Clustering of live network alarms using unsupervised statistical models

Diane Maillot-Tchofo, Ahmed Triki, Maxime Laye, John Puentes

Clustering of Live Network Alarms Using Unsupervised Statistical Models

Diane Maillot-Tchofo[1,2], Ahmed Triki[1], Maxime Laye [1], John Puentes[2]

[1] Orange Innovation, 2 Avenue Pierre Marzin, 22300 Lannion, diane.maillottchofo@orange.com
[2] IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest, France

Abstract *An unsupervised topology and time-based clustering model is proposed to regroup alarms according to their failure events. The different modes and settings of the model are assessed using topology and alarm-related data extracted from a live network as part of a field trial.* ©2023 The Author(s)

Introduction

To ensure high service availability in optical networks, many research efforts have been put to manage failures, including fault prediction and root cause analysis. In a live network, thousands of failure events occur every year. Given the relative opacity and the complexity of the optical systems, maintenance experts estimate the average time required to take charge of one failure event at around 30 minutes. In this context, we believe that statistical-based approaches like machine learning (ML) and deep learning (DL) methods could assist operators to manage failures, thus leading to reduce fault handling time and increase operational efficiency.

Authors in[1] compared several machine learning models to identify correlated alarms related to unexpected power attenuation failure and then localized the position of this failure in the network. Data were labelled and extracted from a Network Management System (NMS) connected to an experimental set-up. In[2], an unsupervised approach based on autoencoder was investigated with experimental data to identify alarms raised by one type of failure.

Authors in[3],[4] and[5] use approaches based on autoencoders, Support Vector Machine, and Gaussian Processes respectively, for failure detection and identification. Instead of using alarm data, these studies are based on performance monitoring data extracted from experimental testbeds.

In this paper, we show the results of a field trial where we assess the performance of our unsupervised clustering model, which aims to provide clusters of alarms, each corresponding to a specific failure event. The dataset extracted from the NMS of a national metro/core network provides an inventory of the network topology and a list of tens of thousands of alarms raised. Our two-stage model performs in the first stage a topology-based classification and in the second stage a time-based clustering.

The novelty of this paper lies in the application of unsupervised clustering techniques on live-network data. Results were checked by maintenance experts who have expressed their satisfaction with the algorithm execution time and the precision of the obtained clusters.

Field Trial Description

The field trial was performed in a live operational network. The network is managed by a NMS and is composed of thousands of nodes and several tens of thousands of optical services extended over the metro/core national network. We developed a data extractor module that uses the North Bound Interface (NBI) of the NMS to collect data related to: i) the topology (i.e., nodes, network elements cards, and links); ii) services (i.e., the modulation format, bit rate, and physical routes); and iii) the alarms raised during the first quarter of the current year. For each alarm, the respective data describe: the severity, the affected object that could be a node, a card, a port or a service, the probable cause as locally seen by the affected object, the site name, and the time when an alarm is detected. Alarms raised by a service can originate from one of the OTN/SDH layers or the NMS itself. Finally, the obtained raw data are pre-processed by removing the cleared alarms and outliers.

Unsupervised Model for Clustering

The goal of the proposed solution is to group together alarms reported due to the same failure event. After data pre-processing, the alarm classification is performed by our two-stage model as shown in Fig. 1.

