Unavailability Analyses of Hyperscale Data Center Interconnect Optical Networks

Lingling Wang [(1)], Lei Wang [(1)], Chunxiao Wang [(1)], Chongjin Xie [(2)]

[(1)] Alibaba Cloud, Alibaba Group, Beijing, China, *wanglingling.wll@alibaba-inc.com
[(2)] Alibaba Cloud, Alibaba Group, New York, USA

Abstract *With massively collected daily operation data from our production optical networks, we analyse the main factors affecting the unavailability of our inter-data center optical networks. Some suggestions to improve the optical network availability are discussed. ©2023 The Author(s)*

Overview

Reliability and availability are crucial factors for cloud network operators to provide high-quality cloud services. Data center interconnect (DCI) optical networks, which offer ultra-high-speed, low-latency, and high-bandwidth network connection, have been widely used in cloud data centers, and DCI networks in hyperscale data centers are transforming toward openness and disaggregation, resulting in the continuous improvement of network scale and complexity [1].

The production DCI optical network we studied has a hierarchical structure composed of optical transmission sections (OTS), optical multiplex sections (OMS), and optical channels (OCH), as shown in Fig. 1(a). Each OCH in the network contains 4 OMSes, 13 OTSes on average and an average length of 905 km. In such a disaggregated network, various transmission components (e.g., equipment, fiber) are sourced from different vendors, leading to multiple concerns on network reliability and availability.

Furthermore, this network is designed without optical layer protection from the beginning. Instead, the equal cost multi-path protection (ECMP) mechanism is enabled at the internet protocol (IP) layer upon a link fault. However, the switching of ECMP takes several seconds, causing network packet loss or service jitter in a link fault. In extreme cases where all channels in the ECMP group are disconnected, border gateway protocol (BGP) route convergence is required, leading to long service interruption. Therefore, it is important to conduct deep analyses of end-to-end network availability and accurately identify the main factors causing network unavailability to ensure and improve service quality.

In this paper, we analyse unavailability data about 400 OCHes over the past year. We study the main factors that cause network unavailability, using daily network performance data and case tickets collected by Alibaba optical network analytics (AONA) platform. Finally. we provide suggestions and discuss their effectiveness on improving end-to-end network availability.

Alibaba data analytics platform of DCI optical network

Fig. 1(b) shows the data analytics platform architecture of our DCI optical networks for data monitoring and analysing. In the data collection and storage layer, firstly, we periodically collect original data from the optical network components covering all our data centers, including fixed asset, real-time performance reported every 15 minutes, and alarm information, which are temporarily stored in the online collection database (DB). Secondly, the change or failure event will be manually confirmed and submitted through the standardized process, and saved in case ticket system, including network

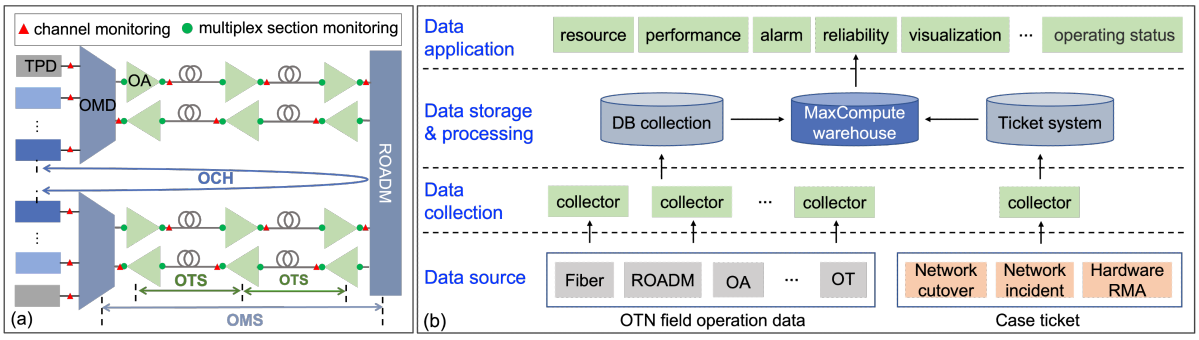

Fig. 1: Architecture of DCI optical network and Alibaba optical network analytics (AONA). (a) architecture of DCI optical network, (b) architecture of AONA. TPD, transponder; OMD, optical multiplexer and demultiplexer; OA, optical amplifier; ROADM, reconfigurable optical add–drop multiplexers; OTS, optical transmission section; OMS, optical multiplex section; OCH, optical channel; OT: optical terminal. DB: database. RMA: return merchandise authorization.

Tab. 1: Unavailable cause information.

