PROACTIVE MAINTENANCE IN A SUBMARINE NETWORK USING ANALYTICS

Andrée Robichaud-Véronneau (Ciena), Brian Hart (Southern Cross), Landry Aguehounde, Radhika Jayaprakash, Lévis Pepin, Catherine Preston-Thomas, Marc-Antoine de Santis, (Ciena)
Email: arobicha@ciena.com

Ciena, 2351 Boul. Alfred Nobel, Suite N-300, Montréal (Saint-Laurent), QC H4S 2A9 Canada.

Abstract: As the world is evolving towards full connectivity between devices and people, reliability of telecommunications networks is the key for success. Unplanned outages, with their induced high costs and SLA[1] violations, are the main showstopper on this road. Network Health Analytics is aimed at solving this problem by providing a risk assessment of individual ports on a network, hence enabling service providers to take proactive maintenance actions in a planned and timely manner and reducing impact for end-customers. The historical equipment performance data from the network, as well as a global database of industry-wide network data collected on a cloud-based infrastructure, is fed into a Machine Learning algorithm to obtain the port risk level. A prioritized list is provided to the customer on a user-friendly dashboard to apply principles of preventive action on risky ports before they have a failure or trigger an alarm. With the collaboration of Southern Cross, the algorithm was successfully demonstrated on submarine repeaters.

1. INTRODUCTION

Unplanned network outages are a main source of concern for telecommunication providers worldwide. Their cost is estimated at an average of 5,600$ per minute of downtime and have been reported as high as of 11,000$ per minute [1]. Furthermore, the non-estimated costs on the long-term of customer dissatisfaction are non-negligible, given the numerous options available to them on the market for a reliable network.

One approach to reduce the number of unplanned outages is through the combined power of big data and machine learning, as implemented in the Network Health tool. This tool has been used since 2017 in a submarine network at Southern Cross, combining submarine and terrestrial equipment data with submarine repeater scans.

Modern telecommunication providers are using optical networks to transport various types of information (such as voice or data) over large distances. The building blocks of an optical network are shown in Fig. 1.

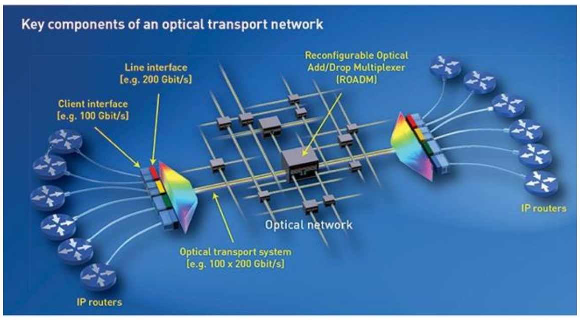

Figure 1: Example of an optical transport network [2].

The various devices required to build an optical transport network are interlinked using optical fiber patch cables connecting to physical communication ports. A physical port can be provisioned with one or many

[1] Service Level Agreement

OAM&P[2] capabilities, or *facilities*. A logical port forms the smallest entity in the network and is therefore identified as the physical port information combined with a single facility identifier, each with associated monitoring parameters.

2. NETWORK HEALTH TOOL DESCRIPTION

The objective of the Network Health (NH) tool is to determine the health of each individual port in a network on the sole basis of Performance Monitoring (PM) parameter values.

The input data for the NH tool includes PM parameter values, inventory details, selected facility and adjacency information, as well as geographical locations for shelves (used on the presentation layer only, not in the analysis). These data are collected from the provider's network once a day. The frequency of the measurements corresponds to the time bin setting for this particular network element (typically 24h or 15min). These data are combined to form unique time series for each port and PM parameter, as shown in Fig. 2.

As a machine learning application, the NH algorithm extracts operational ranges entirely from the power of the data, so no extra documentation or prescription is needed as input. To perform the analysis, the collection of timeseries data, i.e. the dataset, is split into two subsets: the most recent data (last 15 days collected) is used for risk assessment (i.e. the test set) and the rest of the older data is used to train the algorithm (i.e. the training set). Data quality selections are applied to focus the learning on the expected, normal behaviour of the equipment: only ports for which an entry in the inventory data is present and that are carrying traffic ("In Service") are selected. Additionally, a minimum number of three data points are required for a timeseries to be added to the dataset.

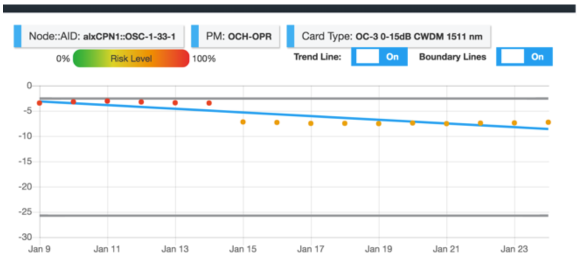

Figure 2: Timeseries for the alxCPN1::OSC-1-33-1 port and the OCH-OPR PM parameter.

The training set data is clustered into logical groups, formed by the product equipment code (PEC), the PM parameter, the facility and, exceptionally in the cases of amplifiers, the port number. The data coming from various sources (customers, laboratory setups, manufacturing) can be combined in each group to augment the statistical power, and therefore reduce the uncertainty on the learned operational ranges (a minimum number of 30 entries is required in each group to be used for the risk assessment). Over 90 million unique data points are currently used in the combined NH training dataset.

