Field-Trial Demonstration of ML Deployment in Optical Networks Using Telemetry and AI Engine

Sen Shen, Ruizhi Yang, Haiyuan Li, Yiran Teng, Rui Wang
Reza Nejabati, Shuangyi Yan, Dimitra Simeonidou

High Performance Networks Group, Smart Internet Lab, University of Bristol, UK sen.shen@bristol.ac.uk

Abstract *We present an AI engine built on OpenFaas utilizing distributed databases and network telemetry to monitor and manage multiple ML models. Two use cases of ML applications in optical networks are demonstrated over the field-trial testbed, showcasing the feasibility and scalability of the proposed scheme.©2023 The Author(s)*

Introduction

Recently optical networks have seen significant advancements with the integration of machine learning (ML) technologies for various applications such as fault management[1], anomaly detection[2], quality of transmission (QoT) prediction[3], and network traffic prediction[4]. These implementations and demonstrations have generated considerable interest and promise for future deployments. However, deploying large numbers of ML models in practical scenarios remains a significant challenge. Especially, efficient and scalable monitoring and data collection solutions with real-time telemetry are essential for all machine learning applications. Deployment of multiple ML models also requires extensive ML model management including training, deployment, inference, and monitoring.

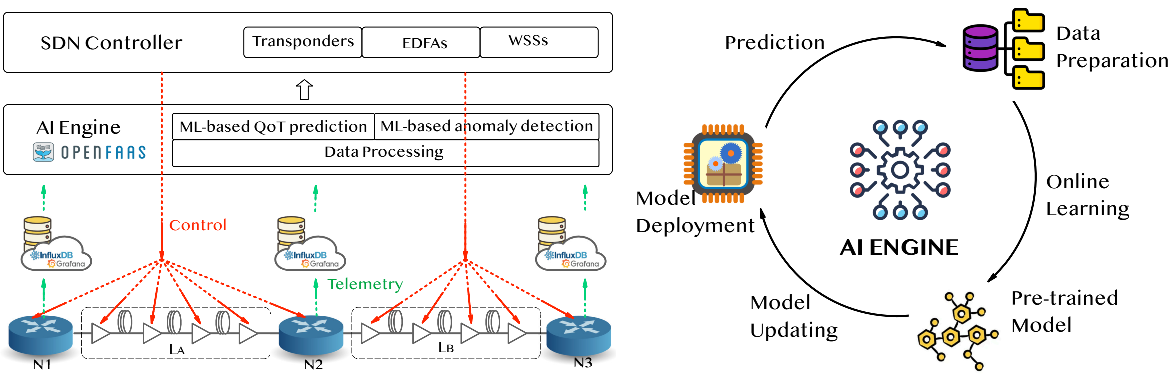
To address this challenge, Machine-Learning-as-a-Service (MLaaS) has been proposed as a framework to allow optical network automation entities to request and quickly access trained ML models with as little human intervention as possible[5]. However, the current work didn't address ML model management and deployment in optical networks and didn't combine it with real-time telemetry to provide a life-cycle optical monitoring platform. In addition, optical network telemetry has been developed supporting both performance monitoring and network state reporting, with monitoring and telemetry protocols, such as gRPC, gNMI, or websockets[6]. As far as authors know, the practical demonstration with telemetry and AI engine has not been reported yet.

This paper presents the implementation and demonstration of an AI engine utilizing distributed databases and network telemetry over a field-trial testbed. The AI engine, built on OpenFaas, is capable of monitoring and managing multiple ML models for different use cases. Time series databases based on InfluxDB are implemented over the UK National Dark Fiber Facility (NDFF) to store collected network status through Flux data streaming. The paper demonstrates the application of the AI engine in two use cases: (1) artificial neural network (ANN)-based QoT predictions for optical signals transversing two fibre links, and (2) link anomaly detection with long-short term memory (LSTM) algorithms. The demonstrated solution integrates optical network telemetry, time-series databases, AI engine with ML applications to provide extra network functions in optical networks. The field-trial demonstration paves a possible way for the practical deployment of ML applications in optical networks.

Architecture of AI Engine in Optical Networks

To fully leverage the potential of ML applications in optical networks, the entire ML life-cycle[7], involving data processing, ML model training and deployment, should be considered. In fact, deployments of ML models require the integration of the database and network controller to combine the models with the network to realize their functions. Figure 1(a) illustrates the proposed network architecture with real-time telemetry and an AI engine, where three essential functions need to be in place:

1) AI Engine: it manages algorithms, models and frameworks used to develop and train ML models. The AI engine performs critical ML tasks, such as training and inference, and facilitates the deployment of ML models in optical networks. It must be able to handle large amounts of data and perform complex computations to generate accurate predictions. The AI engine can be deployed on a dedicated server or cloud infrastructure. It can also be deployed on network devices such as routers or switches to enable real-time pro-



(a) The proposed architecture (b) The designed AI engine

Fig. 1: The schematic view of the proposed network telemetry with an AI engine

cessing and decision-making. In the demonstration of the proposed network architecture, OpenFaas is selected to develop the AI engine. As a serverless computing platform, OpenFaas provides a cloud-native environment for deploying an environment for ML model deployments, enabling real-time prediction and decision-making in optical networks.

