Suspect fault screen assisted graph aggregation network for intra-/inter-node failure localization in ROADM-based optical networks [Invited]

RUIKUN WANG[1], JIAWEI ZHANG[1,*], SHUANGYI YAN[2], CHUIDIAN ZENG[1], HAO YU[3], ZHIQUN GU[1], BOJUN ZHANG[1], TARIK TALEB[3], AND YUEFENG JI[1]

[1] State Key Lab of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing, China
[2] High Performance Networks Group, Smart Internet Lab, University of Bristol, Bristol, UK
[3] Centre for Wireless Communications, University of Oulu, Oulu, Finland
[*] Corresponding author: zjw@bupt.edu.cn

Compiled February 18, 2023

In optical networks, failure localization is essential to stable operation and service restoration. Several approaches were presented to achieve accurate failure localization of nodes and inter-nodes. However, due to the increasing of traffic and demanding for flexibility, the reconfigurable optical add/drop multiplexer (ROADM) is evolving towards a multi-degree architecture. Therefore, each ROADM is composed of multiple devices, which makes the failures of intra-node become more complex. In this context, failure localization of intra-node can effectively reduce the pressure on network operators to further find specific devices. In this work, we redefine the failure model of intra-/inter-node for multi-degree ROADM-based optical networks and propose a suspect fault screen assisted graph aggregation network (SFS-GRN) for intra-/inter-node failure localization. SFS is responsible for screening out suspect fault devices from all devices and reducing the number of candidate devices. GRN is used to analyze these monitoring data from optical performance monitoring (OPM) in node- and network- wide, and determine the most likely failure device. The proposed scheme is evaluated in a 9-node simulated network and a 3-node testbed network, respectively. Extensive results show that SFS-GRN achieves higher accuracy compared with existing methods under different percentage of OPM deployment, number of service requests and failure types. SFS can remove more than 98% of devices, which is beneficial to further detect and repair for network operators. Moreover, the proposed strategy takes about $10ms$ to detect a potential failure and it has the potential to be applied to a real scenario.

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

1. INTRODUCTION

In the fifth-generation fixed network (F5G) and beyond era, optical networks have been considered as the important physical infrastructures to accommodate massive service requests [1]. Different from other network domains (e.g., IP-level), optical networks expect higher reliability since failures in optical networks may lead to many data losses or even service interruption. Thus, failure management is crucial to optical networks, and it becomes an imperative concern for network operators. Specifically, failure localization focuses on accurately finding specific fault objects (e.g., a fiber span), which is beneficial to fault repairing and service recovering. However, failure localization still depends on human intervention, where operation experts analyze alarm logs and contexts to locate faults [2, 3]. This is a time-consuming and inefficient way for large-scale optical networks. On the other hand, with the development of several advanced technologies, including software defined network (SDN) [4], network telemetry [5], and artificial intelligence (AI) [6], the automation and intelligence levels of optical networks have been improved. In this context, failure localization is still desiring the more advanced techniques to find faults accurately and automatically.

Currently, most optical networks are based on reconfigurable optical add/drop multiplexers (ROADMs) since ROADMs enable network operators to flexibly establish and release lightpaths [7]. However, the existing ROADM is composed of several devices, such as wavelength selective switch (WSS), arrayed waveguide grating (AWG), erbium-doped fiber amplifier (EDFA) and splitter. Meanwhile, to meet the fast growth of traffic demand, ROADM node is evolving towards multi-degree archi-

tectures, e.g., 32-degree [8–10]. Thus, each ROADM consists of multiple devices, which makes the failures of intra-node become more complex. In this condition, failure localization of intra-node can effectively reduce the pressure on network operators to further locate specific devices. Moreover, it also ensures that other degrees (i.e., no-failure devices) within ROADM node are available when service re-routing. The above advantages indicate that the optical network with multi-degree ROADMs requires a more fine-grained failure localization scheme (i.e., intra-/inter-node), where the failures of inter-node include several fiber spans and EDFAs.

To locate failures of intra-/inter-node, it is an effective way to deploy many optical performance monitoring (OPM) into intra-/inter-node for monitoring the states of the physical layer, such as signal power [11]. However, network operators expect to deploy fewer OPM for reducing CAPEX, which may result in ambiguity in some failure locations. Meanwhile, because of flexible traffic exchange between different degrees (i.e., node-level) and different nodes (i.e., network-level), it reinforces the complex dependency between monitoring data. Therefore, how to achieve accurate failure localization of intra-/inter-node at low-OPM ratio becomes a crucial problem.

A. Failure Localization Schemes on a ROADM-based Optical Network

Recently, several research efforts have been made to achieve accurate failure localization in optical networks, including rule-based reasoning with routing and alarms, as well as the analysis of alarms and monitoring data through machine learning (ML).

For rule-based reasoning, Delezoide *et al.* [12] localized failures through the analysis of routing matrices and alarm vectors based on monitoring data from coherent transponders. This work is an important reference for failure localization, but ambiguous cases are not well resolved, especially for complex topologies. Xin *et al.* [13] proposed a harmony search algorithm to achieve multi-faults localization, where it obtained the location accuracy of 97.5% within 0.94*s*. Yang *et al.* [14] presented a fault propagation model based on low-density check matrices to solve fault location problem, and this model was validated in the communication network of China Southern Grid. P. Vela *et al.* [15] focused on bit error rate (BER) degradation detection and failure identification in elastic optical networks. Wu *et al.* [16] demonstrated a failure management system, where power attenuation was monitored for locating failures.

For ML-based analysis, S. Mayer *et al.* [17] evaluated an artificial neural network (ANN)-based soft-failure localization framework with partial telemetry, where principal component analysis (PCA) was used to accelerate the training process. Lun *et al.* [18] combined ANN with Gaussian process regression (GPR) to localize the soft failure location and estimate anomaly value. Jia *et al.* [3] introduced a transformer-based alarm context-vectorization representation technique for alarm root cause identification and correlation analysis. Li *et al.* [19] presented a graph neural network (GNN) to perform relational reasoning on alarm knowledge graphs (KGs) and locate the network faults. Musumeci *et al.* [20] focused on domain adaptation and transfer learning for failure localization across different lightpaths with real optical signal-to-noise ratio (OSNR), and this work achieved satisfactory accuracy with the limited data on target domain. Zhang *et al.* [21, 22] proposed an attention-mechanism-based failure localization scheme, in which the average accuracy of 98.73% was obtained. Yang *et al.* [23] introduced a deep neural evolution network to extract deep-hidden fault features from massive collected alarm data, which achieved global search without the constraint of gradient. Lun *et al.* [24] proposed a generative adversarial network (GAN)-based soft failure detection and identification algorithm. Abdelli *et al.* [25] presented a novel method, combining a denoising convolutional autoencoder (DCAE) and a bidirectional long short-term memory (BiLSTM), to locate fiber faults. The approach was applied to noisy optical time-domain reflectometry (OTDR) signals and achieved a diagnostic accuracy of 96.7%. Babbar *et al.* [26] compared several ML models to classify alarms and localize the source of failures.

B. Motivation and Our Contributions

Although these previous works have achieved significant improvement about failure localization, they were mainly investigating the failure localization at node-level (e.g., a ROADM node or a fiber link) rather than a more fine-grained level. As mentioned above, these previous methods will increase the pressure on operations and decrease the utilization efficiency of optical network resources. This paper addresses the problem of failure localization of intra-/inter-node. The main and new contributions are summarized as follows.

1) This work studies the failure localization of intra-/inter-node and analyses the potential failure scenarios for multi-degree ROADM-based optical networks. In addition, we redefine a new failure model and analyze the influence of different devices.

2) We propose a **S**uspect **F**ault **S**creen assisted **G**raph agg**R**egation **N**etwork (SFS-GRN) to locate failures of intra-/inter-node. SFS is used to analyze routing and wavelength assignment (RWA) results for all services and screen out potential fault devices from all devices. GRN is responsible for exploring monitoring data collected from OPM and determining the most likely failure location between potential fault devices.

3) The proposed scheme is evaluated in a 9-node simulation network and a 3-node testbed network, respectively. Extensive simulation and experimental results show that our scheme achieves higher localization accuracy under different percentages of OPM deployment. It also demonstrates that the proposed solution has no strong correlation with the types of failures and the number of service requests. Moreover, SFS-GRN can obtain reliable results within an acceptable time.

In this paper, we extend the previous work presented at the European Conference on Optical Communication (ECOC) in 2022 [27]. The rest of this paper is organized as follows. Section 2 discusses the network scenario and failure model of intra-/inter-node for the addressed problem. Section 3 details the procedure of SFS-GRN. In Section 4, the experimental setup and results are provided and discussed. Finally, Section 5 concludes this paper.

2. PROBLEM STATEMENT

In this section, we firstly introduce the network scenario, including multi-degree ROADMs, bidirectional links, service requests and OPM deployment. Moreover, we redefine a new failure model of intra-/inter-node for this network scenario.