In each learning data group, the Network Health algorithm considers the risk associated to the features of the timeseries data as well as to the time trends and patterns. A linear regression fit is applied to the entire dataset (both the training and test sets), by splitting it into segments based on the scale of the data point value difference and fitting them independently.

Four components are considered to evaluate the global risk level of a port:

- Individual data point values
- Difference between consecutive data point values
- Slope values obtained from the linear regression fit to the data

[2] Operations, Administration, Maintenance & Provisioning

- Times required for the linear regression fit extrapolation to cross the PM parameter values operational bounds.

A gaussian anomaly detection algorithm is used for each risk component to evaluate the average value and the operational ranges. The average is set to 0 for the difference and slope components, where the expectation for a normal behaviour is such that no change is expected. The operational bounds are derived symmetrically as three times the standard deviation of each gaussian distribution.

The data from the test set, as its name states, are therefore tested against those learned features. Some assumptions are made in the risk assessment. In the case where a data point value was outside of its range and the next value moves back inside the range, no risk is associated with the difference in values. Similarly, if a linear regression fit is pointing towards the learned average value for its appropriate group, no risk is associated to the slope value. The risk associated to the time extrapolation component is computed non-symmetrically, in the positive direction. For all other cases, the risk is evaluated linearly and symmetrically from the average value of each distribution.

The risks are combined using weights to account for average and maximum risk values in each component (to detect spurious effects and long-term variations) as well as a specific weight assigned to each component to obtain a global risk value for a timeseries. These timeseries risks are averaged over the PM parameters for each port to form a global risk value. The ports are ranked according to their computed global risk value and classified in three categories: high, medium and low risk. The details of the risk values for each timeseries (before and after weighting) are displayed on a web application, as well as the raw data points for the test set.

3. NETWORK HEALTH FOR A SUBMARINE NETWORK

The Network Health algorithm being agnostic to the data origin, is by construction multilayer and multivendor. In the case of submarine networks, this means that data from terrestrial landing station nodes can be analysed together with submarine repeater scan data in a single tool. By adapting the logical port identifier to make use of the repeater identification details, namely the segment, fiber pair and direction, the learning algorithm can form logical groups for repeater data.

The result of such and analysis can be seen in Fig. 3. Given the scarcity of submarine data, the timing windows to distinguish between the learning and the test set was adjusted from 15 to 1500 days. A ranked list of all the repeater ports is presented at the center of the dashboard, and the top risky port timeseries is shown for the gain parameter on the graph at the bottom. Two occurrences of variations in the gain values have been detected for this port in the test set time range. Also, the values themselves are outside of the operational range detected for this group of ports. Four PM parameters are used for submarine repeater ports: gain, pump current, output and input power.

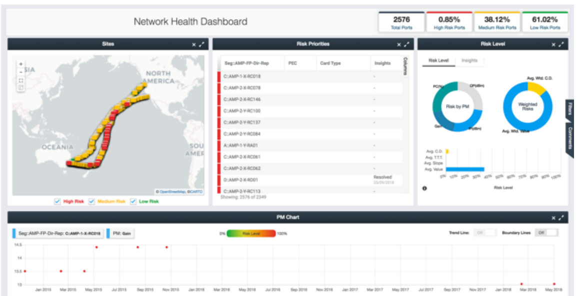

Figure 3: Network Health presentation layer for a submarine network.

The case of submarine repeater data poses the interesting challenge of detecting long term patterns in lowly populated datasets. A use case of major interest for submarine network operators is to detect the degradation of the amplification gain over the lifetime of the repeater. By detecting the end of life patterns for a repeater, a proactive replacement operation can be performed in a controlled manner.

Landing station data can also highlight potential issues in the network. As an example, an issue occurred in December 2018 in the Southern Cross network across a particular segment. It was clearly seen that the Laser Bias Current dropped across all ports associated with that segment by approximately 0.5dB. Once resolved, the levels returned to normal as could be seen in the Network Health tool.

The ports highlighted by the NH tool and their abnormal data patterns need to be validated by the user. As the algorithm makes the assumption that a stable behaviour is not risky, any variation will be detected and flagged. In cases where the fluctuations are known to be caused by normal network operations or addition of new services, the assessed risk should be ignored.

4. CONCLUSION AND OUTLOOK

Machine learning and artificial intelligence are the stepping stones on the road towards more automated and adaptive networks. By taking full advantage of the mass of data that their optical equipment provides, network operators can optimise their operations and their costs with better planning and proactive maintenance. With the advent of 5G technology and higher data traffic, the Network Health tool can enable operators to avoid unplanned outages and ensure smooth services for the benefit of all users. Looking ahead, the anomaly detection algorithm that is Network Health could be enriched by adding inputs to improve the quality of the results.

5. REFERENCES

[1] D. J. Cappuccio, 'Ensure Cost Balances Out With Risk in High-Availability Data Centers', 'Gartner report G00238137', 11 February 2013.

[2] P.J. Winzer, 'Scaling Optical Fiber Networks: Challenges and Solutions', 'Optics and Photonics News', 1 March 2015.