End-to-end QoT Predictions enhanced by GNPy-based Digital Twin with Network Telemetry

S. Shen, H. Li, A. Tyrovolas, Y. Teng, R. Nejabati, S. Yan and D. Simeonidou
High Performance Networks Group, Smart Internet Lab, University of Bristol, UK
sen.shen@bristol.ac.uk

Abstract: Digital twin for dynamic optical networks is implemented using GNPy, network telemetry and databases. It enhances ML-based End-to-End QoT predictions in field trials by supporting model pre-training and minimizing data requirements through the AI engine.

1. Introduction

Machine learning (ML) technology has demonstrated significant potential in enhancing optical networks, including fault management, anomaly detection, quality of transmission (QoT) estimation, and network traffic prediction [1]. These implementations and demonstrations have generated considerable interest and show great potential for effectively managing and optimising network performance. One of the challenges of ML application deployment is the requirement of a substantial amount of data for model training. Such a challenge is particularly evident in a dynamic environment, such as optical networks that require frequent model updating and retraining. Especially, a sudden change in optical networks necessitates the updating of the existing model instantly.

However, it is unrealistic to obtain substantial data from practical networks within a short period, especially in dynamic optical networks. Also, simulation tools need to obtain real-time network configuration information to evaluate network performance.

This paper presents the development of a purpose-built digital twin for optical networks based on GNPy [2] to facilitate multi-channel end-to-end ML-based quality-of-transmission (QoT) predictions regarding training and updating of ML models. The digital twin, based on GNPy, utilizes real-time network telemetry and distributed time-series databases to create a virtual image of the physical network facilities. This includes a 986-km field-trial testbed, ensuring an accurate representation of the optical networks. With real-time network telemetry, the proposed digital twin duplicates the physical testbed with current network configurations and generates reliable synthetic data for ML applications when retraining is required and there is insufficient available data. Two cascaded multi-channel ANN models are developed using data generated by the digital twin and transfer learning to achieve high end-to-end prediction with the precision error within 0.25dB and reduce practical data requirements by 80%. The GNPy-based digital twin with network telemetry provides a feasible solution to do end-to-end QoT predictions in dynamic optical networks in terms of model updating.

2. Architecture of Digital Twin in Optical Networks

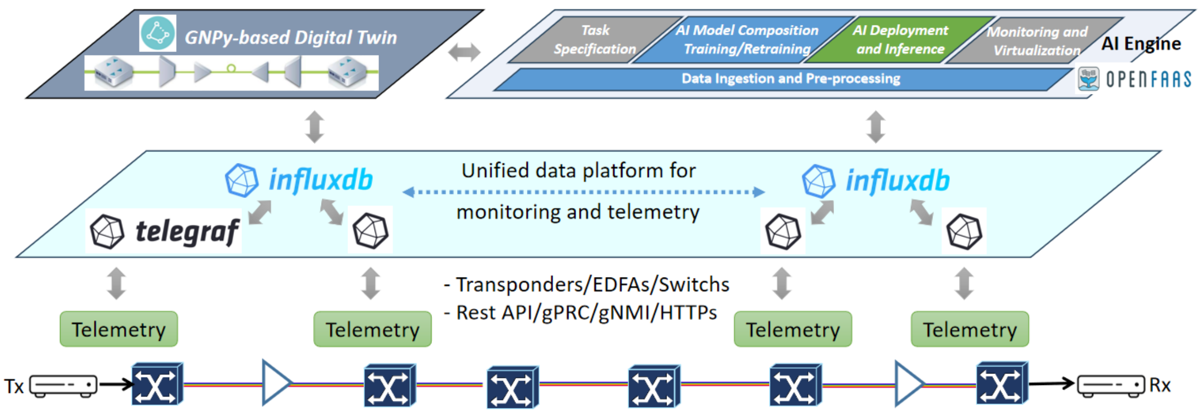

Fig. 1: The schematic view of the GNPy-based digital twin with network telemetry

The network architecture of AI engine-enhanced network intelligence is presented in Figure 1(a), incorporating with real-time telemetry and digital twin [3]. The bottom is the data plane with all network elements (NEs), including transponders, fibre links, wavelength selective switching (WSSs) and EDFAs. In the middle, distributed time-series databases are implemented based on InfluxDB in each node. The databases are responsible for collecting, sampling, and aggregating the data, such as BER values of transponders, optical power setting of EDFAs, and port power values of optical switches from the physical layer, through network telemetry, such as Rest API, gNMI module, and HTTP protocol. Also, Telegraf plugins in the InlfuxDB platform could be used to better collect, process, aggregate, and write metrics from various physical facilities into InfluxDB. The unified data platform monitors real-time network performance and stores network configuration information and performance parameters of NEs in the database. The time-series database associates each data point with a timestamp, enabling the digital twin and AI to efficiently retrieve and analyze the network's current status. The unified monitoring and databases support both the AI engine that manages the life cycle of multiple AI applications, and the GNPy-based digital twin.