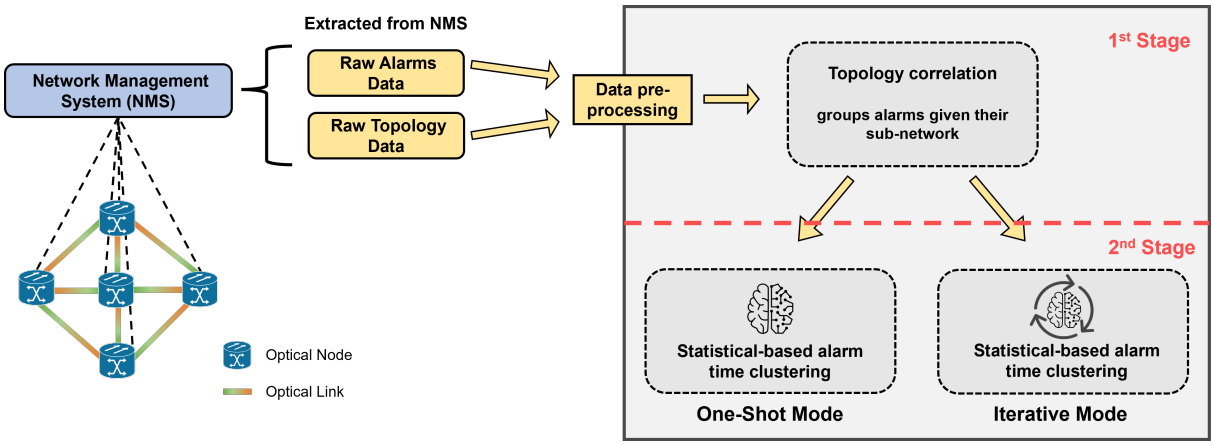

Fig. 1: Proposed two-stage model for alarm classification

In the first stage, a topology-based classification is performed. This classification benefits from the hierarchical structure of the network where several services are established within the same sub-network. Thus, two network objects are considered topologically correlated if the failure occurring in one network object can probably affect the second one. This correlation can be manifested by a physical or logical relationship such as cards belonging to the same equipment or ports involved in the same service route. At the end of this stage, *topological classes* are obtained.

In the second stage, we perform a time-based clustering within each *topological class* separately. We assess different types of statistical clustering and each of these models is applied to the alarm date (i.e., "detection time" field), once transformed into timestamp with precision to the minutes. We assume that the detection time is reliable and network nodes are time-synchronized, which are fair assumptions as a majority of network objects are time synchronous.

For each topological class, it is possible to carry out the alarm time clustering according to the one-shot mode or the iterative mode. In the first mode, the clustering algorithm is applied only once. In the second mode, the clustering algorithm is executed iteratively until one of the following two conditions is met:

- The longest inter-arrival time between alarms within the same cluster is lower than a specific time threshold.
- The list of alarms within all the obtained cluster has not changed between two consecutive iterations.

Moreover, single-alarm clusters are not allowed if the inter-arrival time between a given cluster and its neighbor is inferior to a threshold. It should be noted that the inter-arrival time between two clusters is the time duration between the latest alarm of the first cluster and the earliest alarm of the second cluster. Within a *topological class*, each cluster has a *neighbor* that is the cluster with respect to which it has the shortest inter-arrival time.

Results and Discussion

In Fig. 2, we assess the performance of our model by comparing different clustering methods for the second stage. To do so we cover the main unsupervised classes of algorithms, namely : density-based clustering in OPTICS (Ordering Points To Identify the Clustering Structure)[6] and Mean Shift[7], hierarchical-based clustering in AHC (Agglomerative Hierarchical Clustering)[8], and finally the well-known vector-quantization based clustering in K-Means[9]. Moreover, we consider three empirical time thresholds: 15 minutes, 30 minutes, and 60 minutes.

Regarding the *within-cluster* standard deviation shown in Fig. 2b, the smaller it is, the better our clustering is considered to be, and conversely for *between-cluster* standard deviation depicted in Fig. 2a. Otherwise, there is a clear trend with *within-cluster* distances. When the threshold is reduced, the standard deviation decreases significantly, especially between the 30 and 60 minutes thresholds, with a 150 seconds difference. These two plots inform us that the applied algorithms evolve in the same way throughout thresholds and clustering modes. Moreover, the *within-cluster* standard deviation gap between the iterative and one-shot clustering gets larger with longer time thresholds.

