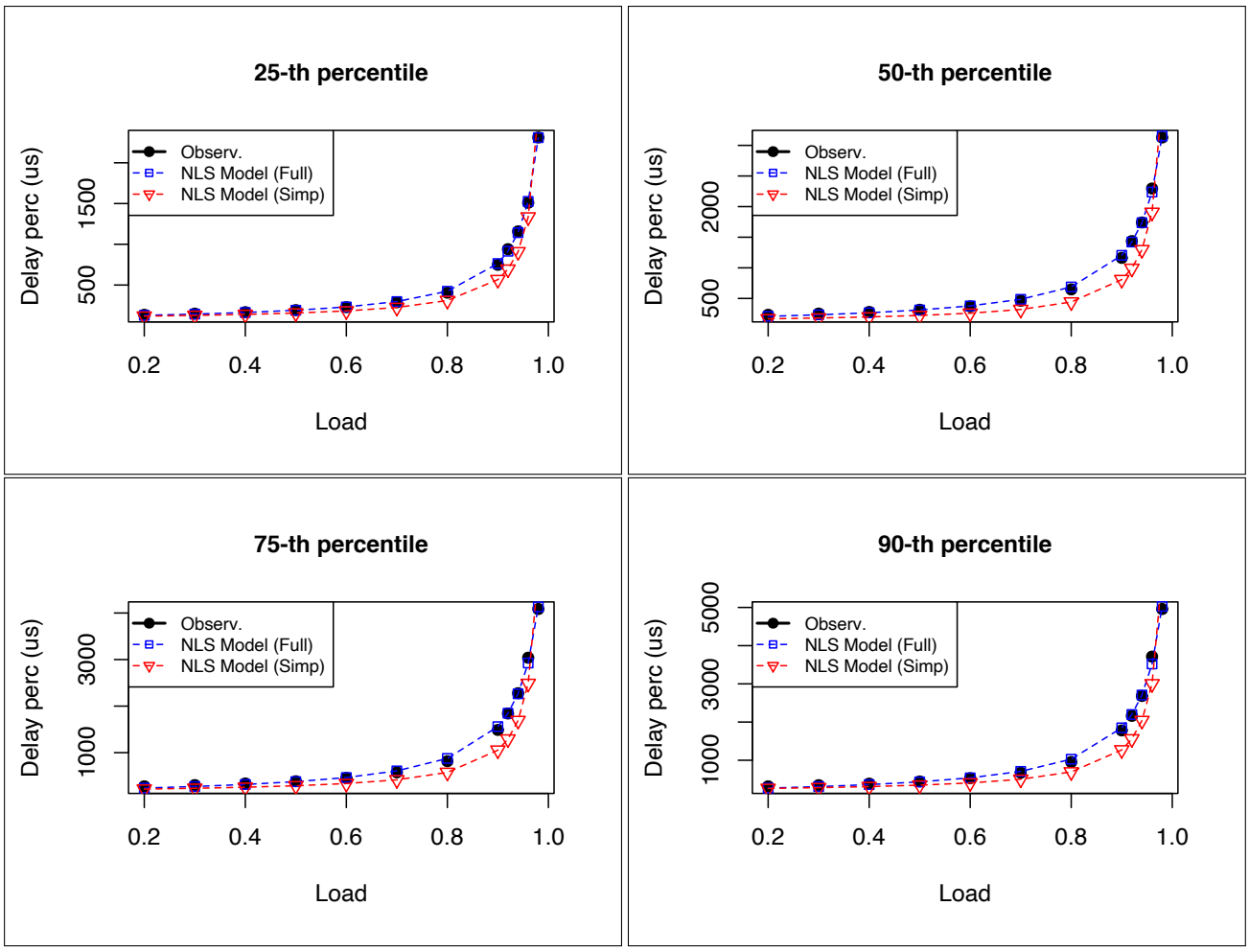

Fig. 3. ML model fit for delay percentiles (Full and Simple models)

Table 2. ML model fit (simple and full) for different delay percentiles, 1G and 10G EPONs (Poisson traffic)