A. Network Scenario

We present an ROADM-based optical network in Fig. 1(a), where several service requests are carried between different nodes and links. As shown in Fig. 1(b), each ROADM is designed with broadcast-and-selected (B&S) architecture, and it is composed of $1 \times K$ WSS, EDFA, AWG and splitter for traffic exchanging on

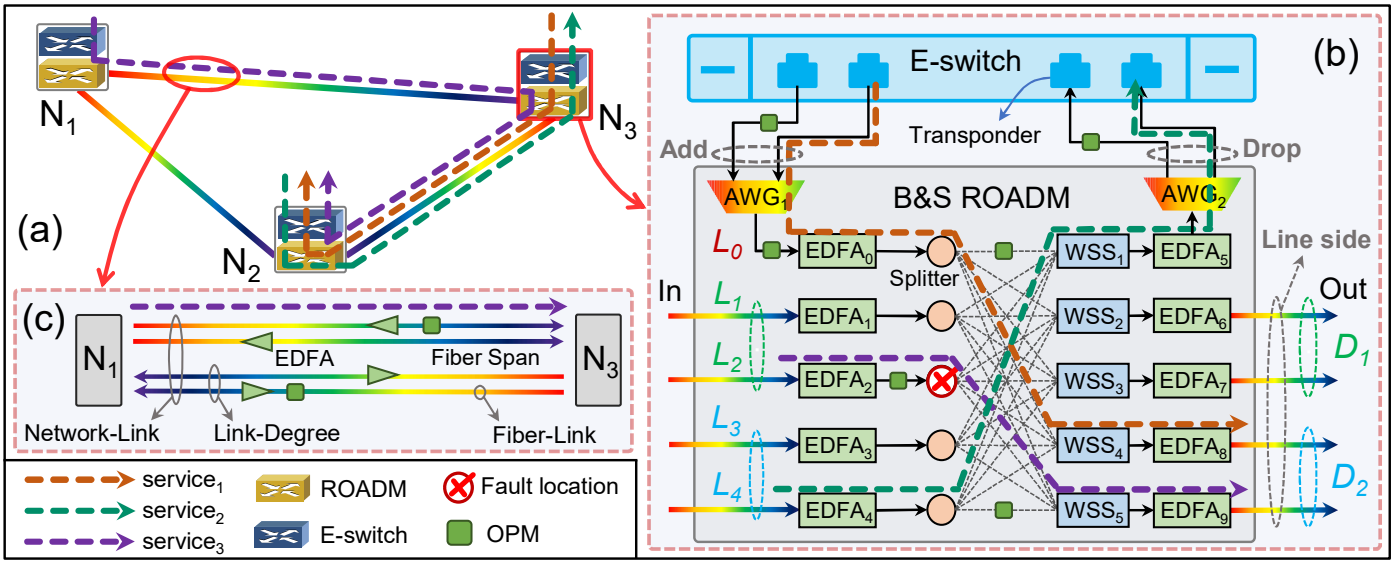

Fig. 1. Network scenario: (a) ROADM-based optical network; (b) multi-degree B&S ROADM and E-switch in each network node; (c) bidirectional link between two network nodes.

optical domain, where K denotes the size of WSS and its common values are limited to 9, 25, 32, etc. [28]. In the line side, the ROADM achieves flexible wavelength-based switching between different link-degrees (e.g., $L_1 \sim L_2$ or $L_3 \sim L_4$) and different direction-degrees (e.g., $D_1 \sim D_2$). The link-degree denotes the number of single-direction fibers between two network nodes, while the direction-degree indicates the number of other nodes connected with this node. In the add/drop side, it connects with an electrical switch (E-switch) to achieve traffic switching and grooming, and this side is considered as an extra link-degree, i.e., L_0 shown in Fig. 1(b). Besides, the E-switch is equipped with several transponders for sending and receiving traffic. Note that there is internal contention for same wavelengths in add/drop side, i.e., wavelength blocking [29]. Several methods can address this issue, such as adding optical transform unit (OTU) or $M \times N$ WSS [10]. However, how to achieve contentionless is out of the scope of our paper, and we assume the same wavelengths can be separated in add/drop side. For each network link, bidirectional links are provided for traffic transmission, and the number of fiber-links in each single-direction link equals link-degree between two adjacent nodes. Each fiber-link includes two fiber spans and one EDFA shown in Fig. 1(c). Meanwhile, a triple is used to represent each service request, i.e., [*src*, *dest*, *band*], where *src* and *dest* indicate source/destination pairs, while *band* denotes bandwidth demand.

As shown in Fig. 2(a), we consider both sides of each device as candidate positions for OPM deployment. In this condition, the number of all available positions is positively correlated with the number of devices, and it can be calculated as:

$$N_{opm} = \sum_{i,j\in\mathcal{N},i\neq j} \left(4 \cdot H_{i,j}\right) + \sum_{i\in\mathcal{N}} \left[2 \cdot T_i + 2 + 2 \cdot \left(1 + \sum_{d\in D_i} L_{i,d}\right) + \left(1 + \sum_{d\in D_i} L_{i,d}\right)^2\right] \quad (1)$$

where N_{opm} denotes total candidate positions, $\mathcal{N}$ indicates the set of network nodes, $H_{i,j}$ denotes the number of fiber-links

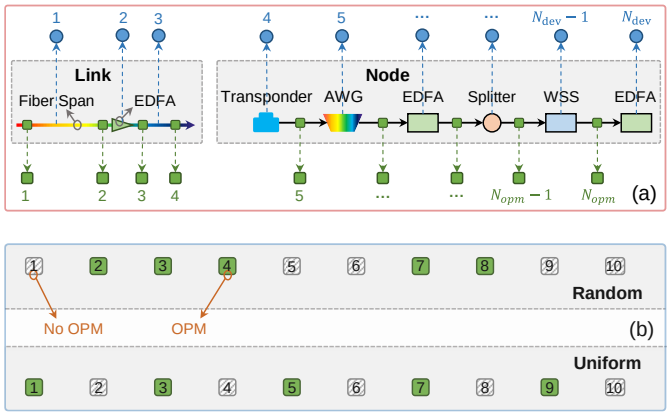

Fig. 2. (a) Total OPM and devices; (b) example about random and uniform OPM deployment.

(i.e., link-degree) between i^{th} node and j^{th} node, T_i denotes the number of transponders in i^{th} node, D_i represents the direction-degree in i^{th} node and $L_{i,d}$ denotes the link-degree in d^{th} direction of i^{th} node. In Equation (1), "4" denotes the number of candidate positions in each fiber-link, while $2 \cdot T_i$, 2, $1 + \sum_{d\in D_i} L_{i,d}$, $(1 + \sum_{d\in D_i} L_{i,d})^2$ and $1 + \sum_{d\in D_i} L_{i,d}$ denote the number of candidate positions between transponder, AWG, EDFA, splitter, WSS and EDFA in i^{th} node, respectively. Equation (1) shows that it is a large candidate space for OPM placement. Thus, how to select partial positions for OPM deployment is essential to failure localization, and we consider two different ways to place OPM. The detailed steps are introduced as follows. Firstly, the total candidate positions of OPM (i.e., N_{opm}) are encoded through a list. Then, we place OPM according to different deployment strategies and percentages. For random way, we randomly select a certain number of candidate positions for placing OPM. For uniform way, the OPM are placed at fixed intervals. Fig. 2(b) shows an example about how to deploy OPM with 50% OPM, where the random way selects five positions (i.e., 2-4 and 7-8) to place OPM, while the uniform way selects the odd positions to

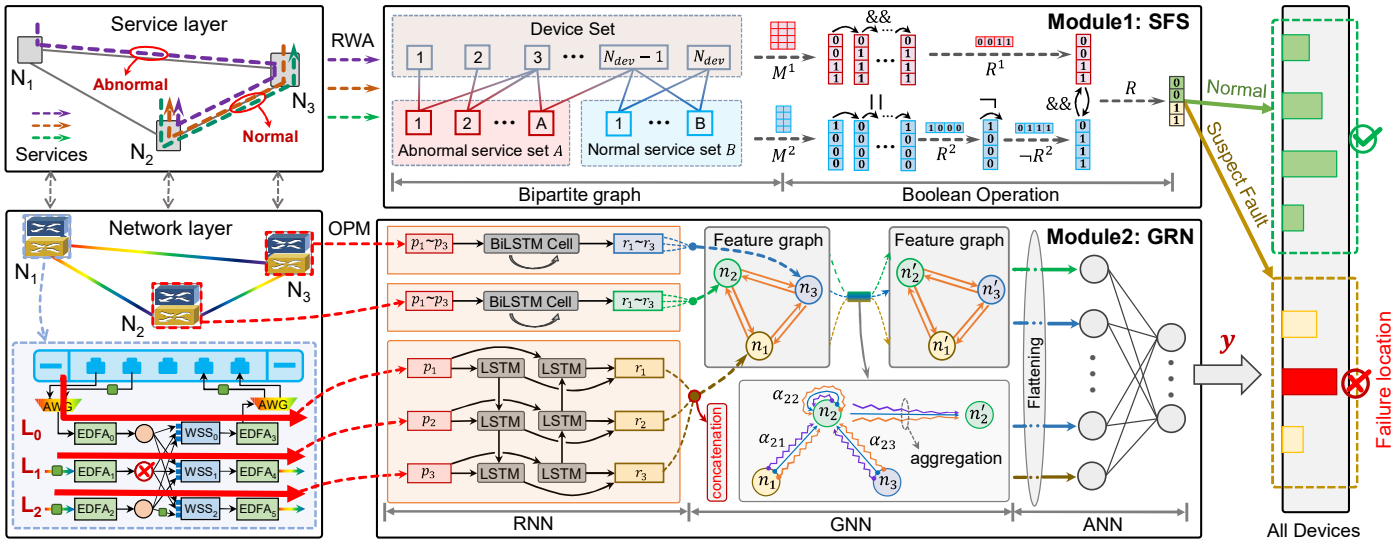

Fig. 3. Intra-/inter-node failure localization scheme based on SFS-GRN.

place OPM.