2) Telemetry: it is responsible for collecting, processing, and aggregating the data from the physical layer and making it available to the AI engine. Telemetry data in optical networks can be obtained from various sources, such as optical performance monitoring, optical power levels, and optical signal-to-noise ratio (OSNR) values. Telemetry protocols such as gRPC, gNMI, or websockets can be used to facilitate the collection and transmission of telemetry data in optical networks. In our demonstration, time-series databases are implemented based on InfluxDB in each node to store the collected data from the physical layer through InfluxDBClient and Flux, which is a query language designed specifically for working with time-series data in InfluxData. Through the integration with the visualization tool Grafana, time-series databases provide operators with real-time insights into network performance shown in Figure 3(a). The board shows real-time BER values received by one of the voyager transponders.

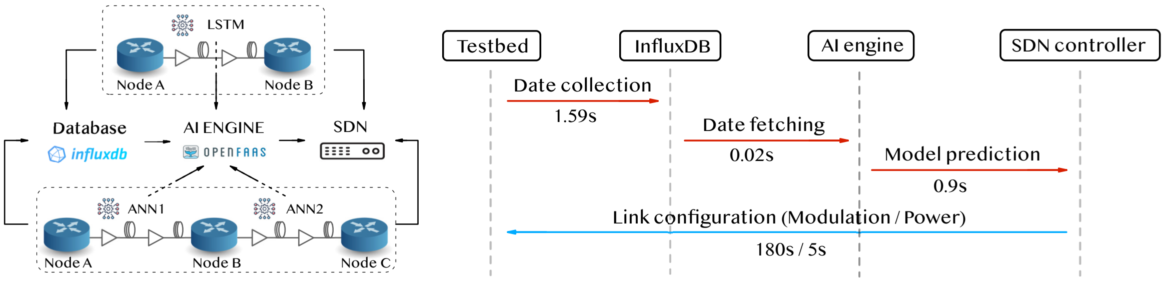
3) SDN controller: it is responsible for network control and management, and plays a critical role in the deployment of ML models in optical networks. The SDN controller can communicate with network devices using various protocols such as OpenFlow, NETCONF, or REST APIs to configure the network and make intelligent decisions for network optimization and fault detection according to the prediction results provided by the AI engine. Experimentally, an SDN controller is developed to control all the devices such as Wavelength selective switching (WSSs), Voyager Transponders, and facilities through REST-APIs and HTTP POST.

With distributed time-series databases and the SDN controller, the AI engine is applied to predict network traffic and detect network anomalies. ML models inside the AI engine are pre-trained offline with part of previously collected data, and deployed on OpenFaas. With ML models deployed, the AI engine automatically fetches the data from InfluxDB to make real-time predictions with the results sent to the SDN controller. Then, the SDN controller will do link configuration according to the prediction results. As shown in Figure 1(b), when the link is reconfigured or the network topology is changed, the AI engine will fetch the latest data, retrain the models and make re-deployments. This approach ensures that the ML models remain relevant and effective in managing the network, even as the network topology and traffic patterns change over time. By updating the models, the AI engine can adapt to new conditions and improve its predictions, allowing the network to be more efficiently managed.

Use cases and results

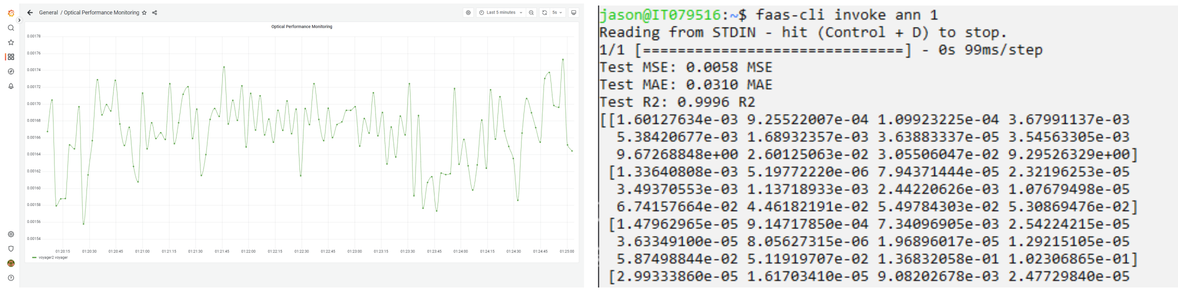
To validate the feasibility of the proposed architecture, we demonstrated it in the field-trial testbed over part of NDFF. Figure 2(a) depicts the link setup of two use cases in the field-trial demonstration.

