Network Failure Avoidance for Software Defined Optical Networks Using Distributed In-band Network Telemetry

Badr Mochizuki（The Kyoto College of Graduate Studies for Informatics）

Abstract

This paper introduces implementation strategies for realizing fast network failure avoidance based on distributed in-band network telemetry orchestration with P4 for optical switching networks. The main idea consists in collecting telemetry data and store it locally at the switch in order to estimate in real-time the quality of transmission (QoT) of the currently established lightpaths. If the estimated QoT is larger than the acceptable threshold, a lightpath may be rerouted through an alternate path and avoid a potential failure.

1. Introduction

This paper aims to examine how In-band Network Telemetry (INT) [1]—which involves embedding network state data into live flow packets as header fields—can be utilized to avert network failures in real-time whenever the quality of transmission (QoT) on a lightpath surpasses a specific threshold. The primary focus is on leveraging the low latency of INT to sustain high availability in optical switching networks. Two principal objectives are identified: first, the development of a flexible and efficient framework for dynamically orchestrating INT metadata collection without adversely affecting the network's overall performance, particularly in terms of QoT, bandwidth usage and quality of service (QoS); second, the design of a module that computes and enforces recovery routes after a QoT degradation is detected. Notably, designing the INT metadata collection framework is essential, as it will provide a foundation for both estimating the QoT and rerouting lightpath for achieve failure avoidance.

The structure of the paper is as follows: Section 2 introduces relevant background information, and Section 3 provides a summary of current research trends in network failure detection and localization.

Sections 4 discusses strategies for the orchestration of distributed INT data collection aimed at achieving effective failure avoidance in optical core networks. The paper concludes with a summary of the potential impact of INT in enhancing the reliability of optical networks.

2. Background

A. Optical switching networks

In optical switching networks, data is transmitted using light over optical fibers at the speed of light, enabling ultra-fast data transfer rates, ideal for networks spanning long distances [2]. As internet bandwidth demand continues to rise, largely driven by the growth of cloud computing, the Internet of Things (IoT), and 5G mobile communications, network traffic has become increasingly intense and unpredictable. To keep up with these demands, network operators frequently update their equipment, but adding new infrastructure is both costly and limited by physical constraints.

To address these challenges more efficiently, Elastic Optical Networks (EONs) have been proposed as a cost-effective solution. EONs use optical circuit switching, establishing a temporary lightpath for data transmission, through which data travels entirely optically, and then releasing the lightpath when data transfer is complete [3]. These lightpaths in EONs are dynamically set up and dismantled, offering flexibility that enhances the network's adaptability to fluctuating traffic patterns. Data in these networks is carried through optical channels, known as spectrum grids, within the optical fibers [4].

A major advantage of EONs over traditional optical core networks is their ability to support higher data bit rates more efficiently by dividing a single wavelength into multiple frequency slots (FSs). In EONs, these FSs must be allocated in adjacent slots due to the spectrum contiguity constraint [5], ensuring that the spectrum is used effectively.

Figure 1 highlights the primary difference between EONs and traditional optical core networks. In conventional networks, the spectrum grid size is fixed at 50 GHz, meaning that if a lightpath needs 70 GHz, it must occupy two 50 GHz grids, leaving 30 GHz underutilized and unavailable to other paths. Additionally, conventional systems require two guard-bands per grid, even if multiple grids are used by the same lightpath, which further limits efficient spectrum use.

EONs, on the other hand, allow for variable grid sizes. In these networks, the optical spectrum can be allocated per FS unit, and only two guard-bands are necessary per lightpath. This flexibility enables EONs to use the available spectrum much more efficiently than traditional networks, providing greater adaptability and reducing wasted bandwidth in the network.