Table 1. Failure Devices of Intra-/Inter-Node and Corresponding Influences

Types	Devices	Influences
Amplification	EDFA	Insufficient amplification
Attenuation	Fiber span Splitter WSS AWG	Extra attenuation
Launch	Transponder	Insufficient launch power

B. Redefinition of Intra-/Inter-Node Failure Model

In previous failure models, network topology is an important criterion for failure classification and the faults are usually divided into nodes and links. However, the existing failure models are not suitable for multi-degree ROADM-based optical networks since the intra-/inter-node contains many devices. In this work, we redefine a failure model according to their function of different devices.

The redefined failure model is shown in Table 1, where all failures are divided into three categories, i.e., amplification, attenuation and launch failures. Amplification failure includes all EDFAs of intra-/inter-node, and they will provide insufficient amplification to optical signals. Attenuation failure contains all WSS, AWG, splitter, and fiber spans, where they bring extra attenuation for optical signals. Launch failure includes all transponders, and they cannot ensure sufficient launch power. The redefined failure model shows the impact of different failures, which is beneficial to failure localization. In addition, Fig. 2(a) shows how to calculate the total devices (i.e., potential failures). For each fiber-link, it consists of three devices and the number of devices in total links is $\sum_{i\neq j}(3\cdot H_{i,j})$. For each ROADM-node, the number of transponders and AWGs is $2\cdot T_i$ and 2, while other devices are $1+\sum_{d\in D_i} L_i, d$, respectively. Thus, the number of total devices is calculated as follows.

$$N_{dev}=\sum_{i,j\in\mathcal{N},i\neq j}\left(3\cdot H_{i,j}\right)+\sum_{i\in\mathcal{N}}\left[2\cdot T_i+2+4\cdot\left(1+\sum_{d\in D_i}L_{i,d}\right)\right] \quad (2)$$

where N_{dev} denotes total number of devices and other variables are consistent with Equation (1). In addition, multiple devices usually do not fail simultaneously. Thus, we focus on single-failure localization.

3. SUSPECT FAULT SCREEN ASSISTED GRAPH AGGREGATION NETWORK

In this section, we propose a collaborative scheme for intra-/inter-node failure localization. The whole scheme consists of two modules shown in Fig. 3.

A. SFS for Reducing Candidate Devices

Equation (2) shows that failure localization of intra-/inter-node will bring a large candidate space. Thus, we present a heuristic scheme to analyze RWA results for reducing the number of candidate devices. This module is introduced in detail.

We establish a bipartite graph between all service requests and devices based on RWA results, in which each link indicates a service request passing through a corresponding device. For example, link 2-3 denotes the second service request passing through third device. Then, total service requests are divided into abnormal service set A and normal service set B according to whether they have alarms or service interruption. Meanwhile, these links are further transformed into two connection matrices, where M^1 denotes the links between abnormal service set A and device set, while M^2 denotes the links between normal service set B and device set.

Moreover, several boolean operations are designed to distinguish between all devices that are normal or suspect faulty devices. This design is based on two principles: 1) failure is one of devices through which all abnormal services pass; 2) failure is not a device that normal services pass through. These operations can be expressed as follows:

$$R^1 = M^1_{:,1} \&\& M^1_{:,j} (\forall 1 < j \leq |A|) \tag{3}$$

$$R^2 = M^2_{:,1} || M^2_{:,j} (\forall 1 < j \leq |B|) \tag{4}$$

$$R = R^1_i \&\& (\neg R^2_i) (\forall 1 \leq i \leq N_{dev}) \tag{5}$$

where &&, || and ¬ denote logical "AND", "OR" and "NOT" operators, $|A|$ and $|B|$ indicate the size of abnormal service set A and normal service set B, N_{dev} indicates the total number of devices, while Equation (3) and (4) correspond to principle 1 and 2, respectively. The final output R is a N_{dev}-dimensional boolean vector, in which "1" and "0" indicate whether each device is suspect fault or normality, respectively.

B. GRN for Locating Failure Devices

After screening by SFS, the failure location will be limited to a small space, but it is usually unable to directly find a specific failure location. Therefore, GRN is designed to explore monitoring data from OPM and further locate the failure device. The module consists of recurrent neural network (RNN), graph neural network (GNN) and ANN. We describe the proposed scheme as follows.

B.1. RNN-based Aggregation in Node-Wide

Firstly, these monitoring results of each network node and its adjacent links are split into multiple vectors, e.g., $p_1 \sim p_3$ in Fig. 3, where the number of vectors depends on the link-degree and direction-degree of this ROADM, i.e., $1 + \sum_{d \in D_i} L_{i,d}$. There is complex sequence-dependency among multiple vectors due to traffic exchanging between different link-degrees. Meanwhile, the RNN model is an effective method for series-dependent processing. Thus, RNN is used for aggregating and analyzing these monitoring data in node-level. Specifically, we take BiLSTM to achieve monitoring data aggregation in double directions due to its benefits in sequence analysis [30, 31].

For each LSTM cell, it is composed of a forget gate, input gate and output gate. The forget gate determines how much information of the previous state can be reused to the current state, and the calculation formula is as follows:

$$f_t = \sigma(w_f \cdot [h_{t-1}, p_t] + b_f) \tag{6}$$

where f_t represents how much information is forgotten ($f_t \in [0,1]$), $\sigma(x) = 1/(1+e^{-x})$ is an activation function, $w_f \in \mathcal{C}^{H_t \times (P_t + H_{t-1})}$ is a weight matrix, $b_f \in \mathcal{C}^{H_t}$ is a bias vector, h_{t-1} denotes the output at the previous sequence, p_t is an input vector for current sequence, while P_t, H_t and H_{t-1} denote the size of sequence p_t, h_t and h_{t-1}.

The input gate determines how much information in current sequence p_t can be stored in neurons and it consists of two parts. The first part is to obtain the input weight i_t according to the previous hidden sequence h_{t-1} and the current input sequence p_t, where z_t indicates new knowledge learned from the current sequence p_t. The second part is to integrate the old information $f_t * c_{t-1}$ and new information $i_t * z_t$ for updating the current state c_t. The calculation formulas are as follows:

$$i_t = \sigma\left(w_i \cdot [h_{t-1}, p_t] + b_i\right) \tag{7}$$

$$z_t = \tanh(w_z \cdot [h_{t-1}, p_t] + b_z) \tag{8}$$

$$c_t = f_t * c_{t-1} + i_t * z_t \tag{9}$$

where $w_i \in \mathcal{C}^{H_t \times (P_t + H_{t-1})}$ and $w_z \in \mathcal{C}^{H_t \times (P_t + H_{t-1})}$ are weight matrices, $b_i \in \mathcal{C}^{H_t}$ and $b_z \in \mathcal{C}^{H_t}$ are bias vectors, $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ is an activation function, while $*$ denotes Hadamard product.

The output gate is used to control the output h_t and transfer information to the next hidden neurons. It firstly obtains the output weight o_t by combining the previous hidden state h_{t-1} and the current input p_t, and then generates a new output h_t.

$$o_t = \sigma(w_o \cdot [h_{t-1}, p_t] + b_o) \tag{10}$$

$$h_t = o_t * \tanh(c_t) \tag{11}$$

where $w_o \in \mathcal{C}^{H_t \times (P_t + H_{t-1})}$ and $b_o \in \mathcal{C}^{H_t}$ are weight matrix and bias vector for output gate, respectively.

The BiLSTM network constructs two LSTM layers in opposite directions for each network node and its adjacent links. The forward $\overrightarrow{h_{i,t}}$ starts from the first sequence and stops at the final sequence, while the backward $\overleftarrow{h_{i,t}}$ starts from the end sequence until the starting sequence. The hidden state in both directions is merged into the output $r_{i,t}$.

$$\begin{aligned} \overrightarrow{h_{i,t}} &= \overrightarrow{LSTM}(p_{i,t}, h_{i,t-1}) \\ \overleftarrow{h_{i,t}} &= \overleftarrow{LSTM}(p_{i,t}, h_{i,t+1}) \end{aligned} \tag{12}$$

$$r_{i,t} = \overrightarrow{h_{i,t}} + \overleftarrow{h_{i,t}} \tag{13}$$

where $r_{i,t}$ denotes the BiLSTM output for t^{th} sequence of i^{th} network node.

