

Network-wide QoT Estimation Using SGD with Gradient Transfer Between Wavelengths

Kayol S. Mayer¹, Jonathan A. Soares¹, Marcos P. A. Dal Maso²,
Christian E. Rothenberg², Dalton S. Arantes¹, Darli A. A. Mello¹,

⁽¹⁾ DECOM, ⁽²⁾ DCA, Universidade Estadual de Campinas, Av. Albert Einstein 400, Campinas, SP, Brazil.

kayol@unicamp.br

Abstract: We propose an SGD-based QoT estimation technique that operates on a network-wide scale by transferring gradients among neighboring wavelengths. Simulation results indicate effective and low-complexity QoT estimation using only transponder SNR telemetry. © 2024 The Author(s)

OCIS codes: (060.4256) Networks, network optimization. (060.4258) Networks, network topology.

1. Introduction

Estimating the quality of transmission (QoT) of unestablished lightpaths has been an active field of research in recent years [1]. Contrasting with the common practice of estimating QoT based on analytical formulas and applying margins, new approaches leverage advances in network operation to improve QoT estimates based on the telemetry of installed devices. Most recent QoT estimation algorithms, based on machine learning, can be divided into two main classes. Class-one algorithms use black-box-like techniques, such as neural networks, which preclude analytical formulas for channel modeling [2]. Class-two algorithms partially rely on analytical formulas, optimizing parameters by the stochastic gradient descent (SGD) algorithm.

Representing the first class, Ibrahim et al. [3] estimate the generalized signal-to-noise ratio (GSNR) distribution of unestablished lightpaths using regression. Morette et al. [4] and Yang et al. [5] exploit the input refinement method to improve SNR prediction in single- and multi-band systems, respectively. Diaz-Montiel et al. [6] improve QoT accuracy up to 3 dB using Mininet-Optical with optical performance monitoring nodes at periodic locations. Mahajan et al. [7] employ support vector machines (SVMs) to reduce to nearly 1 dB the required GSNR margins. Kruse et al. [8] implement agnostic QoT estimation with long short-term memory (LSTM) neural networks, obtaining 1.1 dB improvement in established lightpaths. Amirabadi et al. [9] compare deep neural network (DNN) regressors with other well-known ML algorithms for GSNR estimation. Ayoub et al. [10] extract insights and inspect misclassifications of extreme gradient boosting for supervised binary classification of bit error rate acceptability. Müller et al. [11] implements a path computation element by using gradient boosting trained with enhanced Gaussian noise (EGN) model data. As a second-class representative, Seve et al. [2] propose a QoT estimation scheme using a gradient descent algorithm trained with analytical data.

This paper proposes a QoT algorithm that fits in between classes one and two. We consider each link between ROADMs as an element that impairs the SNR¹, aggregating the contribution of amplified spontaneous emission (ASE) noise from amplifiers and the nonlinear interference generated in fibers. Unlike the work of Seve et al. [2], in which imperfect noise figure and amplifier output powers are estimated using a gradient descent technique applied to the Gaussian-noise model equations, we solely rely on transponder SNR telemetry to optimize equivalent link SNRs, where a link is an abstraction that interconnects two ROADMs. The main feature of our approach is to accomplish network-wide SNR estimation, in which the gradients calculated for an established lightpath are transferred to neighboring channels. When applied to several source-destination pairs, this approach collectively optimizes the estimation process on a network-wide scale.

2. Proposed Framework

We consider an optical network topology with r ROADMs $R_r = \{1, 2, \dots, r\}$ and links $L \in \{\{a, b\} | a, b \in R_r^2 \text{ and } a \neq b\}$ interconnecting the ROADMs. The framework is based on an optical network digital twin (NDT_{ON}) [12, 13], which maintains a digital database of lightpath SNR estimates. Eventually, the lightpath SNRs recorded in the NDT will be different from the SNRs retrieved by telemetry from the real network, creating an error signal that is fed into the SGD algorithm. The algorithm is then repeated until convergence is reached. As shown in Fig. 1a, the QoT estimation workflow starts by creating a database with experimental and synthetic SNR values. The experimental SNRs are read from the physical network, and the synthetic SNRs are loaded from the NDT_{ON}. The NDT_{ON} initial guesses are generated by applying the GN model considering full link spectral loading [14]. The NDT_{ON} is subsequently converted into a simpler digital twin (NDT_{SGD}), where the cascaded spans

¹We denote SNR as the optical GSNR computed assuming an optical noise bandwidth equal to the channel bandwidth.