Cause	Data source	Description
Fiber break	Performance/ Ticket/Alarm	Fiber loss exceeds the break threshold; Alarms and fiber case tickets are also triggered synchronously.
Network cutover	Ticket	Anticipated network cutover with details recorded through ticket.
Fiber loss degradation	Performance/ Ticket/Alarm	Fiber loss is larger 3dB than expected loss, or the fluctuation of loss exceeds 3dB within a statistical period of 15 minutes; Alarms and fiber case tickets are also triggered synchronously.
Hardware incidents	Ticket	Equipment hardware failure, with the start time and finish time recorded.
Channel degradation	/	This part mainly corresponds to the single channel unavailability.

cutover, fiber failure incidents, hardware return merchandise authorization (RMA), etc. In the processing layer, data in DB and case ticket system are transferred to the MaxCompute platform, a home-grown big data analysis platform, every day for long-term data storage and automatic analysis. Finally, MaxCompute provides data support for various services at the application layer, such as resource management, performance and alarm analysis, reliability evaluation, data visualization, and network operation status monitoring.

End-to-end network unavailability analysis

In our platform, as shown in Fig. 1(a), OCH can provide channel operational states and performance monitors, such as input power, output power, error seconds (ES), and unavailable seconds (UAS) [2]. End-to-end network unavailability is defined as abnormal seconds (AS) generated within a 15-minute period on an OCH. Note that, $\mathrm{AS} = \mathrm{ES} + \mathrm{UAS}$. Moreover, optical amplifier (OA) and reconfigurable optical add–drop multiplexers (ROADM) can report power, gain, and attenuation, which can be used to calculate real-time fiber loss by the powers at both ends of a span. Meanwhile, we record fiber lengths and expected span losses are updated regularly. Combined with aforesaid performance data and case tickets, we classify the causes of OCH abnormal seconds, as shown in Tab.1.

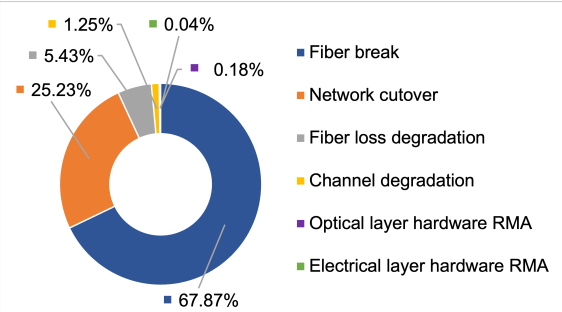

Fig. 2: The distribution of abnormal seconds caused by various causes.

As shown in Fig. 2, we counted all OCH abnormal seconds over the past year and plotted the distribution of abnormal seconds caused by various causes. Furthermore, we studied the abnormal duration in continuous statistical periods about various causes and plotted the cumulative distribution functions (CDFs) separately, as shown in Fig. 3. There are five types of causes over the OCH abnormal seconds: 1) Fiber breaks. As shown in Fig. 2, it accounts for 67.87% of all OCH abnormal seconds. Fiber breaks can be caused by several reasons, especially some emergencies, such as construction excavation, natural disasters, man-made destruction, etc. As they occur unexpectedly, they usually take a relatively long time to recover. As can be seen from Fig. 3 (a),

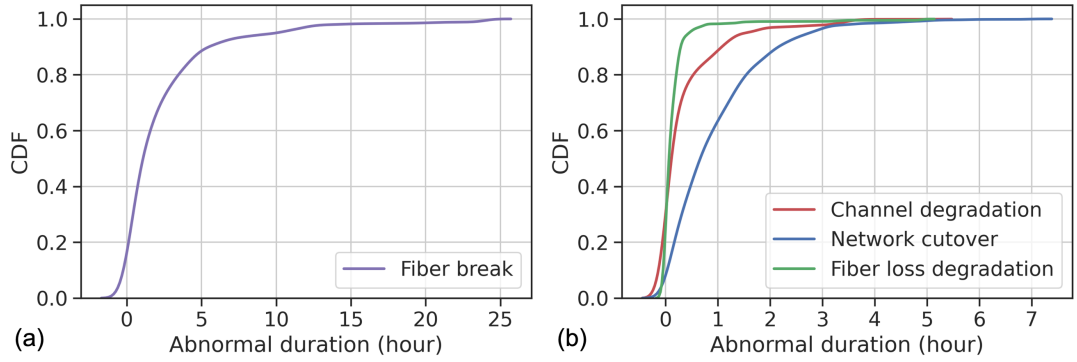

Fig. 3: Abnormal duration distributions caused by various causes. (a) fiber break, (b) Network cutover / fiber loss degradation / Channel degradation. CDF, cumulative distribution function.