Finally, we flatten all BiLSTM outputs of i^{th} network node as a network node's features:

$$n_i = [r_{i,1}, r_{i,2}, ..., r_{i,t}] \tag{14}$$

where $n_i \in \mathcal{C}^{L^i H_t}$ denotes the features of i^{th} network node. These node features constitute a feature graph for further aggregation and analysis.

B.2. GNN-based Aggregation in Network-Wide

There is complex dependency between different network nodes (i.e., network-level) because of flexible routing. Thus, we take a variant of GNN model [i.e., graph attention network (GAT)] to achieve feature aggregating and analysis [32].

Consider a graph with $\mathcal{N}$ nodes, let $\mathbf{n} = \{n_1, n_2, ..., n_{\mathcal{N}}\}$, $n_i \in \mathcal{C}^F$ be a set of node features, where $F\ (= L^i H_t)$ is the number of features in each node. The attention coefficients e_{ij} can be calculated as:

$$e_{ij} = a(w \cdot n_i, w \cdot n_j) \tag{15}$$

where e_{ij} indicates the importance of j^{th} node to i^{th} node, $a(\cdot)$ is a single-layer feedforward neural network, and $w \in \mathcal{C}^{F \times F'}$ is a shared weight matrix.

To make attention coefficients easily comparable across different nodes, e_{ij} can be normalized by the softmax function:

$$\alpha_{ij} = \frac{exp[\varphi(e_{ij})]}{\sum_{k \in \mathcal{N}_i} exp\left[\varphi(e_{ik})\right]} \tag{16}$$

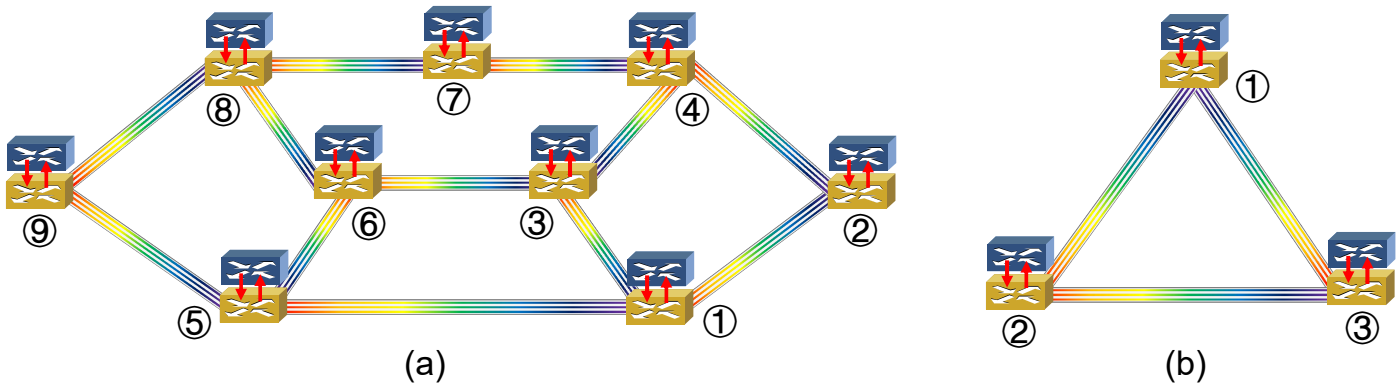

Fig. 4. Evaluation networks: (a) simulation topology for a 9-node network; (b) testbed topology for a 3-node network.

where α_{ij} indicates the normalized attention coefficients, $\varphi(\cdot)$ is a LeakyReLU function, and $\mathcal{N}_i$ is some neighborhood of i^{th} node.

The normalized attention coefficients are used to compute the output features for each node, where the multi-head attention is used to stabilize the learning process. The aggregation operation is defined as follows:

$$n_i^{'} = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j\in\mathcal{N}_i} \alpha_{ij}^{k} \cdot w^{k} \cdot n_j\right) \tag{17}$$

where $n_i^{'} \in \mathcal{C}^{F^{'}}$ denotes the features of next layer of i^{th} node, K is the number of multi-head attention, α_{ij}^{k} is the normalized attention coefficient computed by the k^{th} attention mechanism $a^{k}(\cdot)$, and w^{k} is the corresponding weight matrix.

B.3. ANN-based Failure Localization

After aggregating in node- and network-wide, these features are flattened into a vector g for failure classification. The flatting operation is designed as follows:

$$g = [n_1^{'}, n_2^{'}, ..., n_{\mathcal{N}}^{'}] \tag{18}$$

In this work, we take ANN to analyze these features and locate a failure between suspect fault devices. The formula is as follows:

$$y = \mu(w_a \cdot g + b_a) \tag{19}$$

where y denotes the failure localization results, $\mu(\cdot)$ is a softmax function for normalizing output results, while $w_a \in \mathcal{C}^{N_{dev}\times|\mathcal{N}|F^{'}}$ and $b_a \in \mathcal{C}^{N_{dev}}$ represent weight matrix and bias vector, respectively.

C. Complexity Analysis

SFS-GRN consists of two modules and their time complexity can be analyzed independently.

SFS is a heuristic scheme, and its complexity comes from boolean operators. Equation (3) will be repeated for $|A|-1$ for each device and its complexity is $O[N_{dev}\cdot(|A|-1)]$. Similarly, the complexity of equation (4) is $O[N_{dev}\cdot(|B|-1)]$. Equation (5) will operate local "AND" and "NOT" for a vector with size of N_{dev}, and its complexity is $O(2N_{dev})$. Therefore, the total complexity of SFS is $O[N_{dev}\cdot(|A|+|B|)]$.

Suppose $\mathbf{A} \in \mathcal{C}^{M\times N}$, $\mathbf{B} \in \mathcal{C}^{N\times L}$ and $\mathbf{C} \in \mathcal{C}^{M\times N}$ are arbitrary matrices, while $\mathbf{b} \in \mathcal{C}^{N}$ is an arbitrary vector. Thus, the required number of floating-point operations of **AB** (i.e., matrix-matrix product), **Ab** (i.e., matrix-vector product), **A*C** (i.e., Hadamard product), $\sigma(\mathbf{b})$, $\tanh(\mathbf{b})$ and LeakyReLU($\mathbf{b}$) are $O(2MNL-ML)$, $O(2MN-M)$, $O(MN)$, $O(4N)$, $O(9N)$ and $O(N)$, respectively [4, 33]. We will analyze the complexity of GRN based on the above formulas. Equation (6), (7) and (10) all contain a matrix-matrix product (e.g., $w_f \cdot [h_{t-1}, p_t]$) and an activation function $\sigma(x)$, and every complexity is $O[L^i\cdot(2H_t\cdot(P_t+H_{t-1})+3H_t)]$, where $L^i = 1+\sum_{d\in D_i} L_{i,d}$ denotes the link-degree of i^{th} node. Similarly, the complexity of equation (8) is $O[L^i\cdot(2H_t\cdot(P_t+H_{t-1})+8H_t)]$. Equation (9) consists of two Hadamard products and its complexity is $O(2L^iH_t)$, while Equation (11) is $O(10L^iH_t)$. Equation (12) and (13) represent the operations of BiLSTM and they are twice as complex as LSTM. Equation (15) consists of two matrix-matrix products and a feedforward neural network, and its complexity is $O[K|\mathcal{N}_i|\cdot(2FF^{'}+3F^{'}-1)]$, where $|\mathcal{N}_i|$ denotes the number of neighborhoods of i^{th} node. Equation (16) contains a LeakyReLU and a softmax function, and its complexity is $O[K\cdot(2|\mathcal{N}_i|+2)]$. Equation (17) defines the aggregation operation and its complexity is $O[|\mathcal{N}_i|\cdot(2FF^{'}-F^{'}+1)+4F^{'}]$. Equation (19) is an ANN layer and its complexity is $O[2\cdot N_{dev}|\mathcal{N}|F^{'}-|\mathcal{N}|F^{'}+N_{dev}+(1+N_{dev})^2]$. Thus, the total complexity of GRN is $O[\sum_{i\in\mathcal{N}}(2L^iH_t\cdot(8(P_t+H_{t-1})+29)+K|\mathcal{N}_i|\cdot(2FF^{'}+3F^{'}+1)+2K+|\mathcal{N}_i|\cdot(2FF^{'}-F^{'}+1)+4F^{'})+|\mathcal{N}|F^{'}\cdot(2N_{dev}-1)+N_{dev}+(1+N_{dev})^2]$.

4. NUMERICAL RESULTS AND DISCUSSION

In this section, we first present our experimental setup, including simulation and testbed networks. In addition, we take the existing ANN-based method as the benchmark algorithms and then compare SFS-GRN with ANN and GRN in terms of localization accuracy and execution time under different number of OPM deployment, service requests, failure types and samples.