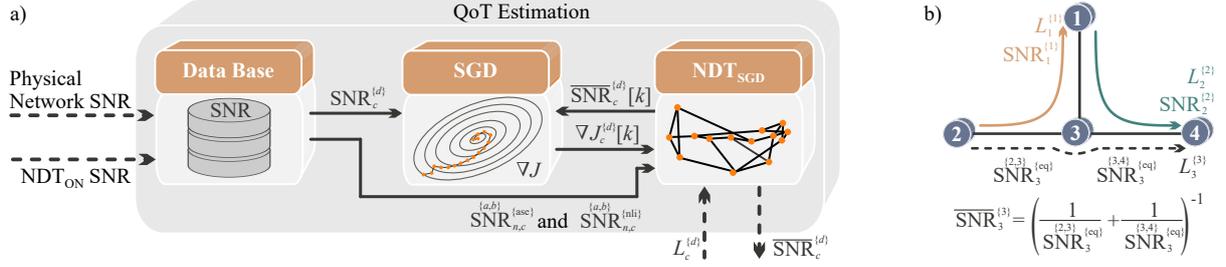


Fig. 1. a) Proposed QoT estimation using a simplified network digital twin (NDT_{SGD}) optimized via SGD. After NDT_{SGD} training, we estimate the desired $\overline{\text{SNR}}_c^{\{d\}}$ of lightpath d traversing the set of links $L_c^{\{d\}}$. b) Example of SNR estimation of a simple 4-node network. Based on the lightpath 1 (light brown) and lightpath 2 (green), we use the SGD algorithm to estimate $\overline{\text{SNR}}_3^{\{3\}}$ of an unestablished lightpath 3 (black dashed line) using optimized $\overline{\text{SNR}}_3^{\{2,3\}\{\text{eq}\}}$ and $\overline{\text{SNR}}_3^{\{3,4\}\{\text{eq}\}}$.

between two neighboring ROADMs are replaced by only one equivalent weighted edge with per-channel SNR given by

$$\overline{\text{SNR}}_c^{\{a,b\}\{\text{eq}\}} = \left[\sum_{n=0}^{N_s} \frac{1}{\overline{\text{SNR}}_{n,c}^{\{a,b\}\{\text{ase}\}}} + \frac{1}{\overline{\text{SNR}}_{n,c}^{\{a,b\}\{\text{nli}\}}} \right]^{-1}, \quad (1)$$

in which N_s is the number of spans between two neighboring ROADMs a and b , and $\overline{\text{SNR}}_{n,c}^{\{a,b\}\{\text{ase}\}}$ and $\overline{\text{SNR}}_{n,c}^{\{a,b\}\{\text{nli}\}}$ are, respectively, the SNRs related to the ASE noise and nonlinear interference, for each span n and channel index c of $\{a,b\}$. In the model, each span is composed of a segment of fiber followed by an amplifier. Index 0 accounts for the booster, where no nonlinear distortion is added.

After computing the equivalent SNRs of all links and channels, we can fit the NDT_{SGD} to the practical network. Using the experimental data (i.e., the SNRs available on transponders), we update the equivalent SNRs of the NDT_{SGD} using the SGD algorithm

$$\overline{\text{SNR}}_{c+s}^{\{a,b\}\{\text{eq}\}}[k+1] = \overline{\text{SNR}}_{c+s}^{\{a,b\}\{\text{eq}\}}[k] - \eta \nabla J_{c+s}^{\{d\}}[k], \quad (2)$$

where k is the training index, η is the learning rate and ∇ is the gradient operator. Note that gradient updating is applied not only to the central channel c , but is transferred to neighboring channels $c+s \in [c-\Delta c, \dots, c+\Delta c]$, where Δc is the number of neighboring channels used during training.

The rationale behind this feature is that neighboring channels should yield equivalent performance. The quadratic cost function $J_{c+s}^{\{d\}}[k]$ of the $(c+s)$ -th channel adjacent to lightpath d is given by

$$J_{c+s}^{\{d\}}[k] = \frac{1}{2} \left(\overline{\text{SNR}}_{c+s}^{\{d\}}[k] - \text{SNR}_c^{\{d\}} \right)^2, \quad (3) \quad \overline{\text{SNR}}_{c+s}^{\{d\}}[k] = \left[\sum_{\{a,b\} \in L_{c+s}^{\{d\}}} \frac{1}{\overline{\text{SNR}}_{c+s}^{\{a,b\}\{\text{eq}\}}[k]} \right]^{-1}, \quad (4)$$

in which $\text{SNR}_c^{\{d\}}$ is read from the optical network (i.e., ground truth) and $\overline{\text{SNR}}_{c+s}^{\{d\}}[k]$ is computed in the NDT_{SGD}, and $L_c^{\{d\}} = L_{c+s}^{\{d\}} \subset L$ is the set of links traversed by lightpath d .