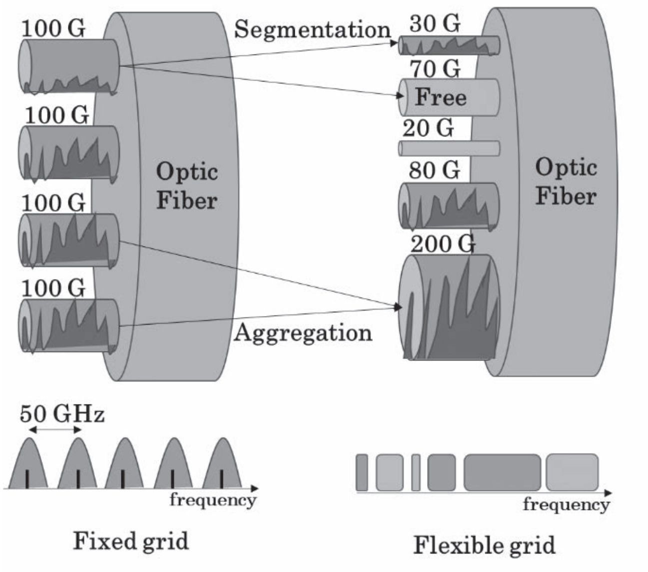

Fig. 1: Conventional optical networks vs. EONs.

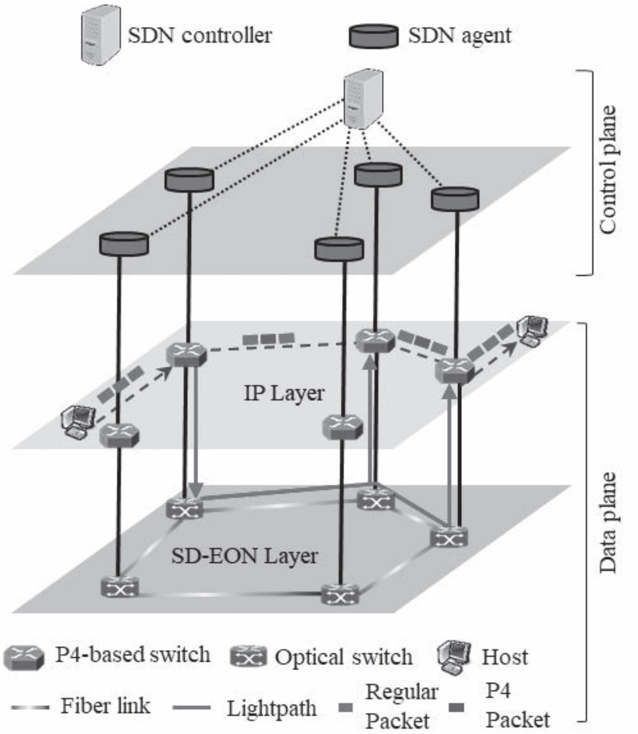

Fig. 2: Software defined EONs.

B. Software Defined EONs

The rapid growth of data centers and the surge in internet traffic have rendered traditional management frameworks inadequate for controlling optical switching networks. To address this issue, Software Defined Networking (SDN) has emerged as a viable solution, offering network operators greater programming flexibility. In SDN, the data plane, responsible for forwarding packets, is separated from the control plane, which manages network rules from a central location. This centralized control plane can add, delete, or modify forwarding rules on SDN switches [6]. As shown in Figure 2 for software-defined EONs (SD-EON), the data plane enables optical switches to send packets over fiber links according to established rules, while the control plane consists of an SDN controller. This controller converts management policies into specific packet-forwarding rules, known as flow entries, which are then installed in the SD-EON switches' flow tables. Communication between the SDN controller and SD-EON switches occurs via protocols such as OpenFlow [7], allowing for efficient, responsive network management in these advanced optical networks.

C. In-band Network Telemetry

In-band network telemetry (INT) is foundational for applications like virtual and augmented reality, health monitoring, and autonomous vehicles. This emerging network monitoring framework operates directly within the data plane and utilizes programmable network devices to gather real-time data on network conditions. INT collects metadata, including both the internal states of network devices (e.g., switch IDs, queue occupancy) and key performance metrics like latency. This data is encapsulated as header fields within the packets, which serve as instructions to network devices, prompting them to gather and forward INT metadata.