A. Simulation and Experimental Setup

A.1. Simulation Scenario

Fig. 4(a) shows a 9-node simulated network, which consists of nine network nodes and twelve bidirectional links. For each network node, the E-switch contains three transponders, and the launch power of each transponder is -1 dBm. The gain of each internal EDFA is 15 dB to amplify signal power. Besides, the insertion loss of each WSS, Splitter and AWG are set as 6 dB, 2 dB and 6 dB, respectively. The number of WSSs in each direction-degree (i.e., link-degree) is set as 1 or 5. For each single-direction

Table 2. Simulation and Experimental Parameters

Parameters	Values
Attenuation of fiber	0.2 dB/km
Amplifier gain of EDFA in intra-node	15 dB
Amplifier gain of EDFA in inter-node	20 dB
Bandwidth demand of each service	Real dataset
Capacity of each wavelength	10 $Gbps$
Insertion loss of each WSS	6 dB
Insertion loss of each AWG	6 dB
Insertion loss of each Splitter	2 dB
Length of each fiber span	$20 \sim 60$ km
Launch power of each transponder	-1 dBm
Number of wavelengths in each fiber	3
Number of services in simulation	$20 \sim 100$
Number of services in testbed	5
Number of WSSs in each direction-degree	1 or 5
Source/destination pairs	Random
Threshold of amplification failures	$0 \sim 50\%$
Threshold of attenuation failures	$150\% \sim 200\%$
Threshold of launch failures	$0 \sim 50\%$

network link, it includes several fiber-links, where the number of fiber-links corresponds to the link-degree of neighboring nodes. These fiber spans range from 20 km to 60 km with 0.2 dB/km attenuation, while the amplifier gain of each inter-node EDFA is set as 20 dB to withstand fiber attenuation. Three wavelengths with total capacity of 3×10 $Gbps$ are used to accommodate service requests. In addition, we give three thresholds to simulate network faults. For amplification and launch devices, it is considered as a failure if their parameters reach $0 \sim 50\%$ of normal values. For attenuation devices, the abnormal parameters range from 150% to 200% of normal values. Meanwhile, we generate $20 \sim 100$ service requests with random source/destination pairs, while their bandwidth demands are collected from real geographically distributed areas [34]. Auxiliary graph (AG) model with minimizing the number of wavelength-links (MinWL) performs RWA for each service request [35, 36]. In our simulation, we ignore the nonlinear effect of optical fibers and insertion loss of connectors.

A.2. Experimental Scenario

We also developed a 3-node testbed network to further validate our scheme shown in Fig. 4(b). It is a ROADM-based testbed presented in our previous work [34], and the experimental scenario is shown in Fig. 5. The traffic generator and analyzer (TGA) is connected to E-switch for injecting live traffic, while variable optical attenuator (VOA) is used to simulate failures. These real data are collected from OPM and reported to network management system (NMS) for validating our schemes. In our experiment, the number of service requests is set as 5, while other parameters are similar with simulation setup. The above parameters are summarized in Table 2.

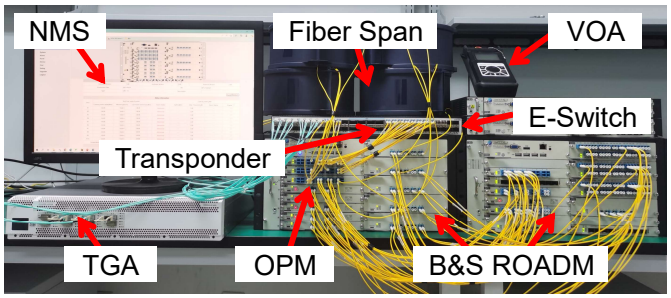

Fig. 5. Testbed of ROADM-based optical networks.

Table 3. Structure of Training and Testing Samples

Service layer	Network layer	Failure index
$\{R_{1,1}, R_{1,2}, ..., R_{1,J}\}$	$\{P_{1,1}, P_{1,2}, ..., P_{1,N_{opm}}\}$	$1 \sim N_{dev}$
$\{R_{2,1}, R_{2,2}, ..., R_{2,J}\}$	$\{P_{2,1}, P_{2,2}, ..., P_{2,N_{opm}}\}$	$1 \sim N_{dev}$
...	...	...
$\{R_{S,1}, R_{S,2}, ..., R_{S,J}\}$	$\{P_{S,1}, P_{S,2}, ..., P_{S,N_{opm}}\}$	$1 \sim N_{dev}$

A.3. Samples and Algorithms Parameters

As shown in Table 3, we design a data structure to describe our training and testing samples. For each sample, it consists of three parts. 1) Service layer includes all routing results, and it corresponds to the bipartite graph in Fig. 3, where S denotes the total number of samples, J indicates the number of service requests, and $R_{i,j}$ describes the routing results for j^{th} service request in i^{th} sample. Specially, $R_{i,j}$ can be represented through a vector, i.e., $R_{i,j} = [r_1^{i,j}, r_2^{i,j}, ..., r_{N_{dev}}^{i,j}, F^{i,j}]$, where $r_n^{i,j}$ is a boolean variable denoting whether the j^{th} service request passes through n^{th} device in i^{th} sample, while $F^{i,j}$ denotes whether the j^{th} service request has alarm or service interruption in i^{th} sample. 2) Network layer presents all OPM monitoring results, and each sample can be indicated by a vector, i.e., $[P_{i,1}, P_{i,2}, ..., P_{i,N_{opm}}]$, where $P_{i,j}$ denotes the monitoring result of j^{th} OPM in i^{th} sample, and N_{opm} denotes the total number of OPM. 3) Failure index ranges from 1 to N_{dev} and denotes the failure location. It is further represented by one-hot encoding.

Meanwhile, we present the relevant parameters of SFS-GRN shown in Table 4. For feature aggregation in node-wide, we first develop a one-layer BiLSTM with $L_{max} \times 15$ neurons, where $L_{max} = max\{|L_{i,d}|\}, d \in D_i, i \in \mathcal{N}$ denotes the maximum number of devices in all link-degrees. These output results are concatenated into $[n_1, n_2, ..., n_{\mathcal{N}}]$ and further exploited by a GAT model. The neurons of GAT model are set as $N_{max} \times 60$, where $N_{max} = max\{|n_i|\}, i \in \mathcal{N}$ indicates the longest sequence of all node's features in Equation (14). Then, the outputs of GAT model are flattened into ANN with $N_{ann} \times N_{dev}$ neurons, where N_{dev} denotes the number of devices, while N_{ann} is the product of the number of network nodes (i.e. 3 or 9), the number of multi-head attention (i.e., $K = 3$) and the number of output neurons in GAT (i.e., 60). In our experiment, the learning rate, batch size and training epoch are set as 0.0001, 256 and 1000, respectively. The total samples (i.e., S) are 5000 and 250 for simulation and testbed, respectively.

The training samples account for 80% of total samples, while the remaining samples are used for testing. Thus, training sam-

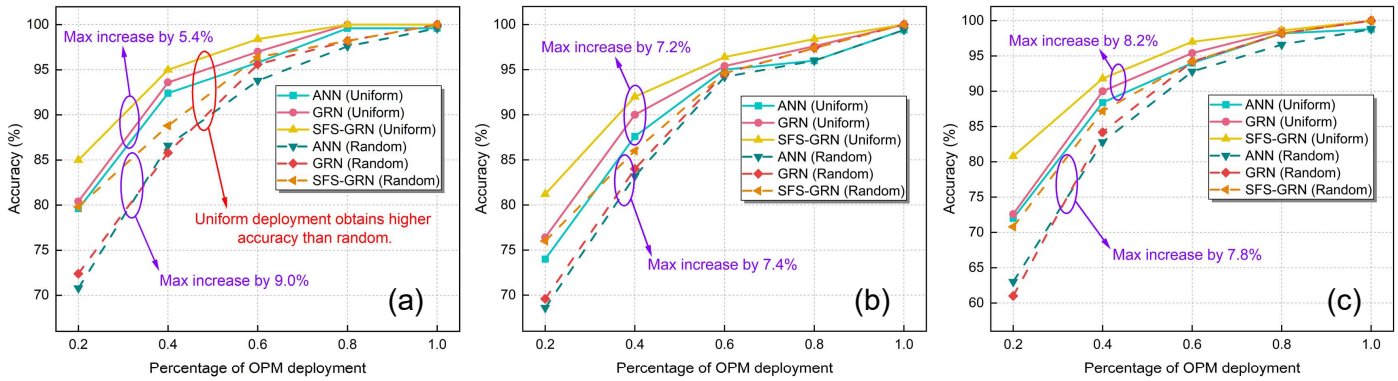

Fig. 6. Percentage of OPM deployment vs. localization accuracy under different number of service requests: (a) Number of service requests is 20; (b) Number of service requests is 40; (c) Number of service requests is 60.

Table 4. Experimental Parameters of SFS-GRN

Parameters	Values
Batch size	256
Learning rate	0.0001
Layers and neurons in BiLSTM	$L_{max} \times 15$
Layers and neurons in GAT	$N_{max} \times 60$
Layers and neurons in ANN	$N_{ann} \times N_{dev}$
Number of training epoch	1000
Number of multi-head attention, i.e, K	3
Number of training samples in simulation	4000
Number of testing samples in simulation	1000
Number of training samples in testbed	200
Number of testing samples in testbed	50

ples in simulation and testbed are 4000 and 200, while testing samples are 1000 and 50, respectively. To show the effectiveness of SFS-GRN for failure localization, we present GRN and ANN for comparison. GRN strategy will remove SFS module and others are consistent with SFS-GRN. ANN is composed of fully connected neural networks with $N_{opm} \times 2000 \times 1000 \times N_{dev}$ neurons, where N_{opm} and N_{dev} denote the number of OPM candidate positions and devices, respectively. The above schemes and data generating are implemented using Python and PyTorch.