Model	1G EPON Equation	R^2	10G EPON Equation	R^2
ML Model (Simple)	$D_{.25} = 0.8 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.92	$D_{.25} = 0.633 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.823
ML Model (FULL)	$D_{.25} = 2\tau\frac{3.9-0.36\rho}{1.03-\rho} + 2.7\ (\mu s)$	0.999	$D_{.25} = 2\tau\frac{3.46+0.97\rho}{1.07-\rho} + 2.7\ (\mu s)$	0.995
ML Model (Simple)	$D_{.50} = 1.04 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.95	$D_{.50} = 0.83 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.887
ML Model (FULL)	$D_{.50} = 2\tau\frac{4.5-0.66\rho}{1.02-\rho} + 2.7\ (\mu s)$	0.999	$D_{.50} = 2\tau\frac{4.13+0.24\rho}{1.04-\rho} + 2.7\ (\mu s)$	0.998
ML Model (Simple)	$D_{.75} = 1.31 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.969	$D_{.75} = 1.03 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.917
ML Model (FULL)	$D_{.75} = 2\tau\frac{4.97-0.65\rho}{1.016-\rho} + 2.7\ (\mu s)$	0.999	$D_{.75} = 2\tau\frac{4.58+0.12\rho}{1.03-\rho} + 2.7\ (\mu s)$	0.999
ML Model (Simple)	$D_{.90} = 1.60 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.972	$D_{.90} = 1.22 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.940
ML Model (FULL)	$D_{.90} = 2\tau\frac{4.87+0.17\rho}{1.014-\rho} + 2.7\ (\mu s)$	0.997	$D_{.90} = 2\tau\frac{4.85+0.01\rho}{1.02-\rho} + 2.7\ (\mu s)$	0.999
ML Model (Simple)	$D_{.95} = 1.74 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.966	$D_{.95} = 1.31 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.95
ML Model (FULL)	$D_{.95} = 2\tau\frac{4.72+0.94\rho}{1.015-\rho} + 2.7\ (\mu s)$	0.996	$D_{.95} = 2\tau\frac{4.91+0.05\rho}{1.02-\rho} + 2.7\ (\mu s)$	0.999
ML Model (Simple)	$D_{.99} = 1.98 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.95	$D_{.99} = 1.45 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.964
ML Model (FULL)	$D_{.99} = 2\tau\frac{4.55+2.40\rho}{1.018-\rho} + 2.7\ (\mu s)$	0.990	$D_{.99} = 2\tau\frac{5.3-0.32\rho}{1.02-\rho} + 2.7\ (\mu s)$	0.9995

where $d \in (0,1)$ and the initial state is x_0 for $n = 0$. Parameter d is a decision threshold in the data generation or non-generation (i.e. ON/OFF). In this sense, data is transmitted (ON) if x_n satisfies $d < x_n < 1$, that is:

$$y_n = \begin{cases} 0, & 0 < x_n \leq d \quad (OFF) \\ 1, & d < x_n < 1 \quad (ON) \end{cases} \tag{19}$$

In the simulations, the following parameters have been used: $m_1 = m_2 = 1$ and starting value $x_0 = 0.1$ (i.e. starting in OFF state); multiple ON/OFF periods are generated with discrete time resolution of 0.5 μs, that is, each ON/OFF period lasts for 0.5 μs, giving that, the smallest possible packet at 1 Gb/s is $0.5 \cdot 10^{-6} \times 10^9 = 500$ bits. The simulation shows highly-variable ON periods (either very long or very short, like in a Pareto ON source) alternating with highly-variable OFF periods (again either very-short or very-long). Parameter d is adjusted to the load ρ in order to simulate different load scenarios, as required to build the experiments. Figs. 4 show the Auto-Correlation Function (ACF) of both LRD traffic and Poisson traffic with respect to the lag (0.5 μs) to verify the LRD nature of the former (chaotic-map traffic model) and SRD nature of the latter (Poisson traffic model). As shown, the ACF for the chaotic-map model decays very slowly (slower than exponential) with time (each lag-unit is 0.5 μs) proving the LRD nature of the chaotic-map model simulated traffic. Given the very-long time required to generate representative traffic traces, only 1G EPON simulations are included in this study.