Once collected, the metadata in the packet headers is extracted and sent to a monitoring system. This system analyzes the INT data, enabling timely detection of issues within the network. By providing real-time network-wide visibility, INT can detect and address network problems, such as estimating the quality of transmission (QoT), network failures, promptly, improving reliability and responsiveness across the network [1].

D. Data plane programming using P4 language

In traditional software-defined networks (SDNs), while the control plane's behavior can be adjusted, the data plane remains fixed, limiting flexibility in executing actions associated with routing policies. Although routing policies can be defined with some freedom, corresponding actions in the data plane cannot. To address this limitation, a programming language for the data plane, called P4 (Programming Protocol-independent Packet Processors), was developed by the creators of OpenFlow [8]. P4 allows for customizing how SDN switches process incoming packets without modifying the control plane, making it an attractive tool for researchers due to its open-source nature and permissive licensing.

P4 operates through a match-action pipeline structure, as shown in Figure 3. In SDN switches, packet forwarding is achieved through table lookups and predefined actions. Using P4 programming, components like the parser, ingress pipeline, egress pipeline, and deparser can be modified. Initially, an incoming packet is processed by the parser, which extracts the packet header while buffering the payload. The header is then directed to the ingress pipeline, where the programmed lookup keys and actions process it, determining an egress port and a queue for the packet. The packet then moves to the egress pipeline, where additional processing can be applied based on specific instructions. Finally, the deparser reassembles the packet's header and payload before forwarding it to the designated egress port.

This approach allows SDN operators to adapt packet handling within the data plane, enhancing flexibility and enabling more sophisticated,

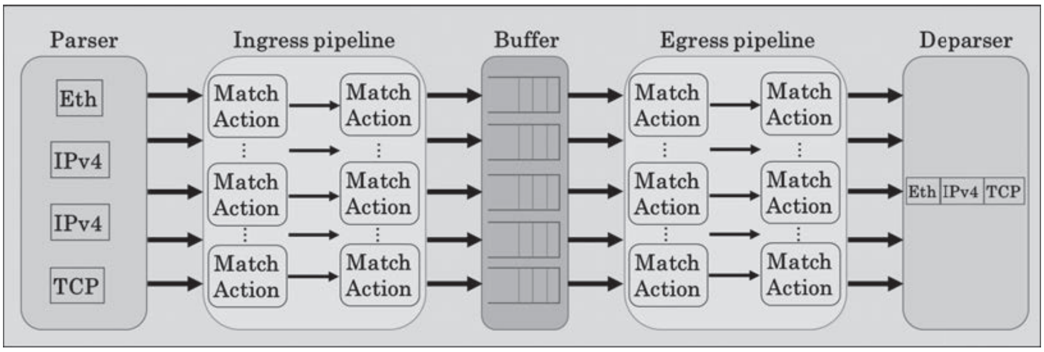

Fig. 3: P4 forwarding model.

customized actions to meet varied network requirements effectively.

3. Related Research

The detection and recovery of network failures in optical networks are essential to maintaining high-speed, reliable data transmission. In-band network telemetry (INT) has gained attention as an effective approach to monitoring and detecting failures in real time. Unlike traditional monitoring systems, which often require separate communication channels, INT embeds telemetry data, such as latency, packet loss, and device states, directly into packet headers as they traverse the network. This enables rapid, in-path collection of diagnostic data, which is then used to identify and address faults efficiently without out-of-band signaling [9].

The use of INT for failure detection allows for a more immediate view of network conditions and enhances the ability to detect issues as they arise. Studies show that INT reduces the mean time to detect (MTTD) failures by providing detailed insights into network performance at various points in the transmission path. This is crucial in optical networks where any downtime can disrupt large volumes of data. With INT, systems can identify deviations in performance metrics, enabling network administrators to locate faults rapidly [10].