B. Simulation Results and Analysis

In this part, we present the simulation results under different settings and discuss the impact of different factors on failure localization.

B.1. Localization Accuracy under Different Percentage of OPM Deployment

We first evaluate the localization accuracy with the increasing of OPM deployment in a 9-node simulation network [i.e., Fig. 4(a)]. In this simulation, OPM are placed in a random or uniform way, and the link-degree is 1, i.e., total number of WSSs in each network node is 3 or 4 in Fig. 4(a). Simulation results are shown in Fig. 6.

Fig. 6(a) shows the performance comparisons under different percentage of OPM deployment, where the number of service requests is set as 20. It can be observed that SFS-GRN achieves higher localization accuracy followed by GRN and ANN, in which the biggest increases are 9.0% and 5.4%, respectively. In addition, uniform deployment obtains higher localization accuracy than random. This is because random placement may cause OPM to be centrally placed on a few network nodes and links, which is not beneficial to feature exploration and fault localization. Thus, it is also an effective way to improve localization accuracy for choosing an appropriate OPM placement way. Meanwhile, Fig. 6(b) and Fig. 6(c) present localization accuracy when the number of service requests are 40 and 60, respectively. The results show that the proposed algorithm can obtain higher localization accuracy regardless of the number of service requests. This is beneficial to fault localization because the number of service requests is changing in a real network. The above results suggest that the proposed SFS-GRN has more potential than traditional schemes in failure localization of intra-/inter-node.

B.2. Localization Accuracy under Different Number of Service Requests

To show generality of the proposed strategy, we compare SFS-GRN with GRN and ANN under different number of service requests, where OPM are uniformly deployed, while other simulation parameters are same with Section B.1. The corresponding results are shown in Fig. 7.

Fig. 7(a) shows the variation trend of localization accuracy with the increasing of service requests, where the OPM ratio is 20%. The simulation results indicate that the proposed scheme improves localization accuracy by 9.2% at most. Moreover, it also shows that SFS-GRN has no strong correlation with the number of service requests. Similar results can be found in Fig. 7(b) and Fig. 7(c), where the percentage of OPM placement is 40% and 60%, respectively. We conclude that the proposed SFS-GRN has strong generality and can be applied to different service requests.

B.3. Localization Accuracy under High-Degree ROADM-based Optical Networks

As shown in Fig. 8, we compare the proposed strategy with GRN and ANN under high-degree ROADM-based optical networks, where the number of service requests is 20, OPMs are deployed with uniform way and the link-degree is 5, i.e., total number of WSSs in each network node is 11 or 16 in Fig. 4(a).

The simulation results show that the accuracy of all schemes

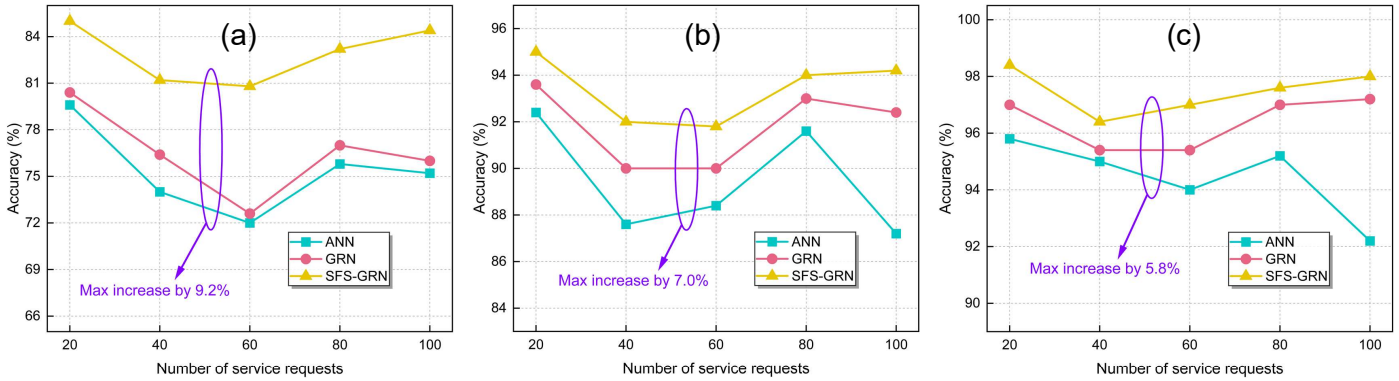

Fig. 7. Number of service requests vs. localization accuracy under different percentage of OPM deployment: (a) Percentage of OPM deployment is 20%; (b) Percentage of OPM deployment is 40%; (c) Percentage of OPM deployment is 60%.

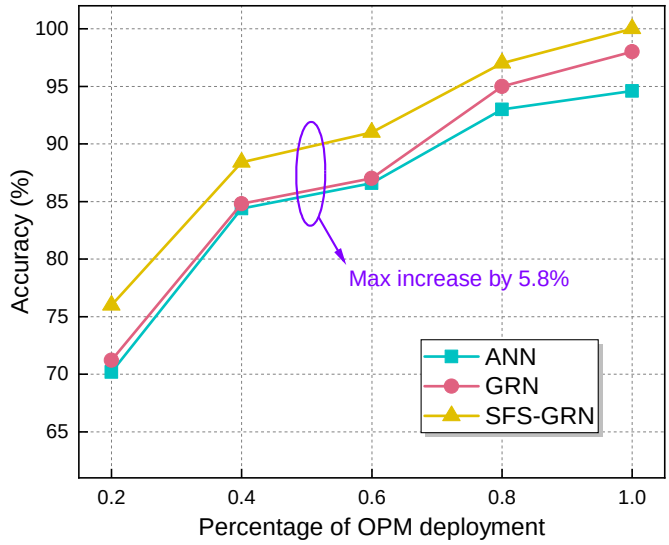

Fig. 8. Percentage of OPM deployment vs. localization accuracy under high-degree ROADM-based optical networks.

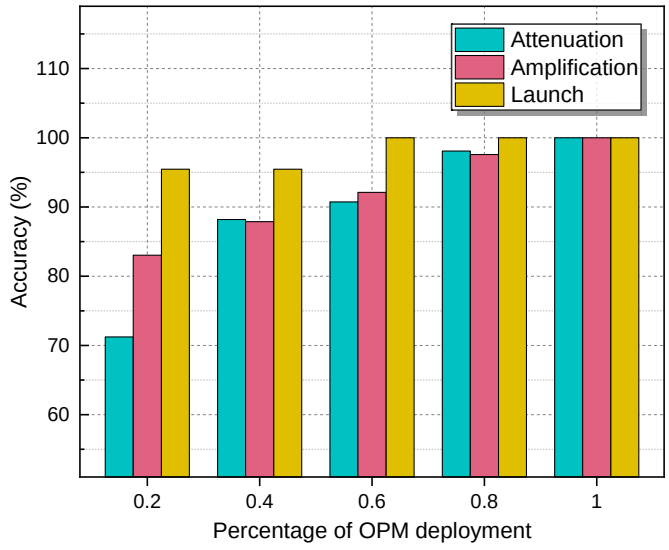

Fig. 9. Percentage of OPM deployment vs. localization accuracy of different failure types with SFS-GRN scheme.

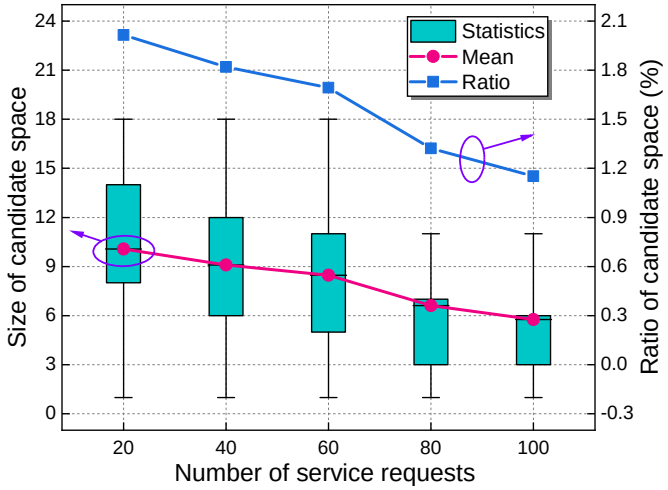

Fig. 10. Number of service requests vs. number and ratio of suspect fault devices with SFS.

is slightly decreased compared with the low-degrees [see Fig. 5(a)]. For instance, the accuracy of SFS-GRN dropped by about 5% when OPM ratio is 20%. This is because the network scenario of high-degrees becomes more complex than low-degrees, where the intra/inter-node contain more devices. Meanwhile, the proposed strategy also achieves a higher accuracy improvement, in which the biggest increase is 5.8%. The above results indicate that our scheme has the potential for achieving accurate failure localization in high-degree ROADM-based optical networks.