Table 3 shows the models and goodness of fit for both simple and full models when the ONUs offer self-similar traffic to the OLT. As shown, both simple and full ML models show similar performance results on various percentile models, in the range of 0.65 to 0.8 R^2. This is a consequence of the extremely highly-variability of the traffic, which shows variable delay percentile values for similar network loads, as observed in Fig. 5.

As an example, Fig. 5 shows both simple and full ML models on the LRD trace generated using the chaotic map model explained above, for estimating the 90-th delay percentile. As

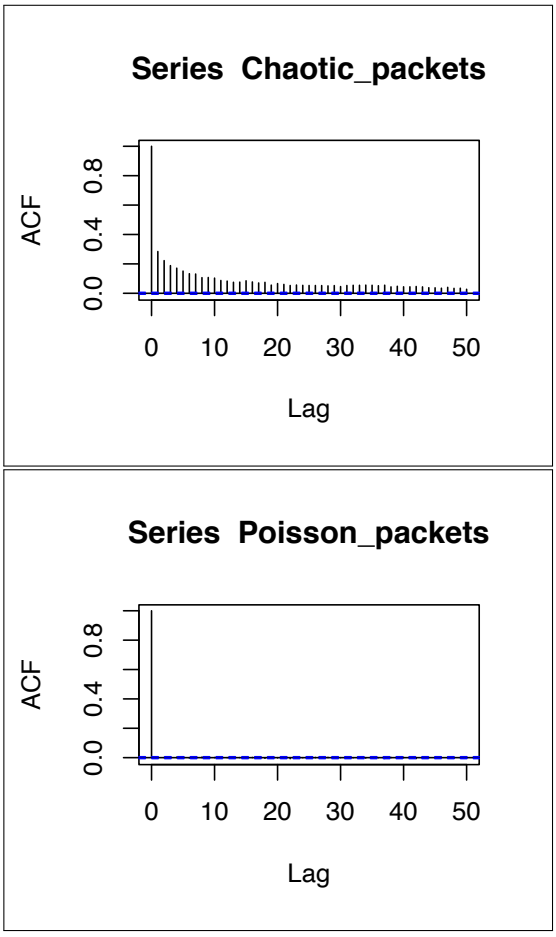

Fig. 4. Auto-Correlation Function of LRD traffic generated with the chaotic map model (top) vs Poisson traffic (bottom).

Table 3. ML model fit (simple and full) for different delay percentiles, 1G EPON (Self-similar traffic)

Model	1G EPON Equation	R^2
ML Model (Simple)	$D_{.25} = 6.7 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.76
ML Model (FULL)	$D_{.25} = 2\tau\frac{-1.8+12.5\rho}{0.95-\rho} + 2.7\ (\mu s)$	0.82
ML Model (Simple)	$D_{.50} = 12.0 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.75
ML Model (FULL)	$D_{.50} = 2\tau\frac{-0.23+21.8\rho}{0.96-\rho} + 2.7\ (\mu s)$	0.78
ML Model (Simple)	$D_{.75} = 21.4 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.66
ML Model (FULL)	$D_{.75} = 2\tau\frac{7.31+32.0\rho}{0.97-\rho} + 2.7\ (\mu s)$	0.68
ML Model (Simple)	$D_{.90} = 28.5 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.71
ML Model (FULL)	$D_{.90} = 2\tau\frac{27.6+25.0\rho}{0.97-\rho} + 2.7\ (\mu s)$	0.72
ML Model (Simple)	$D_{.95} = 31.6 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.75
ML Model (FULL)	$D_{.95} = 2\tau\frac{41.5+17.8\rho}{0.97-\rho} + 2.7\ (\mu s)$	0.76
ML Model (Simple)	$D_{.99} = 37.4 \cdot 2\tau\frac{2-\rho}{1-\rho}\ (\mu s)$	0.74
ML Model (FULL)	$D_{.99} = 2\tau\frac{77.3-19.9\rho}{0.97-\rho} + 2.7\ (\mu s)$	0.76

shown, both models capture the dynamics of delay percentile curve, being the full model slightly better in terms of R^2.