First, we demonstrated the proposed system in LSTM-based link anomaly detection[8]. From node A to node B, 4 equalized 50GHz-spaced 32Gbaud PM-16QAM signals are generated by 4 Voyager Transponders and launched into the 493km NDFF link. An LSTM model is deployed in the AI engine, reading the previous data stored



(a) Two different demonstration cases (b) Time and sequence diagram of the proposed architecture

Fig. 2: Field-trial demonstration setup and the workflow



(a) Optical performance monitoring based on InfluxDB (b) AI engine prediction deployed on OpenFaas

Fig. 3: Results of field-trial demonstration

in InfluxDB to make short-term QoT predictions. As shown in Figure 2(b), the proposed architecture has a fast response to internal messaging. In this case, it takes a short time for Influxdb to collect data from the testbed and for the AI engine to read data from the database, with the time cost of 1.59s and 0.02s respectively. Generally, the time to write and query data in InfluxDB is within milliseconds, mainly depending on the amount of data, the writing and querying method. The time of AI inference is less than 1s. Due to the factors such as the size and complexity of the function, the load on the OpenFaaS cluster and network latency, it takes a few seconds to invoke OpenFaaS and run the function. The response time from data acquisition to model prediction within several seconds proves that the system can provide real-time network performance monitoring and prediction. However, it will take a bit of time for the completion of the link configuration since devices need a certain amount of time to respond to commands sent by the SDN controller. For instance, it will take around 180s for the modulation format configuration of Facebook Voyager transponders and ADVA FSP3000 Teraflex devices we used, and around 5s for the power configuration.

In addition, the proposed system is demonstrated on cascaded ANN-based link-penalty models for QoT prediction[9]. 12 optical channels with 50GHz-spaced 32Gbaud PM-16QAM signals are generated from two Facebook Voyager devices each with four transponders and two ADVA FSP3000 Teraflex platforms, which support four transponders in total. They are aggregated into the same fibre via a wavelength selective switch (WSS) from node A to node C through two NDFF sections of 493km. In node B, a ROADM is deployed to switch, add and drop optical channels. For each link, it has an ANN model deployed in the AI engine to predict the Q-factor of the current link, which requires the AI engine to be able to support the deployment of multiple models simultaneously. As shown in Figure 3(b), quick AI inference with a time of 99ms and high prediction accuracy demonstrates the robustness and scalability of the designed AI engine.

Conclusion

We have successfully demonstrated the deployment of an AI engine with ML applications, integrating optical network telemetry and time-series databases over a field-trial testbed. . The proposed architecture offers great advantages for the practical deployment of ML applications in optical networks.

Acknowledgements

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

References

[1] X. Chen, C.-Y. Liu, R. Proietti, J. Yin, Z. Li, and S. J. B. Yoo, “On cooperative fault management in multi-domain optical networks using hybrid learning”, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 4: Mach. Learn. in Photon. Commun. and Meas. Syst. Pp. 1–9, 2022. DOI: 10.1109/JSTQE.2022.3151878.

[2] K. Abdelli, J. Y. Cho, F. Azendorf, H. Griesser, C. Tropschug, and S. Pachnicke, “Machine-learning-based anomaly detection in optical fiber monitoring”, *Journal of Optical Communications and Networking*, vol. 14, no. 5, pp. 365–375, 2022. DOI: 10.1364/JOCN.451289.

[3] Z. Gao, S. Yan, J. Zhang, *et al.*, “Ann-based multi-channel qot-prediction over a 563.4-km field-trial testbed”, *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2646–2655, 2020. DOI: 10.1109/JLT.2020.2971104.

[4] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, “Quality of transmission estimation and short-term performance forecast of lightpaths”, *Journal of Lightwave Technology*, vol. 38, no. 10, pp. 2807–2814, 2020. DOI: 10.1109/JLT.2020.2975179.

[5] C. Natalino, N. Mohammadiha, and A. Panahi, “Machine-learning-as-a-service for optical network automation”, in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, W4G.3. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2023-W4G.3.

[6] R. Vilalta, R. Casellas, R. MartÃnez, R. MuÃ$\pm oz$, A. GonzÃ¡lez-MuÃ$\pm iz$, and J. FernÃ¡ndez-Palacios, “Optical network telemetry with streaming mechanisms using transport api and kafka”, in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4. DOI: 10.1109/ECOC52684.2021.9606002.

[7] P. Paudyal, S. Shen, S. Yan, and D. Simeonidou, “Toward deployments of ml applications in optical networks”, *IEEE Photonics Technology Letters*, vol. 33, no. 11, pp. 537–540, 2021. DOI: 10.1109/LPT.2021.3074586.

[8] S. Shen, R. Yang, H. Li, *et al.*, “Lstm assisted optical transmission performance analysis over a 493-km field-trial”, in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, W4G.4. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2023-W4G.4.

[9] R. Yang, S. Shen, H. Li, *et al.*, “986 km field trial of cascaded ann-based link-penalty models for qot prediction”, in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, W4G.5. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2023-W4G.5.