Scaling Optical Network Fault Management with Decentralized Graph Learning

Qunzhi Lin, Xiaokang Chen, Zhenlin Ouyang, Hanyu Gao, Xiaoliang Chen and Zhaohui Li

School of Electronics and Information Technology & Guangdong Provincial Key Labratory of Optoelectronic Information Processing Chips and Systems, Sun Yat-sen University, Guangzhou, China, and Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China

xlichen@ieee.org

Abstract: We propose a decentralized graph learning framework for scaling cognitive fault management in optical networks. Results show the proposed design achieves $> 96\%$ fault identification and localization accuracy.

1. Introduction

Effective fault management for optical networks is crucial for securing the correct delivery of tens of thousands of Internet services. Traditional approaches typically rely on threshold-based fault detection and manual troubleshooting, which can result in prolonged fault management cycles and considerable traffic losses. Recently, machine learning (ML) has demonstrated compelling prospects in automating and intelligentizing optical network fault management, including soft failure detection [1], identification [2], and localization [3]. However, most of the existing ML solutions were designed to be executed in a centralized control plane with global data accessability. This entails powerful telemetry services that constantly stream local optical performance monitoring (OPM) data to network controllers, adding nonnegligible control plane overheads and survivability concerns. Meanwhile, global ML models often scale with the sizes of topologies or datasets [4], which restricts their applicability to only small-scale networks. For instance, a neural network performing fault location prediction has an output layer whose number of neurons is equal to the total number of optical elements throughout the network. Consequently, frequent model retraining is also necessary upon network upgrades. In this context, distributed intelligence can be seen as a promising solution for meeting the aforementioned challenges [5, 6]. Nevertheless, few studies have reported ML-based distributed fault management designs for optical networks.

In this paper, we present a decentralized graph learning design for scaling cognitive optical network fault management. The proposed design makes use of decentralized fault management (DFM) agents that perform local fault detection and root cause identification. The DFM agents learn cooperatively through repeated message passing and feature aggregation, which allows them to propagate state data topologically and hereby learn the complex patterns of fault propagation. Performance evaluations with data collected from a six-node topology show high fault identification and localization accuracy achieved by the proposed design.

2. Principle

Fig. 1(a) shows the schematic of decentralized fault management, where each optical node is equipped with OPM functions and an AI/ML-empowered DFM agent that performs local decision making on fault detection, identification and localization tasks. Because faults can propagate over networks, utilizing purely local OPM data is often inadequate for identifying the root cause of a fault. For instance, a malfunctioning amplifier in the central node can trigger concurrent anomaly indicators (e.g., deviations of signal quality-of-transmission) at downstream nodes. Rather than all reporting OPM data to the network manager to resort to global data aggregation and inferences, adjacent DFM agents communicate to exchange necessary information (OPM data, gradients, etc.) for pursuing collaborative learning. Each agent thus can employ an identical ML model whose complexity, unlike that of a centralized solution, does not scale with the size of the network. Consequently, the DFM framework prevails its centralized counterparts by achieving high scalability and cost-effectiveness.

Faults propagate with complex graphical patterns and affect the OPM data from different nodes with varying degrees as we route different sets of lightpaths over a network. Learning the correlation between OPM data while incorporating the topological information is essential for capturing such patterns. We take the advantage of graph learning in extracting topological correlations and propose to realize DFM with decentralized operations of graph learning (DeGL), in particular, graph convolutional network (GCN) [7]. Fig. 1(b) illustrates the principle of DeGL over a three-node linear topology. *(1) Initialization*: the state of each DFM agent is initialized to contain local OPM parameters such as the power and noise level of each channel. *(2) Message passing*: each agent advertises