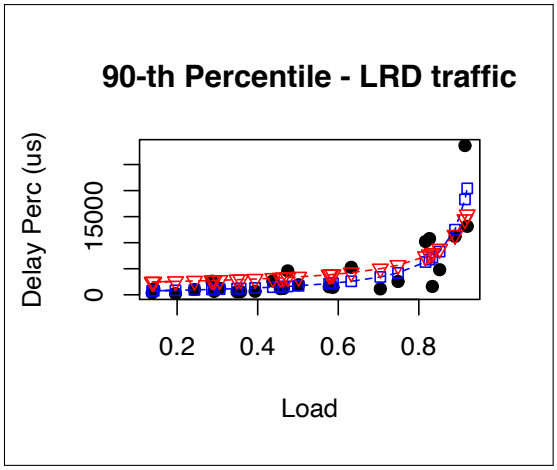

Fig. 5. Simple and Full ML models for LRD traffic, 90-th percentile.

Essentially, if after conducting the NLS regression models for a given scenario with its particular topology and traffic sources, the model scores an R^2 statistic above 70%, it is generally accepted that the model captures most of the variability of the data and fits well. This is the case of the experiments conducted for both Poisson and LRD traffic.

5. USES AND APPLICATIONS OF THE ML-BASED DELAY MODELS

This section focuses on the applicability of the ML models in dimensioning PONs with delay guarantees in different network scenarios.

A. Tactile Internet and Low-Latency Services for Residential Users

Consider a PON located in a dense urban scenario, providing broadband access to 32 residential users and maximum ONU-OLT distance of 4 km (i.e. $\tau = 20\ \mu s$). In this scenario, the network operator is willing to guarantee that a majority of packets (90% of them) should not exceed a given delay threshold, say $D_{thres} = 1\ ms$, on attempts to support certain Tactile Internet services. Following the Simple ML model for Poisson traffic, this goal translates into finding the the maximum traffic load ρ at the upstream channel of the EPON which guarantees $D_{.90} \leq D_{thres}$.

Under the assumption of Poisson traffic, the 90th delay percentile using the Simple ML model in a 1G EPON scenario follows $\beta \cdot 2\tau\frac{2-\rho}{1-rho}$ with $\beta = 1.6$ (see Table 2). Thus, the maximum traffic load at the EPON is:

$$\rho \leq \frac{\frac{D_{thres}}{2\tau} - 2\beta}{\frac{D_{thres}}{2\tau} - \beta} = 0.93 \qquad \textbf{(20)}$$

In other words, the network operator must ensure that the PON utilization does not exceed $\rho_{max} = 0.93$ or 93% load maximum, contributed by all residential ONUs. Thus, a 1 Gb/s EPON should be loaded with no more than 930 Mb/s. However, if the EPON scenario comprises a distance of 8 km between ONU and OLT (i.e. $\tau = 40\ \mu s$), then the maximum load allowed under the same conditions is $\rho \leq 0.85$. For 20 km distance PONs, this maximum traffic load reduces to $\rho \leq 0.53$.

Now, in the case of a symmetric 10G-EPON, the Simple ML model uses $\beta = 1.22$ (see again Table 2), thus the maximum load for a PON with ONU-OLT distance of 4 km is $\rho_{max} = 0.95$ or 9.5 Gb/s at most. The maximum loads at 8 and 20 km are 0.89 and 0.68 respectively (i.e. 8.9 and 6.8 Gb/s).

As shown in the example, providing delay guarantees decreases the amount of network traffic that the EPON can allocate, since very highly-loaded upstream channels will translate into high queuing delays. This is particularly harmful at 1 Gb/s and long-distance EPONs. However, increasing the capacity to 10 Gb/s allows a lot more bandwidth efficiency (from 0.53 to

0.68 for a 20-km PON) and makes the PON more insensitive to long-distance ONUs.