B.4. Localization Accuracy of Different Failure Types

We further present the localization accuracy of different failure types under different percentage of OPM deployment, where SFS-GRN is used to achieve failure localization, and all simulation parameters are same with Section B.3.

Fig. 9 presents the simulation results, where launch failures achieve higher accuracy followed by amplification and attenuation failures. This is because attenuation failures contain more devices (i.e., fiber spans, Splitter, WSS and AWG) than launch failures (i.e., transponders). The simulation results show that the localization accuracy of different failure types is positively correlated with the number of devices.

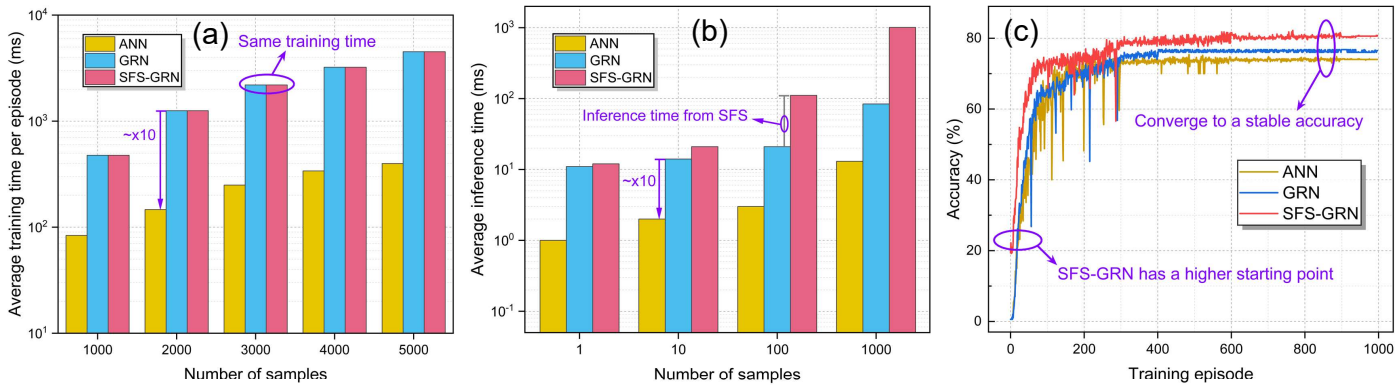

Fig. 11. (a) Average training time per episode versus number of samples; (b) Average inference time versus number of samples; (c) Localization accuracy versus number of training episodes.

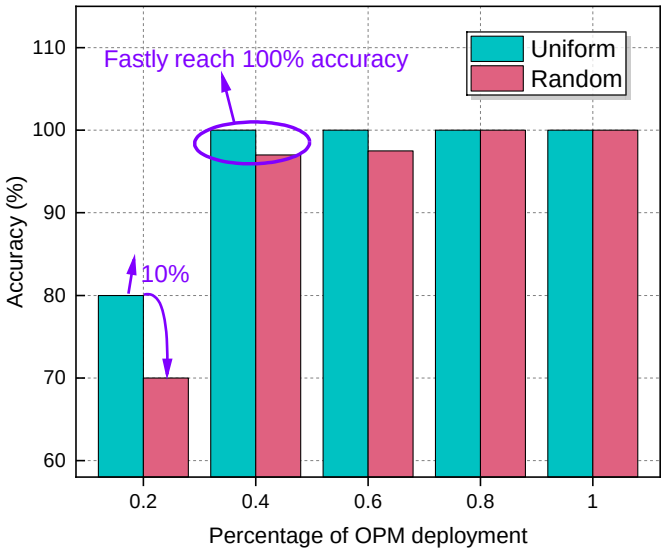

Fig. 12. Percentage of OPM deployment vs. localization accuracy in testbed networks with SFS-GRN.

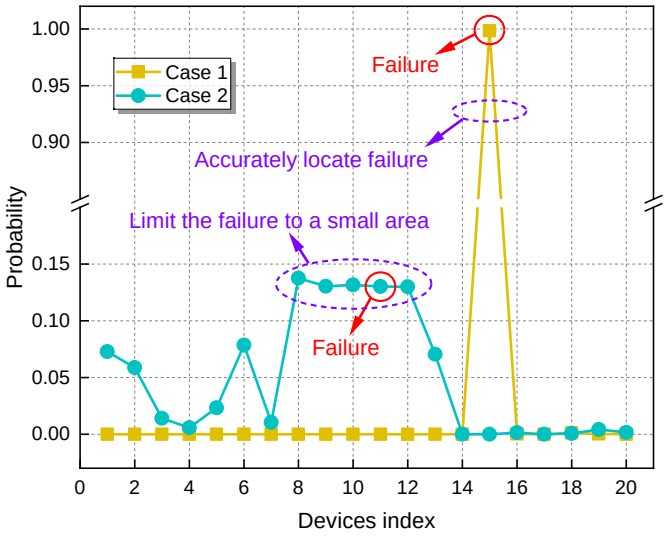

Fig. 13. Locating probability for different devices in testbed networks with SFS-GRN.

B.5. Number of Suspect Fault Devices under Different Number of Service Requests

We also present the number and distribution of suspect fault devices after SFS screening, and the simulation results are shown in Fig. 10.

The results show that the proposed SFS can remove more than 98% of devices under different number of service requests. For example, SFS can reduce the fault location to 1.69% of all devices when the number of service requests is 60. This result is helpful to further detect and repair since network operators only check the suspect fault devices (e.g., 1.69%) rather than all devices (i.e., 100%). It is an effective way to reduce the pressure on network operators. Meanwhile, we find that the distribution and number of suspect fault devices will decrease with the increasing of service requests. This is because routing results will become more diversified as the number of service request increases. The diversified routing is beneficial to analyze suspect fault devices.

B.6. Computing Time of Training and Inference

In this part, we first compare the average computing time of training and inference between different methods, where the number of services is 20, while the percentage of OPM ranges from 0.2 to 1.0 (step by 0.2). Fig. 11(a) shows the average training time per episode with the number of samples. The results show that the training time of GRN is about ten times that of ANN. This is because the updating of BiLSTM and GAT introduces additional training time. Meanwhile, this is obvious that GRN and SFS-GRN have the same training time because SFS is only used in inference period. Besides, Fig. 11(b) shows the average inference time under different number of samples, where ANN and GRN present the similar trend with training period, while SFS-GRN increases the extra inference time for screening out potential fault devices. Moreover, the inference time of the proposed scheme ranges from 10 *ms* to 1000 *ms*, and this is possible to run it in a real system. Fig. 11(c) shows the testing accuracy under different number of training episodes, where the number of service requests is 40 and the percentage of OPM is 20%. The results show that SFS assists GRN to obtain a higher starting point. This is because SFS can determine the failure location to a small range. Finally, these strategies converge to different stability accuracies.

C. Experimental Results and Analysis

In this part, we present the experimental results under different settings and discuss the impact on failure localization.

C.1. Localization Accuracy in Testbed Networks

We evaluate the proposed SFS-GRN in a 3-node testbed network [see Fig. 4(b)], and the experimental results are shown in Fig. 12. The experimental results show that uniform deployment is a more useful way to place OPM than random. For instance, it improves localization accuracy of 10% when the percentage of OPM deployment is 20%. Moreover, it achieves localization accuracy of 100% at low OPM ratio (i.e., 40%). The above results indicate that our scheme achieves similar performance in both simulated and real network scenarios.

C.2. Locating Probability in Testbed networks

Fig. 13 further shows the locating probability for different devices, where the OPM is uniformly deployed at 20%. The results present two cases: 1) The proposed scheme can accurately localize failure with close to 1.0 probability. 2) SFS-GRN can't locate the specific failure location, but it can limit the failure to a small area. The reason is that SFS analyzes the RWA results for suspecting potential fault devices from all devices. The above cases indicate that SFS-GRN first limits the fault location to a certain range and further determines the specific location, which improves its reliability.

5. CONCLUSIONS

In this work, we studied failure localization of intra-/inter-node in multi-degree ROADM-based optical networks. Different from the previous failure models, we redefined a failure model according to their function of different devices. Moreover, we proposed a collaborative scheme to achieve accurate failure localization, where SFS is used to reduce the number of failure locations through screening out potential fault devices from all devices, while GRN is responsible for determining the most likely failure location between potential fault devices. The proposed SFS-GRN was evaluated on a 9-node simulated network and a 3-node testbed network, respectively. Extensive simulation and experimental results showed that our scheme can achieve higher accuracy than existing strategies under different percentage of OPM deployment, different number of service requests and different failure types. This approach has the potential to support automated failure management for multi-degree ROADM-based optical networks.

In the future, we are planning to investigate how to deploy OPM for large-scale optical networks and balance the tradeoff between the number of OPM and localization accuracy. Meanwhile, we also extend our strategy to more complex ROADM architectures, such as colorless and directionless (CD) or colorless, directionless and contentionless (CDC)-ROADM. In this condition, more complex failure models and schemes should be designed to support these architectures. Finally, the fine-grained failure localization scheme depends on complicated network orchestrating and management. The detailed control procedure should be further studied to make it come true.