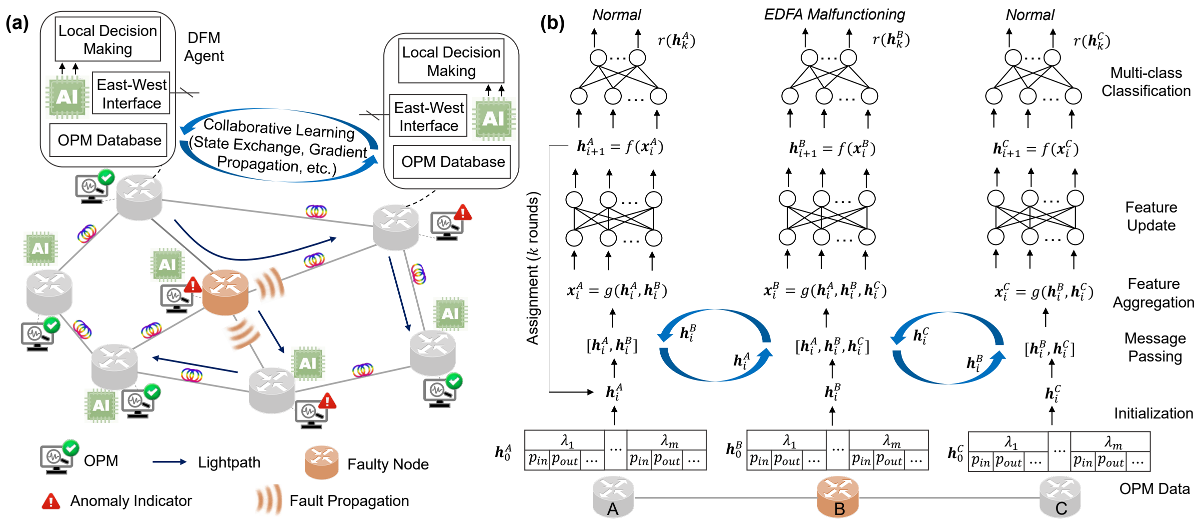


Fig. 1: (a) Schematic of decentralized fault management and (b) illustration of decentralized graph learning over a three-node linear topology. DFM: decentralized fault management. OPM: optical performance monitoring.

its state to the neighboring agents, for instance, *Node B* receives $\mathbf{h}_i^A$ and $\mathbf{h}_i^C$ while *Node A* receives $\mathbf{h}_i^B$. *(3) Feature aggregation*: the local and received states are aggregated by,

$$[\mathbf{x}_i^A\mathbf{x}_i^B\mathbf{x}_i^C]' = g([\mathbf{h}_i^A\mathbf{h}_i^B\mathbf{h}_i^C]') = D^{-\frac{1}{2}}\tilde{G}D^{-\frac{1}{2}}[\mathbf{h}_i^A\mathbf{h}_i^B\mathbf{h}_i^C]', \tilde{G} = D^{-\frac{1}{2}}(G+I))D^{-\frac{1}{2}}, D_{ii} = \sum_j (G+I)_{i,j}, \tag{1}$$

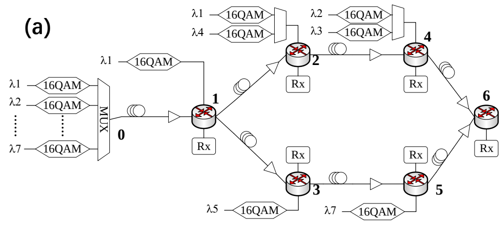
where G is the adjacent matrix of the topology, I is the identity matrix and D is a diagonal matrix. The rationale behind Eq. 1 is to weight states based on node degrees to avoid gradient vanishing or exploding. Afterward, the features of all the agents have the same dimension. *(4) Feature update*: each agent makes use of a neural network $f_{\theta_f}(\cdot)$ to generate an updated state from the aggregated features, where θ_f is the set of trainable weights. Steps *(1)-(4)* are repeated for multiple rounds (e.g., k), allowing for each node's state being disseminated and exploited throughout the network. This way, the patterns of fault propagation can potentially be learned. *(5) Multi-class classification*: finally, each agent predicts with a multi-class classification layer $r_{\theta_r}(\cdot)$ whether a local fault presents and if so, the root cause of the fault (e.g., EDFA malfunctioning). *(6) Training*: the agents (more specifically, θ_f and θ_r) can be trained cooperatively by a gradient descend method that minimizes the cross-entropy function.

3. Performance Evaluation

We evaluated the performance of the proposed decentralized fault management design with simulation data collected from a six-node topology shown by Fig. 2(a). Each fiber link is of 50 kilometers and amplified by a EDFA. We set up 12 lightpaths in parallel and Fig. 1(b) summarizes the configuration of the lightpaths. Each transmitter operated at 56 Gbps and one of the seven wavelength channels ranging from 193.1 to 193.7 THz. All the lightpaths adopted the 16-QAM modulation. We emulated three types of faults by manipulating the optical fiber attenuation, amplifier noise figures and insertion losses of ROADM nodes. The ranges of these parameters and the choices of fault threshold are presented in Fig. 1(c).

We considered the case of single faults, i.e., only a single fault was introduced to one of the six nodes at a certain point. In total, 1,900 samples were collected, including 100 samples for each fault type in each node and 100 normal ones. Note that, for each node, only 300 samples out of the entire set belong to the faulty category. The state of each node is composed by the power, noise levels and OSNR of the seven channels. With DeGL, each agent predicts whether a sample is normal or corresponds to one of the three types of faults. We compared DeGL with a multi-layer perception (MLP) model of two hidden layers (95 and 57 neurons, respectively). The MLP model can be treated as a centralized solution that takes the states of all the nodes as input and performs a 19-class (one normal class plus three types of faults for each node) classification task. To explore the impact of message passing on the performance of DeGL, we configured DeGL with the number of message passing layers ranging from one to three. We trained the DeGL models using the Adam algorithm with a learning rate of 0.001 and a batch size of 30. To avoid over-fitting, we set a heuristic patience threshold of 40, which refers to an early-stopping method that terminates training when the validation loss remains stable for 40 epochs. For MLP, we set the learning rate and batch size to 0.007 and 190, respectively. The data set was divided into the training, validation and testing sets with a ratio of 3 : 1 : 1.