The digital twin is developed based on GNPy to support ML model training and retraining by interacting with AI engine. To capture physical network changes, the GNPy-based digital twin is connected with unified monitoring and databases to obtain practical network information in real-time. The digital twin is specifically designed to generate synthetic data for the AI engine platform to support ML model training and retraining, with practical network configurations and the parameters of NEs such as channel status, transmit power, EDFA gains and fibre lengths feed.

In the defined use case, ML models for QoT prediction need to be retrained due to unexpected network changes. Without enough practical data supporting the retraining, the AI engine requests the digital twin to generate the synthetic data for ML model pre-training based on the current network status. Subsequently, the GNPy-based digital twin duplicates the physical link setup with the information provided by the data platform and generates synthetic data for ML model pre-training. Then, ML models trained with synthetic data inside the AI engine will leverage practical data stored in InfluxDB for transfer learning to improve the ML model performance. After that, requested by the SDN controller through APIs provided by OpenFaas, AI engine will fetch real-time data from time-series databases, and feed it to the deployed ML models with data pre-processing to make real-time QoT prediction.

Table 1: ANN models trained with synthetic data

Synthetic Data	ANN1	ANN2	End-to-End Prediction
MSE	0.0086	0.0049	0.2454
MAE	0.0543	0.0441	0.2780
R2	0.9993	0.9995	0.9866

Table 2: ANN models trained with practical data

Practical Data	ANN1	ANN2	End-to-End Prediction
MSE	0.0039	0.0073	0.0098
MAE	0.0419	0.0593	0.0649
R2	0.9999	0.9996	0.9995

Table 3: ANN models trained with transfer learning

Transfer learning	ANN1	ANN2	End-to-End Prediction
MSE	0.0050	0.0069	0.0084
MAE	0.0415	0.0491	0.0526
R2	0.9998	0.9996	0.9995

3. Field-trial Demonstration and Results

Figure 2(a) illustrates the 986-km field trial testbed implemented over part of the UK National Dark Fibre Facility (NDFF), comprising three nodes and two optical links from Bristol(UoB) to Power Gate bypassing Bradley Stoke, Froxfield, and Reading. In the experimental setup, 12 $\times$ 32Gbaud PM-16QAM signals are generated by two Facebook Voyager transponders and two ADVA FSP3000 Teraflex platforms and then are aggregated into the same NDFF link via a WSS with power equalization from node A to C, where 12 transmission wavelengths range from 193.45THz to 194.0THz with 50GHz grid respectively. At node B, a two-degree route-and-select reconfigurable optical add-drop multiplexer (ROADM) is deployed using two WSSs to switch, add and drop optical channels. Through configuring the ROADM, some channels transmitted from Node A will be dropped depending on spectrum allocation, and sent back to the corresponding receivers for detection and measurement. New optical channels will be added through the ROADM, joined with the remaining optical channels from Node A, and then forwarded to the second link. After the 493 km fibre link, all signals are separated using the WSS and sent back to the transponders for Q-factor measurements at Node C.

Two ANN-based link penalty estimation models are developed to predict end-to-end QoT performance over two links by cascading two models [4]. In this experiment, due to the link changes, the two models require retraining during the operations. Without enough data after the failure of the previous ANN models, synthetic data are needed to accelerate the training process. So the GNPy-based digital twin is used to generate data based on the new network conditions. The updated information, including the transceiver setup, EDFA gains and fibre parameters are feed into the GNPy-based digital twin to generate all channel statuses of 12 optical channels and the corresponding GSNR values to prepare synthetic data. Subsequently, two ANN models for the two links are trained with 1500 synthetic data samples and cascaded to provide end-to-end QoT predictions. Then, the trained ANN models are transferred to the practical scenario, which is re-trained with 200 samples of practical data. The deployed ANN model structure is presented in [4].