Research work in [11] presents a novel framework utilizing in-network P4 processing for distributed multi-layer telemetry data, enabling rapid soft failure detection and recovery in packet-optical networks, achieving effective strategies enforced within microseconds, enhancing overall network resilience and performance.

The authors of [12] presented their work which focuses on In-band Network Telemetry (INT) for fault-tolerant monitoring, introducing In-Patching for rapid recovery from link failures. It emphasizes autonomous detour application in the data plane, significantly improving recovery speed compared to traditional control plane solutions.

Research work in [13] proposes an event-driven method in P4 that can detect link failure with a single probe packet, and in [14] a method locates the exact switch and link of failure using backward probing, hence reducing network overhead.

However, integrating INT into optical networks is not without challenges. The added telemetry data can introduce additional packet overhead, which may affect network performance, particularly during high traffic. Scalability remains another challenge, as high-capacity networks generate substantial amounts of telemetry data that must be efficiently processed to maintain real-time responsiveness. Research continues to address these issues, exploring ways to reduce overhead while enhancing INT's scalability.

INT offers significant benefits for failure detection and recovery in optical networks by enabling faster, more precise, and automated responses to faults. However, the traditional centralized approach for INT orchestration requires a communication with a dedicated INT server, which increases the reaction time to failures.

In the next section, we discuss a distributed approach for INT metadata collection, namely QoT estimation, and finally a lightpath rerouting framework based on the QoT estimation in order failure avoidance in SD-EONs.

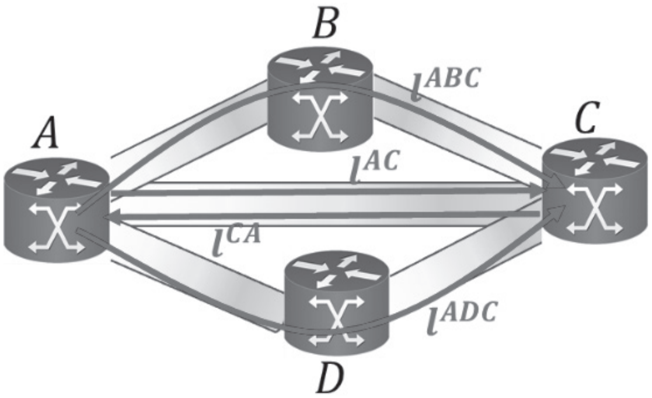

Fig. 4: SD-EON scenario.

4. Implementation strategies

Theme 1: Orchestration of distributed INT metadata collection

This research centers on SD-EON nodes that leverage P4-programmable packet switching technologies for high-speed optical transmission. Such nodes offer the potential for substantial cost reductions, especially in optical metro networks, by eliminating the need for standalone transponders and O/E/O conversions. Additionally, they facilitate seamless integration between the optical and IP layers, paving the way for the deployment of advanced distributed network processing capabilities and enabling innovative distributed processing functions in the data plane.

Figure 4 illustrates an SD-EON comprising four nodes interconnected via optical fibers. The connection between nodes A and C is established by two unidirectional lightpaths, l^{AC} and l^{CA}. Alternative routes linking A and C are also available, traversing either B or D. Each node possesses information regarding the quality of transmission (QoT) for the lightpaths terminated at its coherent receivers, such as the received OSNR, pre-FEC BER, and RX power. Additionally, every node has access to packet-level quality of service (QoS) parameters, including input/output packet or bit rates and queue occupancy. For instance, node C actively monitors the QoT and QoS of optical connections originating from A, B, and D. This allows C to promptly identify and respond to QoT degradations impacting l^{AC}.