Finally, it is worth remarking that current PON deployments are highly underloaded in most residential cities, showing peak traffic values of tens of Mb/s per household at the time of writing (see [45, 46] for further details).

Nevertheless, the IEEE 802.3ca standard for 25G and 50G EPONs has been approved in 2020; and 25G/50G EPON products should be commercially available and massively deployed in the near future.

B. Fronthaul Transport in C-RAN Scenarios

Fronthaul traffic for the deployment of C-RAN based 5G scenarios typically demands high-bandwidth capacity networks and ultra-low latency connectivity, especially for low-level functional splits like CPRI [47].

In this use case, we first consider an EPON providing support to three Remote Radio Heads covering three sectors of a macro cell station, and implementing the Intra-PHY I_{UI} functional split of eCPRI for a 20 MHz radio channel, as specified in 5G New Radio (NR). In this setting, the eCPRI bandwidth required per antenna is 921.6 Mb/s (assuming no MIMO), see [35, 36] for further details.

Again, we consider a scenario where the antennas are 4 km apart from the OLT, and the maximum latency allowed is $D_{thres} = 250\ \mu s$. In a 1G EPON setting, the maximum load allowed to guarantee that 99% of the packet delays fall below D_{thres} is:

$$\rho \leq \frac{\frac{D_{thres}}{2\tau} - 2\beta}{\frac{D_{thres}}{2\tau} - \beta} \tag{21}$$

with $\beta = 1.98$, which yields a maximum load of $\rho \leq 0.91$ or 910 Mb/s, clearly insufficient for carrying such three fronthaul flows (3x0.91 Gb/s). However, the $D_{.99}$ model for a 10G-EPON uses $\beta = 1.45$, hence the maximum load is 0.93. This implies a maximum traffic load of 9.3 Gb/s maximum, which is far below $3 \times 910\ Mb/s$, hence meeting the delay requirements in the upstream channel.

Furthermore, if we compute the $D_{.99}$ delay percentile for a 1:16 split 10G-EPON with say, three fronthaul flows at 910 Mb/s each and 13 residential users with a peak rate of 100 Mb/s each (i.e. total load of $\rho = \frac{3\times 0.91 + 13\times 0.1\ Gb/s}{10\ Gb/s} = 0.403$), then the resulting average packet delay for all users is:

$$E(D) = 2 \cdot 20 \frac{2 - 0.4}{1 - 0.4} = 107\ \mu s \tag{22}$$

and the 99th delay percentile becomes:

$$D_{.99} = 1.45 \cdot 2 \cdot 20 \frac{2 - 0.4}{1 - 0.4} = 155\ \mu s \tag{23}$$

which is below the 250 μs limit. However, when the ONU-OLT distance doubles (i.e. 8 km), then the 250 μs limit is not guaranteed for 99% of the packets.

It is finally worth noting that the model assumptions require that traffic patterns are Poisson, which is typically not true for some functional splits, especially low-level ones, which typically show a Constant Bit Rate (CBR) traffic pattern. Nevertheless, this is an illustrative example of PON dimensioning of strict delay percentile guarantees, that can be used as a first approach toward PON dimensioning in 5G contexts.

C. Impact of LRD traffic in PON dimensioning

To see the impact of highly-variable self-similar traffic in dimensioning PONs with delay guarantees, consider a residential scenario with 32 households and maximum ONU-OLT distance of 4 km (i.e. $\tau = 20\ \mu s$). Let us further assume that the network operator is willing to guarantee a maximum delay of 5 ms for 99% of the packets.

In this case, applying the eq. 20 to the $D_{.99}$ delay model for Poisson traffic gives a maximum load of 98.4% (here $\beta \approx 2$ only). However, if we consider the delay model under self-similar traffic, the maximum load becomes only 57% in eq. 20, since the value of β is substantially larger ($\beta = 37.4$).

This example shows the impact of LRD traffic variability, captured by the ML model as a very high value of the β parameter, 37.4 for LRD vs 2 for Poisson traffic. This highlights the importance of collecting real data measurements by the network operator to build their own ML models and estimates of β parameters for their planning strategies.