6. FUNDING

This work was supported by the National Key R&D Program of China (No. 2022YFB2903700), the National Nature Science Foundation of China Project (No. 61971055) and the BUPT Innovation and Entrepreneurship Support Program (2023-YC-T002).

REFERENCES

1. ETSI, "Fifth generation fixed network (f5g); f5g generation definition release 1," (2020). https://www.etsi.org/deliver/etsi_gr/F5G/001_099/001/01.01.01_60/gr_F5G001v010101p.pdf.
2. F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, and M. Tornatore, "A tutorial on machine learning for failure management in optical networks," J. Light. Technol. **37**, 4125–4139 (2019).
3. J. Jia, D. Wang, C. Zhang, H. Yang, L. Guan, X. Chen, and M. Zhang, "Transformer-based alarm context-vectorization representation for reliable alarm root cause identification in optical networks," in *2021 European Conference on Optical Communication (ECOC),* (2021), pp. 1–4.
4. R. Wang, J. Zhang, Z. Gu, S. Yan, Y. Xiao, and Y. Ji, "Edge-enhanced graph neural network for du-cu placement and lightpath provision in x-haul networks," J. Opt. Commun. Netw. **14**, 828–839 (2022).
5. R. Casellas, R. Martinez, R. Vilalta, R. Munoz, A. Gonzalez-Muniz, O. G. de Dios, and J.-P. Fernandez-Palacios, "Advances in sdn control and telemetry for beyond 100g disaggregated optical networks [invited]," J. Opt. Commun. Netw. **14**, C23–C37 (2022).
6. K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6g: Vision, enabling technologies, and applications," IEEE J. on Sel. Areas Commun. **40**, 5–36 (2022).
7. Z. Zhong, L. Lu, Y. Li, L. Zong, and G. Shen, "Colorless, directionless, and partially contentionless (cdpc) roadm: A new architecture for the best performance to cost ratio (invited)," in *2019 Asia Communications and Photonics Conference (ACP),* (2019), pp. 1–3.
8. D. C. Morão, L. G. Cancela, and J. L. Rebola, "Exploring future large-scale roadm architectures," in *2021 Telecoms Conference (ConfTELE),* (2021), pp. 1–6.
9. D. G. Sequeira, L. G. Cancela, and J. L. Rebola, "Impact of physical layer impairments on multi-degree cdc roadm-based optical networks," in *2018 International Conference on Optical Network Design and Modeling (ONDM),* (2018), pp. 94–99.
10. C. Zhang, J. Li, H. Wang, A. Guo, and C. Janz, "Evaluation of dynamic optical service restoration on a large-scale roadm mesh network," IEEE Commun. Mag. **57**, 138–143 (2019).
11. D. Wang, H. Jiang, G. Liang, Q. Zhan, Y. Mo, Q. Sui, and Z. Li, "Optical performance monitoring of multiple parameters in future optical networks," J. Light. Technol. **39**, 3792–3800 (2021).
12. C. Delezoide, P. Ramantanis, L. Gifre, F. Boitier, and P. Layec, "Field trial of failure localization in a backbone optical network," in *2021 European Conference on Optical Communication (ECOC),* (2021), pp. 1–4.
13. J. Xin, X. Li, Y. Tang, L. Zhang, J. Wei, Y. Zhang, and S. Huang, "Harmony search based multi-faults localization for 5g coexisting radio and optical wireless networks," in *2021 Asia Communications and Photonics Conference (ACP),* (2021), pp. 1–3.
14. Z. Yang, D. Hong, Q. Huang, and Y. Chen, "Fault location method for optical transmission network based on low density check matrix," in *2020 IEEE 20th International Conference on Communication Technology (ICCT),* (2020), pp. 578–582.
15. A. P. Vela, M. Ruiz, F. Fresi, N. Sambo, F. Cugini, G. Meloni, L. Potì, L. Velasco, and P. Castoldi, "Ber degradation detection and failure identification in elastic optical networks," J. Light. Technol. **35**, 4595–4604 (2017).
16. L. Wu, J. Cui, J. Zhang, and Y. Ji, "Demonstration of slice fault monitoring and handling for converged optical-wireless access networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC),* (2021), pp. 1–3.
17. K. S. Mayer, J. A. Soares, R. P. Pinto, C. E. Rothenberg, D. S. Arantes, and D. A. A. Mello, "Machine-learning-based soft-failure localization with partial software-defined networking telemetry," J. Opt. Commun. Netw. **13**, E122–E131 (2021).
18. H. Lun, Y. Wu, M. Cai, X. Liu, R. Gao, M. Fu, L. Yi, W. Hu, and Q. Zhuge, "Roadm-induced anomaly localization and evaluation for optical links based on receiver dsp and ml," J. Light. Technol. **39**, 2696–2703 (2021).
19. Z. Li, Y. Zhao, Y. Li, S. Rahman, F. Wang, X. Xin, and J. Zhang, "Fault localization based on knowledge graph in software-defined optical

networks," J. Light. Technol. **39**, 4236–4246 (2021).
20. F. Musumeci, V. G. Venkata, Y. Hirota, Y. Awaji, S. Xu, M. Shiraiwa, B. Mukherjee, and M. Tornatore, "Domain adaptation and transfer learning for failure detection and failure-cause identification in optical networks across different lightpaths [invited]," J. Opt. Commun. Netw. **14**, A91–A100 (2022).
21. C. Zhang, D. Wang, J. Jia, L. Wang, S. Liu, L. Guan, and M. Zhang, "Attention mechanism-driven potential fault cause identification in optical networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC),* (2021), pp. 1–3.
22. C. Zhang, D. Wang, J. Jia, L. Wang, K. Chen, L. Guan, Z. Liu, Z. Zhang, X. Chen, and M. Zhang, "Potential failure cause identification for optical networks using deep learning with an attention mechanism," J. Opt. Commun. Netw. **14**, A122–A133 (2022).
23. H. Yang, X. Zhao, Q. Yao, A. Yu, J. Zhang, and Y. Ji, "Accurate fault location using deep neural evolution network in cloud data center interconnection," IEEE Transactions on Cloud Comput. **10**, 1402–1412 (2022).
24. H. Lun, X. Liu, M. Cai, Y. Wu, M. Fu, L. Yi, W. Hu, and Q. Zhuge, "Gan based soft failure detection and identification for long-haul coherent transmission systems," in *2021 Optical Fiber Communications Conference and Exhibition (OFC),* (2021), pp. 1–3.
25. K. Abdelli, H. Grießer, C. Tropschug, and S. Pachnicke, "Optical fiber fault detection and localization in a noisy otdr trace based on denoising convolutional autoencoder and bidirectional long short-term memory," J. Light. Technol. **40**, 2254–2264 (2022).
26. J. Babbar, A. Triki, R. Ayassi, and M. Laye, "Machine learning models for alarm classification and failure localization in optical transport networks," J. Opt. Commun. Netw. **14**, 621–628 (2022).
27. R. Wang, J. Zhang, S. Yan, C. Zeng, H. Yu, Z. Gu, B. Zhang, T. Taleb, and Y. Ji, "Suspect fault screening assisted graph aggregation network for intra-/inter-node failure localization in roadm-based optical networks," in *2022 European Conference on Optical Communication (ECOC),* (2022), pp. 1–4.
28. Y. Li, J. Lin, L. Zong, S. K. Bose, B. Mukherjee, and G. Shen, "Colorless, partially directionless, and contentionless architecture for high-degree roadms," J. Opt. Commun. Netw. **14**, 481–492 (2022).
29. L. Lu, Y. Li, L. Zong, B. Mukherjee, and G. Shen, "Asymmetric cdc roadms for efficient support of bi-directionally asymmetric traffic demands," J. Light. Technol. **39**, 4572–4583 (2021).
30. A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," Neural networks **18**, 602–610 (2005).
31. X. Li, H. Wang, P. Xiu, X. Zhou, and F. Meng, "Resource usage prediction based on bilstm-gru combination model," in *2022 IEEE International Conference on Joint Cloud Computing (JCC),* (2022), pp. 9–16.
32. P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," stat **1050**, 20 (2017).
33. A. Mohammad, C. Masouros, and Y. Andreopoulos, "Complexity-scalable neural-network-based mimo detection with learnable weight scaling," IEEE Transactions on Commun. **68**, 6101–6113 (2020).
34. Z. Chen, J. Zhang, B. Zhang, R. Wang, H. Ma, and Y. Ji, "Admire: Demonstration of collaborative data-driven and model-driven intelligent routing engine for ip/optical cross-layer optimization in x-haul networks," in *2022 Optical Fiber Communications Conference and Exhibition (OFC),* (2022), pp. 1–3.
35. H. Zhu, H. Zang, K. Zhu, and B. Mukherjee, "A novel generic graph model for traffic grooming in heterogeneous wdm mesh networks," IEEE/ACM Transactions on Netw. **11**, 285–299 (2003).
36. H. Yu, T. Taleb, J. Zhang, and H. Wang, "Deterministic latency bounded network slice deployment in ip-over-wdm based metro-aggregation networks," IEEE Transactions on Netw. Sci. Eng. **9**, 596–607 (2022).