Tables 1, 2, and 3 summarize the error metrics of two ANN models and the cascaded models trained with

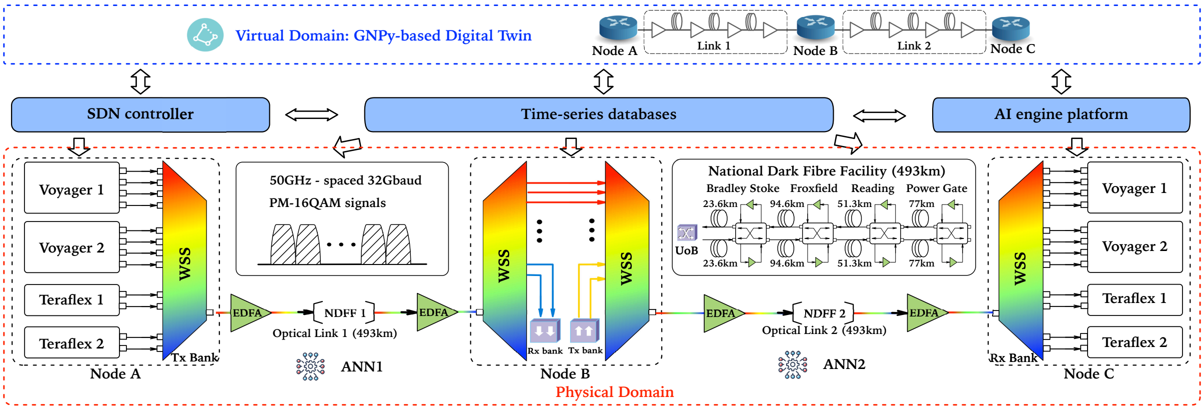

Fig. 2: Experimental setup of the 986-km field-trial with a digital twin

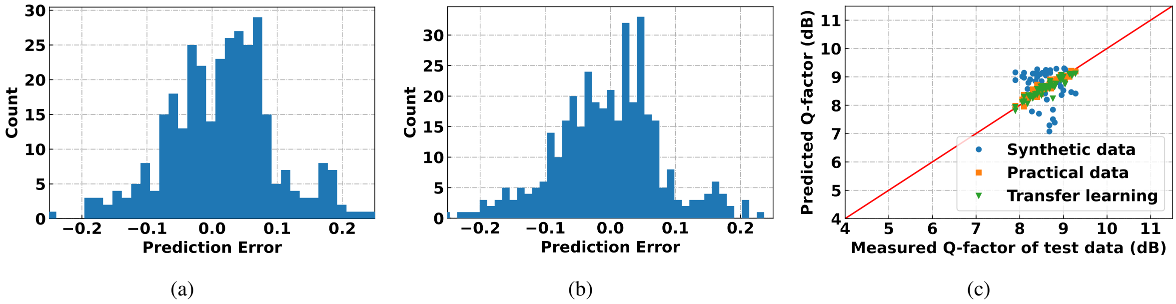

Fig. 3: End-to-end QoT prediction results:(a) Distribution of prediction errors for the cascaded ANN model trained with practical data, (b) Distribution of prediction errors for the cascaded ANN model trained with transfer learning, (c) Predicted Q-factor vs the measured Q-factor for end-to-end performance.

synthetic data, practical data and transfer learning. With the two ANN models for the two links, the end-to-end performance for optical channels transversing both links can be predicted by cascading the two models. It can be seen that ANN models trained with synthetic data generated by GNPy-based digital twin with the physical parameters feed fit well, but with high MSE and MAE in the end-to-end prediction testing on the practical data. By retraining with a few practical data, the cascaded ANN model achieves the same prediction accuracy up to 99.95% with low MSE and MSE, compared with the model trained entirely with practical data. Figure 3(a) depicts the prediction error of the ANN model trained entirely with practical data, whereas Figure 3(b) shows the prediction error of the ANN model trained with transfer learning. Obviously, the ANN model trained with transfer learning performs well with a prediction error within 0.25dB. It can be seen intuitively from Figure 3(c) that GNPy-based digital twin enhanced the end-to-end QoT predictions over multiple links. It is worth noting that the ML approach outperforms the GNPy-only approach.

4. Conclusion

In this paper, we introduced a GNPy-based digital twin designed to bolster end-to-end QoT prediction by seamlessly integrating network telemetry, time-series databases, and an AI engine over a field-trial testbed. Notably, our proposed digital twin produced dependable synthetic data essential for model pre-training. Leveraging transfer learning, the QoT prediction models exhibited marked improvement, achieving remarkable accuracy across multiple links with a prediction error confined to 0.25 dB. This pioneering effort paves the way for the practical deployment of digital twins, setting a new standard for enhancing end-to-end QoT predictions in ever-evolving dynamic optical networks.

5. Acknowledgement

This work was partly supported by the European Commission's Horizon research and innovation program: Allegro project (No. 101092766) and the UK EPSRC project: NDFF (No. S028854).

References

1. Z. Gao, *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2646–2655, 2020.
2. Ferrari, in *2020 International Conference on Optical Network Design and Modeling (ONDM)*, 2020, pp. 1–6.
3. S. Shen, in *49th European Conference on Optical Communications (ECOC 2023)*, 2023, pp. 1–4.
4. R. Yang, in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, 2023, pp. 1–3.