Solving the gradient derivatives of (2) with respect to (3) and (4), yields

$$\overline{\text{SNR}}_{c+s}^{\{a,b\}\{\text{eq}\}}[k+1] = \overline{\text{SNR}}_{c+s}^{\{a,b\}\{\text{eq}\}}[k] - \eta e_{c+s}^{\{d\}}[k] \left[\frac{\overline{\text{SNR}}_{c+s}^{\{d\}}[k]}{\overline{\text{SNR}}_{c+s}^{\{a,b\}\{\text{eq}\}}[k]} \right]^2, \quad (5)$$

where $e_{c+s}^{\{d\}}[k] = \overline{\text{SNR}}_{c+s}^{\{d\}}[k] - \text{SNR}_c^{\{d\}}$ is the SNR error.

Once the NDT_{SGD} is trained, the SNR of an unestablished lightpath d ($\overline{\text{SNR}}_c^{\{d\}}$) can be readily estimated, given as input the set of traversed links $L_c^{\{d\}}$. The scheme is illustrated for a four-node network in Fig. 1b, in which there are two established lightpaths (1 and 2) and the QoT of a third unestablished lightpath (3) arriving in node 4 is estimated.

3. Results

Simulations consider GNet and NSFNet [15] optical networks with 4.8-THz optical wavelength band (C-band), and 12.5-GHz grid. Demands are uniformly distributed in the network, with bandwidth uniformly distributed between 1 to 4 frequency slots (FSs), resulting in a symbol rate of 10 GBaud \times FS. Transponders operate with polarization multiplexing and root-raised cosine shaping filter with $\alpha = 0.15$. Routing and wavelength assignment are performed with Dijkstra and first-fit algorithms, respectively. Route-and-select (R&S) reconfigurable add-drop multiplexers (ROADMs) are equipped with a per-channel power control loop based on optical channel monitors (OCMs) and wavelength selective switches (WSSs), ensuring a -6 -dBm launch power per slot (equivalent to 0-dBm launch power for a 50-GHz FS). We assume 80-km spans with 0.2-dB/km attenuation, except for the last one, which ranges between 50 km and 120 km, to achieve the desired total span length. The idealized NDT_{ON} is simulated considering uniform inline amplifier (ILA) gains and noise figures (NFs) equal to 16 dB

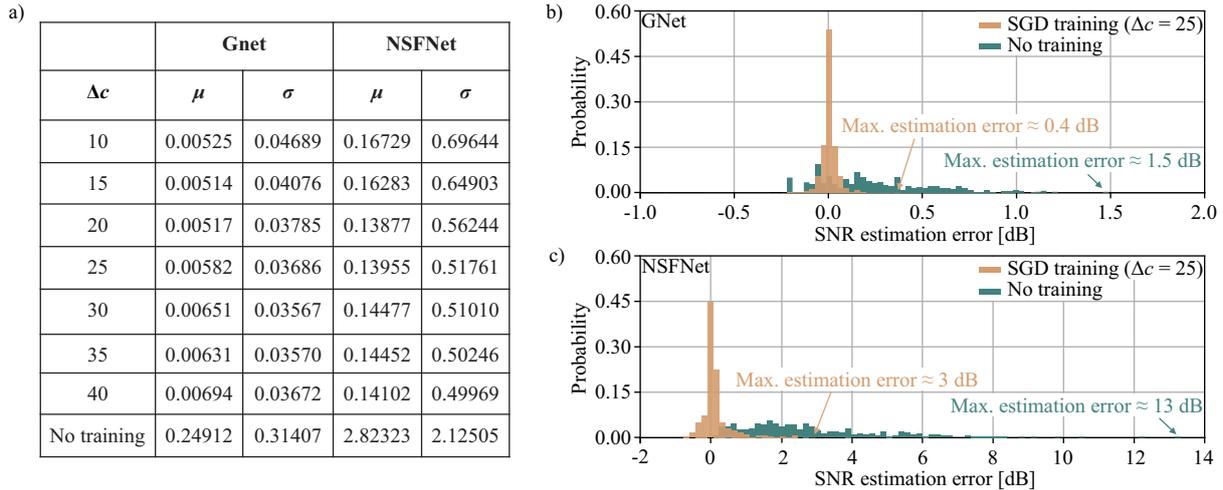


Fig. 2. a) Simulation results and probability distribution of the estimation errors before (green bars) and after training (orange bars), with $\Delta c = 25$. b) GNet, c) NSFNet.

and 5.5 dB, respectively. To emulate real network imperfections, the experimental data set is generated considering inline amplifiers with wavelength-dependent gain. To create wavelength dependency, the ILA gains $g(\lambda_1)$, $g(\lambda_n)$, $g(\lambda_{N_{ch}}) \sim \mathcal{U}\{14, 16\}$ dB, in which n is randomly chosen in the interval $(1, N_{ch})$. The in-between gains are quadratically interpolated, and white Gaussian noise ($\mathcal{N}\{0, 0.0001\}$ dB) is added to each gain. The noise figures (NFs) are also a function of the gains, i.e., $NF(\lambda_n) = [5.5 + 16 - g(\lambda_n)]$ dB, $n \in [1, \dots, N_{ch}]$.