The Calinski-Harabasz (CH) scores[10] obtained on clusters using the different combination of modes and time thresholds can be seen in Fig. 2c. The CH score is defined as the ratio of the *between-cluster* variance to the sum

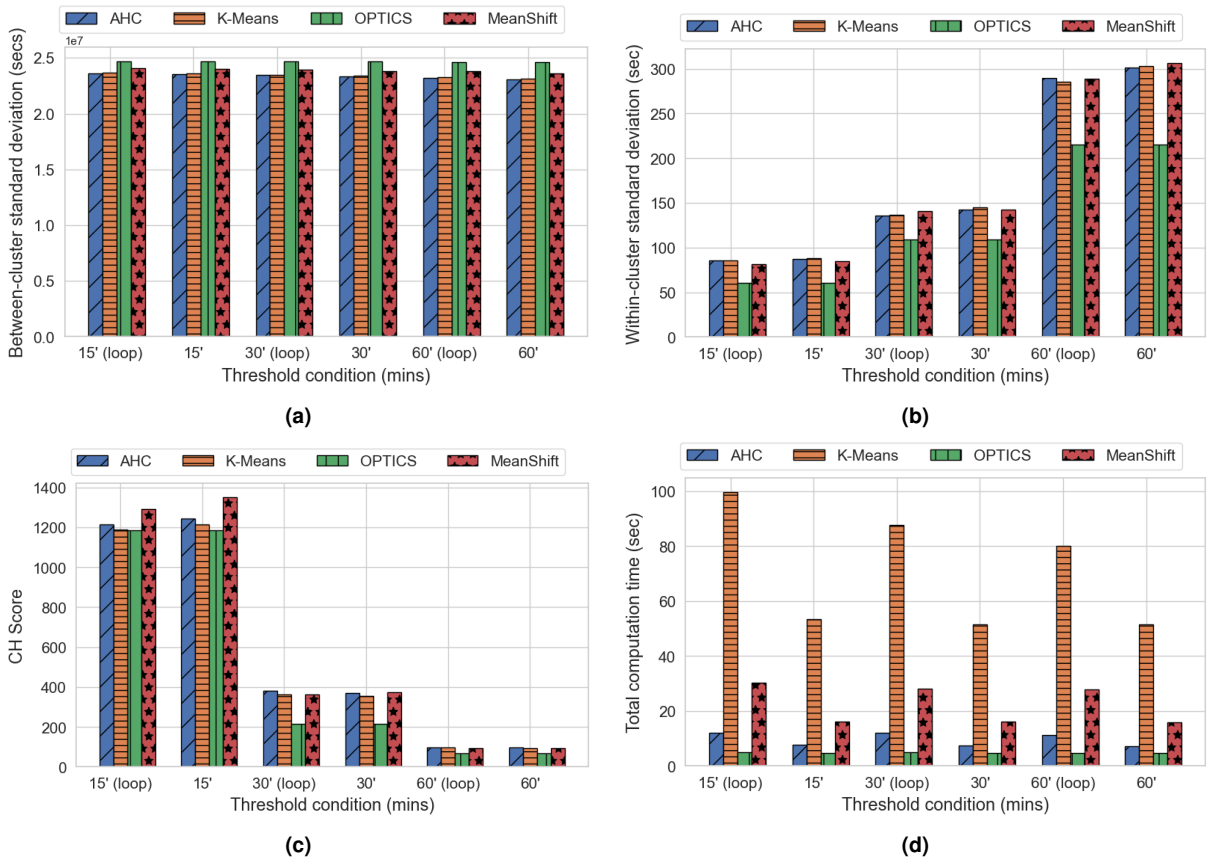

Fig. 2: Clustering results of the different modes and thresholds. (a) CH Score per model given the threshold and algorithm type. (b) *within-cluster* standard deviation per model given the threshold and algorithm type. (c) *between-cluster* standard deviation per model given the threshold and algorithm type. (d) Computation times per model given the threshold and algorithm type.

of *within-cluster* variances multiplied by a weight. The higher the CH score is, the better the separation and cluster density are.