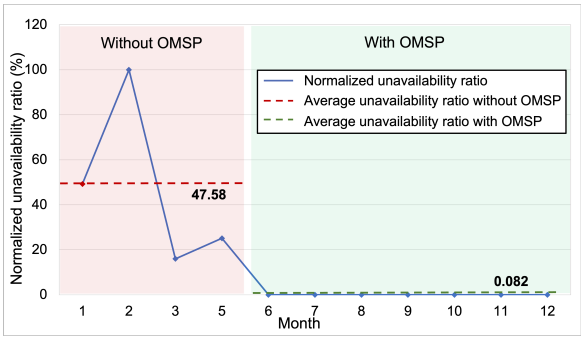

Fig. 4: Unavailability ratio improvement of OCHs by OMSP.

only 76.77% of abnormal duration caused by fiber break is less than 3 hours. In addition, 4.98% of them are more than 10 hours. 2) Network cutover. Fig. 2 shows that it causes about 25.23% of abnormal seconds. As shown in Fig. 3 (b), 96.47% of abnormal duration caused by network cutover is within 3 hours, and 99.36% is within 5 hours. Thus, it can be seen that the impact of network cutover planned ahead by operators is less than unexpected fiber breaks. 3) Fiber loss degradation. As shown in Fig. 3 (b), 98.27% of the abnormal duration caused by fiber loss degradation is within 1 hour and 86.32% within 15 minutes, and a considerable part of them can recover after a short period of jitter. 4) Equipment hardware RMA. RMA mainly corresponds to the failure and replacement of service cards in the optical layer or electrical layer. As shown in Fig. 2, abnormal seconds caused by RMA are few. As the distribution is not statistically significant, we do not plot the CDF. 5) Channel degradation. The amount of this part is about 1.24% of the total abnormal seconds, and we do not further identify the causes of channel degradation now. In theory, for single channel unavailability, we can roughly confirm whether the channel unavailability is due to optical layer or electrical layer problems by evaluating the channel generalized signal-to-noise ratio (GSNR) using optical channel monitors (OCM) data [3]. As 87.21% of the abnormal duration of this type is within 1 hour, as shown in Fig. 3(b), and the collection period of OCM data is also 1 hour, it is difficult to cover such short-term unavailability. We will work this out using telemetry OCM data pushed every minute in the future.

The above analyses show that most OCH failures are caused by fibers. Some techniques to reduce the failures are suggested and their effectiveness to improve network availability is measured. 1) Optical multiplex section protection (OMSP) [4]. For this scheme, we demonstrated the protection effect on part of spans in the production network, and analysed the unavailability ratio of OCHs only passing through these spans. In this context, unavailability ratio refers to the ratio of abnormal seconds to the total seconds (abnormal & normal) in every month, and the result has been normalized. Theoretically speaking, OMSP switching can be completed within 10 milliseconds, due to the acquisition accuracy, the actual average abnormal seconds caused by OMSP switching is 1.25 seconds. As shown in Fig. 4, we can see that the average unavailability ratio of OCHs without OMSP is 47.58%, while with OMSP is 0.082%, reducing the unavailability ratio by 580 times. This seems to be in line with expectations. In fact, OMSP can cover the majority of OCH unavailability caused by fiber breaks, loss degradation, and even cutover. 2) Automatic power adjustment [5]. We analyzed the OCH unavailability caused by fiber loss degradation, and about 70% of the degradation is still within adjustment margin of OA and variable optical attenuator (VOA). Automatic power adjustment can improve this degradation in the first time without manual intervention and ensure service stability adequately. In the current unprotected scenario, the unavailability caused by fiber loss degradation can be reduced by 3.3 times by automatic power adjustment. 3) Improve network cutover efficiency. As mentioned, network cutover resulted in 25.23% of OCH unavailability. Therefore, we should ask operators to improve cutover efficiency and shorten cutover frequency and duration as much as possible.

Conclusion

In conclusion, the analyses of collected daily operation data for a production optical network have allowed us to identify and study various factors that influence network availability and showed that fiber failures are the main factor for the unavailability of the network. Some techniques to improve network availability are suggested and their performance is investigated and discussed.

Reference

[1] C. Xie, L. Wang, L. Dou, M. Xia, S. Chen, H. Zhang, Z. Sun, and J. Cheng, "Open and disaggregated optical transport networks for data center interconnects [Invited]," in *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C12-C22, June 2020, DOI: 10.1364/JOCN.380721.

[2] "Error performance parameters and objectives for multi-operator international paths within optical transport networks," ITU-T Recommendation G.8201, https://www.itu.int/rec/T-REC-G.8201/.

[3] Y. He, Z. Zhai, L. Wang, Y. Yan, L. Dou, C. Xie, C. Lu, and A. P. T. Lau, "Improved QoT Estimations through Refined Signal Power Measurements in a Disaggregated and Partially-loaded Live Production Network," in Optical Fiber Communication Conference (OFC) 2023, paper Tu2F.5, https://opg.optica.org/abstract.cfm?uri=OFC2023-Tu2F.5.

[4] "Optical transport network: Linear protection," ITU-T Recommendation G.873.1, https://www.itu.int/rec/T-REC-G.873.1/.

[5] H. Zhang, F. Gao, J. Cheng, L. Dou, S. Chen, B. Yan, Z. Sun, L. Wang, and C. Xie, "Demonstration of a

Disaggregated ROADM Network with Automatic Channel Provisioning and Link Power Adjustment," *2021 European Conference on Optical Communication (ECOC)*, Bordeaux, France, 2021, pp. 1-3, DOI: 10.1109/ECOC52684.2021.9606164.