6. SUMMARY AND DISCUSSION

In this article, we show the applicability of using Machine-Learning based non-linear regression techniques in modeling the upstream delay of Ethernet-based TDM PONs. Exact equations for the average delay experienced by packets already exist in the literature, however equations for delay percentiles do not. Delay percentiles are of particular importance in the planning and design of PONs for the support of ultra-low latency applications and services, including the Tactile Internet.

To this end we show the power of non-linear regression techniques to build simple but accurate models for characterizing delay percentiles in PONs, based on both Poisson and LRD simulation data and ML techniques. These models along with the open-source R code available in Github are directly applicable for network designers willing to have a first estimate of delay percentiles in their scenarios. However, the network designer will ultimately have to conduct his/her own experiments with real measurements to verify whether or not their PON settings are suitable in providing certain tight delay guarantees.

In this sense, the application of the ML model in real scenarios require the network operator to collect real-time data measurements and construct their own ML models with accurate estimates of the β parameters, since their value highly depend on the distance ONU-OLT and the underlying traffic model and its variability. In the case of operators using other DBAs including service differentiation, the same methodology may apply, but a model needs to be built (its parameters estimated) per traffic class, since all classes will probably have a similar delay vs load shape as those observed throughout the paper. The accuracy of the models will be determined by their R^2 metric.

7. SUPPLEMENTAL MATERIAL

Both datasets and R code are available at Github for the reader to replicate the experiments[2].

FUNDING

Ministerio de Ciencia, Innovación y Universidades (ACHILLES, PID2019-104207RB-I00); EU H2020 (PASSION project, grant no. 780326).

[2]see https://github.com/josetilos/nls_for_EPONs/, last access 25th May 2020.

ACKNOWLEDGMENTS

The authors would also like to thank Prof. Elaine Wong from University of Melbourne for her valuable comments and feedback on the paper.