Fig. 2(d) shows the results of fault detection accuracy from different models. We can observe DeGL with two message passing layers performs the best among the three configurations. We presume that this is because



(b)	1	2	3	4	5	6	7	8	9	10	11	12
Path	0-1	1-2	2-4-6	0-1-2-4	4-6	0-1-2	0-1-3	0-1-3	3-5-6	0-1-3-5	5-6	0-1-3-5-6
Fiber Length /km	50	50	100	150	50	100	100	100	100	150	50	200
Launch Power /dBm	10											

(c) Optical Fiber Attenuation (dB/km)		Optical Amplifier Noise Figure (dB)		ROADM Insertion loss (dB)	
Range	Fault Threshold	Range	Fault Threshold	Range	Fault Threshold
0.2-0.5	0.5	4-10	10	0-10	10

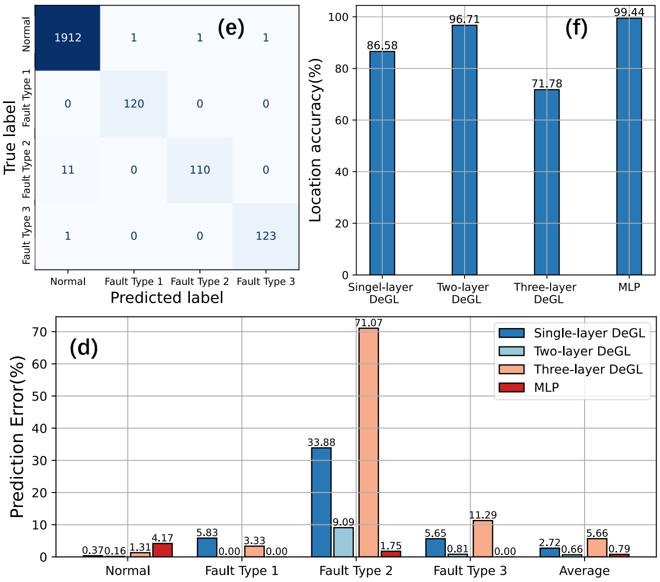


Fig. 2: (a) Six-node simulation testbed; (b) lightpath configuration; (c) configuration of optical parameters; (d) results of fault detection accuracy; (e) confusion matrix of the two-layer DeGL; and (f) results of fault localization accuracy.

increasing the number of layers from one to two helps the agents propagate their state across the network and thus facilitates learning the patterns of fault propagation, whereas further increasing the number of layers makes all the agents possess highly overlapping receptive fields because of the relatively small scale of the network under evaluation. On average, the two-layer DeGL achieves a prediction error of 0.66%, comparable to that from MLP, which has a higher false alarm rate. Fig. 2(e) shows the confusion matrix for the two-layer DeGL, where the vertical and horizontal axes represent the ground truth and predictions, respectively. It can be seen that DeGL achieves approximate 100% detection accuracy for fault types 1 and 3 and performs slightly worse for fault type 2, which corresponds to the introducing of abnormal amplifier noise figures. Fig. 2(f) shows the results of fault localization accuracy. Again, the two-layer DeGL outperforms the single-layer and three-layer configurations with an accuracy of 96.71%, while MLP achieves the best performance among all. Note that, MLP could suffer from scalability issues as its output layer scales up quickly with the size of the network and the number of fault types.

4. Conclusion

This paper presented a scalable optical network fault management framework leveraging decentralized graph learning. Performance evaluations verified the effectiveness of the proposed design.

References

1. X. Chen *et al.*, *IEEE Commun. Mag.* 60, 88–94 (2022).
2. S. Shahkarami *et al.*, *Proc. of Opt. Fiber Commun. Conf.*, paper M3A.5 (2018).
3. K. S. Mayer *et al.*, *J. Opt. Commun. Netw.* 13, E122–E131 (2021).
4. P. Safari *et al.*, *Proc. of Opt. Fiber Commun. Conf.*, paper Th4J.3 (2021).
5. L. Velasco *et al.*, *J. Opt. Commun. Netw.* 15, 676–686 (2023).
6. G. Bernárdez *et al.*, *IEEE Tans. Cogn. Commun. Netw.* 15, 676–686 (2023).
7. T. N. Kipf *et al.*, *Proc. of ICLR*, arXiv:1609.02907 (2017).

This work was supported in part by NSFC under Project 62201627, Guangzhou Basic and Applied Basic Research Foundation (No. SL2022A04J01907), the Guangdong Program (No. 2021QN02X039), and Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. SML2022007).