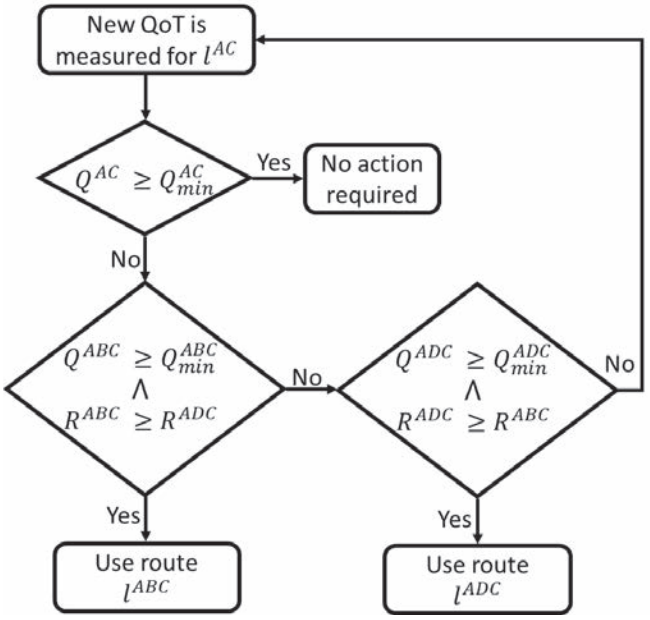

Fig. 5: Failure avoidance algorithm.

Traditional SDN-based mechanisms present challenges for node C in acquiring direct knowledge of the QoT status of the reverse lightpath (l^{CA}). In the event of a QoT degradation on l^{CA}, such as optical equipment malfunction, node C depends on rapid notification from node A. While effective mechanisms exist for handling failures, such as routing protocol alerts and Bidirectional Forwarding Detection (BFD), these approaches are unsuitable for QoT degradations. This is because QoT degradation are characterized by increased pre-FEC BER with occasional post-FEC issues, which are not captured by binary status indicators.

Alternative notification strategies using centralized SDN systems may struggle with scalability and speed, potentially leading to data loss. Furthermore, even if node C identifies the QoT degradation, it lacks a reliable method to quickly find alternative paths with adequate end-to-end QoT and QoS. Node C also has no visibility into the performance of links it is not directly connected to, such as A-B and A-D, which may themselves experience soft failures or insufficient available bandwidth. In such cases, relying on SDN-based intervention may not offer timely solutions for mitigation.

The paper presents a framework designed to address existing limitations by enabling the direct, exchange of INT metadata between SD-EON nodes. These INT metadata contain real-time performance metrics—both QoT and QoS—relevant to the specific links of interest for a given node. For instance, node C can assess the status of outbound links that are one hop away by collecting telemetry data from nodes A, B, and D. This allows node C to select the most suitable connection, whether it be the direct l^{AC} lightpath or alternative routes such as l^{ABC} via node B or l^{ADC} via node D.

The INT metadata are processed at wire speed within the network using the P4 ASIC, which leverages its stateful capabilities to perform operations on the retrieved data. In this framework, the SDN Controller pre-calculates alternative routes to activate in the event of a QoT threshold violation on any active lightpath. These calculations consider the real-time QoT and QoS metrics of each potential route to ensure thresholds are satisfied and QoS levels are maintained. The SDN Controller then encodes these decisions as custom P4 rules, which are deployed to the P4-enabled nodes. Finally, the algorithm enforces the necessary flow rules and thresholds for real-time operation using protocols like P4Runtime or Thrift.

Theme 2: Lightpath rerouting framework for failure avoidance

The algorithm presented in Fig. 5 outlines the P4 rules configured by the SDN Controller at node C to address potential QoT degradations on l^{AC}. Let Q^X represent the current QoT measurements for lightpath X, Q^X_{min} denote the corresponding QoT threshold, and R^X indicate the unused FSs for lightpath X. Initially, the algorithm verifies if l^{AC} meets its QoT threshold; if it does, no action is required. If the threshold is violated, the algorithm identifies a node—either B or D—for re-routing the traffic. This selection process involves ensuring that all links in a potential alternative path meet their QoT thresholds and then choosing the path with the highest available unused FSs.