REFERENCES

1. T. Pfeiffer, "Can pon technologies accelerate 5g deployments?" in *Conference on Optical Network Design and Modelling (ONDM), Workshop on Optical Technologies in the 5G Era,* (2018).
2. C. Behrens, S. Krauss, E. Weis, and D. Breuer, "Technologies for convergence of fixed and mobile access: An operator's perspective [invited]," IEEE/OSA J. Opt. Commun. Netw. **10**, A37–A42 (2018).
3. K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, and M. Dohler, "Toward haptic communications over the 5g tactile internet," IEEE Commun. Surv. & Tutorials **20**, 3034–3059 (2018).
4. S. Mondal, L. Ruan, and E. Wong, "Remote human-to-machine distance emulation through ai-enhanced servers for tactile internet applications," in *2020 Optical Fiber Communications Conference and Exhibition (OFC),* (2020), pp. 1–3.
5. G. Kramer, B. Mukherjee, and G. Pesavento, "Ipact: a dynamic protocol for an ethernet pon (epon)," IEEE Commun. Mag. **40**, 74–80 (2002).
6. M. P. McGarry and M. Reisslein, "Investigation of the dba algorithm design space for epons," IEEE/OSA J. Light. Technol. **30**, 2271–2280 (2012).
7. F. Aurzada, M. Scheutzow, M. Reisslein, N. Ghazisaidi, and M. Maier, "Capacity and delay analysis of next-generation passive optical networks (ng-pons)," IEEE Transactions on Commun. **59**, 1378–1388 (2011).
8. A. D. Hossain and A. R. Hossain, "A distributed control framework for tdm-pon based 5g mobile fronthaul," IEEE Access **7**, 162102–162114 (2019).
9. D. Rafique and L. Velasco, "Machine learning for network automation: overview, architecture, and applications [invited tutorial]," IEEE/OSA J. Opt. Commun. Netw. **10**, D126–D143 (2018).
10. T. Tanaka, A. Hirano, S. Kobayashi, T. Oda, S. Kuwabara, A. Lord, P. Gunning, O. Gonzalez de Dios, V. Lopez, A. M. Lopez de Lerma, and A. Manzalini, "Autonomous network diagnosis from the carrier perspective [invited]," IEEE/OSA J. Opt. Commun. Netw. **12**, A9–A17 (2020).
11. F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "A survey on application of machine learning techniques in optical networks," arXiv preprint arXiv:1803.07976 pp. 1–21 (2018).
12. R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shariar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," J. Internet Serv. Appl. **9**, 1–99 (2018).
13. J. Mata, I. de Miguel, R. J. Duran, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (ai) methods in optical networks: A comprehensive survey," Opt. Switch. Netw. **28**, 43–57 (2018).
14. L. Ruan, M. P. I. Dias, and E. Wong, "Enhancing latency performance through intelligent bandwidth allocation decisions: a survey and comparative study of machine learning techniques," IEEE/OSA J. Opt. Commun. Netw. **12**, B20–B32 (2020).
15. A. Azzouni and G. Pujolle, "A long short-term memory recurrent neural network framework for network traffic matrix prediction," CoRR. **abs/1705.05690** (2017).
16. R. M. Morais and J. ao Pedro, "Machine Learning Models for Estimating Quality of Transmission in DWDM Networks," J. Opt. Commun. Netw. **10**, D84–D99 (2018).
17. I. Martin, S. Troia, J. A. Hernandez, A. Rodriguez, F. Musumeci, G. Maier, R. Alvizu, and O. Gonzalez de Dios, "Machine Learning-Based Routing and Wavelength Assignment in Software-Defined Optical Networks," IEEE Transactions on Netw. Serv. Manag. **16**, 871–883 (2019).
18. D. Côté, "Using machine learning in communication networks (invited)," J. Opt. Commun. Netw. **10**, D100–D109 (2018).
19. J. A. Hernández, "Learning from data: applications of machine learning in optical network design and modeling," in *Proc. of Optical Network Design and Modeling (ONDM),* (2020).
20. R. A. Butt, S. M. Idrus, K. N. Qureshi, N. Zulkifli, and S. H. Mohammad, "Improved dynamic bandwidth allocation algorithm for xgpon," IEEE/OSA J. Opt. Commun. Netw. **9**, 87–97 (2017).
21. S. Bharati and P. Saengudomlert, "Analysis of mean packet delay for dynamic bandwidth allocation algorithms in epons," J. Light. Technol. **28**, 3454–3462 (2010).
22. F. Aurzada, M. Lévesque, M. Maier, and M. Reisslein, "Fiwi access networks based on next-generation pon and gigabit-class wlan technologies: A capacity and delay analysis," IEEE/ACM Transactions on Netw. **22**, 1176–1189 (2014).
23. F. Aurzada, M. Scheutzow, M. Reisslein, N. Ghazisaidi, and M. Maier, "Capacity and delay analysis of next-generation passive optical networks (ng-pons)," IEEE Transactions on Commun. **59**, 1378–1388 (2011).
24. J. García-Reinoso, J. A. Hernández, I. Seoane, and I. Vidal, "On the effect of sudden data bursts in the upstream channel of ethernet pons employing ipact under the gated-service discipline," Opt. Switch. Netw. **13**, 94–102 (2014).
25. I. Seoane, J. A. Hernández, R. Romeral, and D. Larrabeiti, "Analysis and simulation of a delay-based service differentiation algorithm for ipact-based pons," Photonic Netw. Commun. **24**, 228–236 (2012).
26. H. Beyranvand, M. Lévesque, M. Maier, J. A. Salehi, C. Verikoukis, and D. Tipper, "Toward 5g: Fiwi enhanced lte-a hetnets with reliable low-latency fiber backhaul sharing and wifi offloading," IEEE/ACM Transactions on Netw. **25**, 690–707 (2017).
27. I. Ucar, J. A. Hernández, P. Serrano, and A. Azcorra, "Design and analysis of 5g scenarios with 'simmer': An r package for fast des prototyping," IEEE Commun. Mag. **56**, 145–151 (2018).
28. D. Eugui and J. A. Hernández, "Low-latency transmission of fronthaul traffic over xg (s)-pon with fixed-elastic bandwidth reservations," in *Optical Fiber Communication Conference,* (Optical Society of America, 2019), pp. W2A–28.
29. S. Zhou, X. Liu, F. Effenberger, and J. Chao, "Low-latency high-efficiency mobile fronthaul with tdm-pon (mobile-pon)," J. Opt. Commun. Netw. **10**, A20–A26 (2018).
30. H. Zeng, X. Liu, S. Megeed, A. Shen, and F. Effenberger, "Digital signal processing for high-speed fiber-wireless convergence (invited)," J. Opt. Commun. Netw. **11**, A11–A19 (2019).
31. "Amsterdam Internet exchange Ethernet frame size distribution," https://stats.ams-ix.net/sflow/index.html. Online; April, 2020.
32. A. Gowda, J. A. Hernández, D. Larrabeiti, and L. Kazovsky, "Delay analysis of mixed fronthaul and backhaul traffic under strict priority queueing discipline in a 5g packet transport network," Transactions on Emerg. Telecommun. Technol. **28**, e3168 (2017). E3168 ett.3168.
33. Z. Tayq, "Fronthaul integration and monitoring in 5g networks," Ph.D. thesis, Université de Limoges (2017).
34. P. Chanclou, A. Pizznat, D. Yann, and S. Randazzo, "C-ran architecture and fronthaul challenges," in *RAN world conference, Cloud RAN working group,* (2015).
35. G. Otero Pérez, J. A. Hernández, and D. Larrabeiti, "Fronthaul network modeling and dimensioning meeting ultra-low latency requirements for 5g," IEEE/OSA J. Opt. Commun. Netw. **10**, 573–581 (2018).
36. G. Otero Pérez, D. Larrabeiti, and J. A. Hernández, "5g new radio fronthaul network design for ecpri - ieee 802.1cm and extreme latency percentiles," IEEE Access **7**, 82218–82230 (2019).
37. A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, "Things you should know about fronthaul," J. Light. Technol. **33**, 1077–1083 (2015).
38. D. Eugui and J. A. Hernández, "Analysis of a hybrid fixed-elastic dba with guaranteed fronthaul delay in xg(s)-pons," Comput. Networks **164**, 106907 (2019).
39. D. Bates and D. Watts, *Nonlinear regression analysis and its applications*, Wiley series in probability and mathematical statistics (Wiley, New York [u.a.], 1988).
40. D. M. Bates and J. M. Chambers, *Nonlinear models. Chapter 10 of Statistical Models* (Wadsworth & Brooks/Cole, 1992).
41. T. Strutz, *Data Fitting and Uncertainty: A Practical Introduction to*