For the different clustering methods, the scores are largely similar, being the overall trend the same for all models. In addition, the OPTICS model has notably worse CH scores regardless of modes and thresholds. The huge gap between the 15 and 30 minutes thresholds (regardless of modes) can be explained by the large increase of small clusters, when comparing the former threshold to the latter. This leads to smaller *within-cluster* variances and bigger *between-cluster* variance.

When considering the results by class of algorithm, the density-based algorithms do not benefit as much from iterative clustering compared to the other two classes. Indeed, for example, the mean *within-cluster* distances decreased by 9.6% for K-Means clustering with a 60 minutes threshold, and a similar trend with the AHC algorithm (7.32% decrease with the same threshold) was observed. Thus, in this instance, if one chooses OPTICS, the iterative mode is not needed, which is not the case for AHC or even K-Means.

Finally, aside from the K-Means algorithm, regardless of modes and thresholds, the other approaches take less than 30 seconds to compute as shown in Fig. 2d. Furthermore, considering that our data were extracted from a live network, the low computation times suggest that our approach would be scalable, because minimal preprocessing and topology classification times are required.

Conclusions

In this paper, we presented an unsupervised topology and time-based clustering model to regroup alarms related to the same failure event. Data extracted used to evaluate the model are unlabeled and extracted from a live network as part of a field trial. The comparative study taking into account the different modes and settings of our two-stage model shows that the model is able to ensure high CH score with low computation time, especially when the OPTICS algorithm is used with a threshold tuning parameter of 15 minutes. These results are in accordance with experts feedback who verified the relevance of the obtained clusters.

References

[1] J. Babbar, A. Triki, R. Ayassi, and M. Laye, “Machine learning models for alarm classification and failure localization in optical transport networks”, *Journal of Optical Communications and Networking*, vol. 14, no. 8, pp. 621–628, 2022.

[2] L. Z. Khan, A. Triki, M. Laye, and N. Sambo, “Optical network alarms classification using unsupervised machine learning”, in *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC)*, 2022, pp. 1–3. DOI: 10.23919/OECC/PSC53152.2022.9849872.

[3] S. Liu, D. Wang, C. Zhang, L. Wang, and M. Zhang, “Semi-supervised anomaly detection with imbalanced data for failure detection in optical networks”, 2021, Th1A.24. DOI: 10.1364/OFC.2021.Th1A.24. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2021-Th1A.24.

[4] S. Shahkarami, F. Musumeci, F. Cugini, and M. Tornatore, “Machine-learning-based soft-failure detection and identification in optical networks”, 2018, M3A.5. DOI: 10.1364/OFC.2018.M3A.5. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=OFC-2018-M3A.5.

[5] F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, and M. Tornatore, “A tutorial on machine learning for failure management in optical networks”, *Journal of Lightwave Technology*, vol. 37, no. 16, pp. 4125–4139, 2019. DOI: 10.1109/JLT.2019.2922586.

[6] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “Optics: Ordering points to identify the clustering structure”, *Association for Computing Machinery SIGMOD Record*, vol. 28, no. 2, 49â60, 1999. DOI: 10.1145/304181.304187. [Online]. Available: https://doi.org/10.1145/304181.304187.

[7] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition”, *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975. DOI: 10.1109/TIT.1975.1055330.

[8] J. C. Gower and G. J. S. Ross, “Minimum spanning trees and single linkage cluster analysis”, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 18, no. 1, pp. 54–64, 1969. DOI: 10.2307/2346439. [Online]. Available: http://www.jstor.org/stable/2346439.

[9] J. MacQueen, “Some methods for classification and analysis of multivariate observations”, vol. 5, 1967, pp. 281–297.

[10] T.Calinski and J.Harabasz, “A dendrite method for cluster analysis”, *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. DOI: 10.1080/03610927408827101. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610927408827101.