By pre-determining alternative routes, the P4 ASIC can promptly redirect traffic when the QoT threshold is breached. This distributed telemetry framework enables real-time failure avoidance as well as detection of failures, while assessing the performance of viable end-to-end alternative routes. This ensures swift and efficient traffic steering to maintain network reliability and performance.

5. Conclusion and Future work

The strategies mentioned in section 5 will be confirmed through simulations using the network simulator Mininet [15], and INT data collection procedures will be implemented in P4. Furthermore, the P4 program will be written for the V1Model architecture implemented on P4.org's bmv2 software switch. using the experimental setup provided in the P4 tutorial [16]. As a future work, practical experiments using P4 compatible hardware will be conducted in order to validate the results obtained by Mininet simulations.

[References]

[1] L. Tan et al., "In-band network telemetry: a survey", *Computer Networks*, vol.186, pp.1-24, 2021.

[2] B. C. Chatterjee, N. Sarma, P. P. Sahu, and E. Oki, "Limitations of Conventional WDM Optical Networks and Elastic Optical Networks for Possible Solutions", *Routing and Wavelength Assignment for WDM-based Optical Networks*, Springer, Cham, vol. 410, pp. 101-115, 2017.

[3] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic Optical Networking: A New Dawn for the Optical Layer?", *IEEE Communications Magazine*, vol. 50, no. 2, pp. S12-S20, 2012.

[4] B. C. Chatterjee, N. Sarma, and E. Oki, "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1776-1800, 2015.

[5] M. Recalcan, F. Musumeci, M. Tornatore, S. Bregni, and A. Pattavina, "Benefits of Elastic Spectrum Allocation in Optical Networks with Dynamic Traffic", *IEEE Latin America Transactions*, vol. 13, no. 11, pp. 3642-3648, 2015.

[6] B. Nunes, et al., "A survey of software-defined networking: Past, present, future of programmable networks", *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617-1634, 2014.

[7] Open Networking Foundation, https://www.opennetworking.org

[8] P. Bosshart et al., "P4: Programming Protocol-Independent Packet Processors", *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 3, pp. 88-95, 2014.

[9] D. Medhi and K. Ramasamy, "Network Routing: Algorithms, Protocols, and Architectures", Morgan Kaufmann, 2017.

[10] L. Tan, et al., "In-band Network Telemetry: A

Survey", Computer Networks, vol. 186, 107763, 2021.
[11] F. Cugini, C. Natalino, D. Scano, F. Paolucci, and P. Monti, "P4-based Telemetry Processing for Fast Soft Failure Recovery in Packet-Optical Networks", 2023 Optical Fiber Communications Conference and Exhibition (OFC), CA, USA, March 2023. DOI: 10.23919/OFC49934.2023.10117408
[12] A. G. Gastro, et al., "Taking Detours: An In-Network Fault-Tolerant Probing Planning for In-Band Network Telemetry", ICC 2023, Rome, Italy, 2023. DOI: 10.1109/ICC45041.2023.10279199
[13] B. Mochizuki, " Event-driven approach for link failure detection in P4", B-7-16 Information Network, IEICE Society Conference, Saitama, Japan, 2024.
[14] B. Mochizuki, " Precise failure localization with P4 for software-defined network",P1-34, International Conf. on Emerging Technologies for Communications, Fukuoka, Japan, 2024.
[15] Mininet: http://mininet.org/
[16] P4 Tutorial: https://github.com/p4lang/tutorials

◆◆著者紹介

望月バドル　Badr Mochizuki

京都情報大学院大学　准教授
福井大学大学院工学研究科修了　工学博士
京都大学大学院情報科学研究科博士課程単位取得定退学
奈良先端科学技術大学院大学情報科学研究科修了 工学修士
元 CNRS 研究所（フランス）　研究員
Al Akhawayn 大学（モロッコ）　工学士