The learning rate and the number of training epochs were chosen large enough to avoid underfitting. For each investigated topology (GNet and NSFNet), 10% of randomly selected services were reserved for inference. Adapted 10-fold cross-validation was employed to verify the proposed approach's robustness over different training samples. In this adaptation, during each round of cross-validation, the training and validation root mean squared error (RMSE) were calculated for every training epoch, and the corresponding system optimization was recorded. Once the last training epoch was computed, we chose the system that had the lowest validation RMSE for inference in order to prevent overfitting. This procedure was repeated 10 times, once per round. We stored the SNR estimation errors to compute the probability distribution, mean, and standard deviation (depicted in Fig.2.a)), with and without training, at the end of the process. Results were obtained as a function of neighboring channels (Δc).

Results are summarized in Fig. 2.a). Before training, resorting only to the GN model with ideal parameters, the mean estimation error for the NSFNet is significantly higher than for the GNet, owing to its longer link lengths. After training, even with $\Delta c = 10$, the proposed QoT estimation is able to shift the mean estimation error of both topologies to approximately 0 dB. For larger Δc , up to 40 wavelengths, we observe a decrease in the error standard deviation. The shape of the error distribution is depicted in Figs. 2.b) and c), for the GNet and NSFNet before (green bars) and after (orange bars) training, for $\Delta c = 25$. Both figures evidence a clear reduction in the mean and standard deviation for the error, evidencing the algorithm's ability to perform accurate QoT in the presence of severe imperfections.

Acknowledgement. This work is supported by Padtec S/A.

References

- [1] N. Guo et al., "Protection against failure of machine-learning-based ...," *J. Opt. Commun. Netw.* **14**, pp. 572–585, (2022).
- [2] E. Seve et al., "Associating machine-learning and analytical models ...," *J. Opt. Commun. Netw.* **13**, pp. C21–C30, (2021).
- [3] M. Ibrahim et al., "Machine learning regression for QoT ...," *J. Opt. Commun. Netw.* **13**, pp. B92–B101, (2021).
- [4] N. Morette et al., "On the robustness of a ML-based method for QoT tool ...," *Proc. of OFC*, paper M3F.1, (2022).
- [5] X. Yang et al., "QoT estimation improvement with inputs refinement tool for ...," *Proc. of OFC*, paper W4G.6, (2023).
- [6] A. A. Díaz-Montiel et al., "Real-time QoT estimation through ...," *IEEE Photon. Technol. Lett.* **33**, pp. 1050–1053, (2021).
- [7] A. Mahajan et al., "Modeling EDFA gain ripple and filter penalties with ...," *J. Light. Technol.* **38**, pp. 2616–2629, (2020).
- [8] L. E. Kruse et al., "Exact component parameter agnostic QoT estimation using ...," *Proc. of OFC*, paper Th1C.1, (2022).
- [9] M. A. Amirabadi et al., "Deep neural network-based QoT estimation for ...," *J. Light. Technol.* **41**, pp. 1684–1695, (2023).
- [10] O. Ayoub et al., "Towards explainable artificial intelligence in ...," *J. Opt. Commun. Netw.* **15**, pp. A26–A38, (2023).
- [11] J. Müller et al., "QoT estimation using EGN-assisted machine ...," *J. Opt. Commun. Netw.* **14**, pp. 1010–1019, (2022).
- [12] K. S. Mayer et al., "Demonstration of ML-assisted soft-failure ...," *J. Light. Technol.* **40**, pp. 4514–4520, (2022).
- [13] D. A. A. Mello et al., "When digital twins meet optical networks operations," *Proc. of OFC*, paper W4A.3, (2023).
- [14] P. Poggiolini et al., "The LOGON strategy for low-complexity control ...," *Proc. of OFC/NFOEC*, paper OW1H.3, (2013).
- [15] K. S. Mayer et al., "Machine-learning-based soft-failure ...," *J. Opt. Commun. Netw.* **13**, pp. E122–E131, (2021).