Weighted Least Squares and Beyond (Vieweg and Teubner, Wiesbaden, DEU, 2010).
42. S. Gorshe, A. Raghavan, T. Starr, and S. Galli, *Broadband access: wireline and wireless-alternatives for internet services* (John Wiley & Sons, 2014).
43. D. G. Rossiter, *Technical note: Curve fitting with the R Environment for Statistical Computing* (2016).
44. M. Balanici and S. Pachnicke, "Hybrid electro-optical intra-data center networks tailored for different traffic classes," IEEE/OSA J. Opt. Commun. Netw. **10**, 889–901 (2018).
45. J. A. Hernandez, R. Sanchez, I. Martin, and D. Larrabeiti, "Meeting the traffic requirements of residential users in the next decade with current ftth standards: How much? how long?" IEEE Commun. Mag. **57**, 120–125 (2019).
46. R. Sanchez, J. A. Hernández, J. Montalvo García, and D. Larrabeiti, "Provisioning 1 gb/s symmetrical services with next-generation passive optical network technologies," IEEE Commun. Mag. **54**, 72–77 (2016).
47. A. de la Oliva, J. A. Hernández, D. Larrabeiti, and A. Azcorra, "An overview of the cpri specification and its application to c-ran based lte scenarios," IEEE Commun. Mag. **54**, 152–159 (2016).