CHALMERS
UNIVERSITY OF TECHNOLOGY

Machine Learning Analysis of Polarization Signatures for Distinguishing Harmful from Non-harmful Fiber Events

Machine Learning Analysis of Polarization Signatures for Distinguishing Harmful from Non-harmful Fiber Events

Leyla Sadighi[1], Stefan Karlsson[2], Lena Wosinska[1], Marija Furdek[1]

[1]*Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden*

[2]*Swedish Defense Material Administration, 586 63 Linköping, Sweden.*

E-mail: sadighi@chalmers.se

ABSTRACT Secure and reliable data transmission in optical networks is essential for supporting high-speed internet services. Optical fibers, the enabler of global connectivity for millions of users, are vulnerable to various potentially harmful events including mechanical failures, like fiber cuts, and malicious physical layer attacks, such as eavesdropping. These incidents can degrade network performance, breach privacy and integrity through unauthorized access to the transmitted data, and cause significant financial and data loss. It is, therefore, crucial to detect and classify the malicious events. Continuous monitoring of polarization state changes, combined with application of machine learning algorithms, enables detection of deviations in the polarization patterns caused by the harmful events. In this study, we introduce a method that detects and identifies potential harmful events in optical networks. By using a Histogram Gradient Boosting classifier within our machine learning framework, we achieve 97.94% detection accuracy of the harmful and non-harmful events.

Keywords: Polarization state movement, polarization signatures, harmful vibration, eavesdropping, machine learning

1. Introduction

The capacity demand in communication networks is growing exponentially, aiming to support the increasing number of users as well as new and emerging online services. To keep up with this development, fiber optic networks offering ultra-high transmission capacity are considered as the future-proof technology. The optical networks carry a lot of data, including sensitive information, ranging from national defense communications to personal private data. Their crucial role in maintaining both short- and long-distance connectivity underscores the importance of safeguarding the integrity of transmitted information. However, the physical nature of these networks makes them vulnerable to sabotage from various sources, which can lead to degraded signal quality, disrupted communication services, and violated data security.

Various external events, regardless of intentions or aims, result with vibrations that impact signals transmitted over fiber optic installations. Examples include non-harmful events, such as the normal vibrations in buildings, as well as potentially harmful incidents, such as the operation of an excavator near a fiber installation. When an excavator's engine works at 4800 rpm, it generates 80 Hz vibrations, which indicates a fiber cut risk. Another severe security threat is eavesdropping and/or information tampering. While encryption algorithms are commonly employed to secure the data, advancements in computing power and the development of quantum computers threaten the efficiency of these security measures. Optical fibers can be eavesdropped relatively easily, and may be difficult to detect by measuring the received optical power [1]. These vulnerabilities underscore the importance of identification and categorization of harmful events, such as eavesdropping and harmful vibrations, to ensure the privacy and integrity as well as reliability of fiber optic communications.

A common way of detecting external fiber events is by monitoring the State of Polarization (SOP), i.e., tracking the polarization state of propagated light. Changes in the polarization state can indicate physical or environmental stress, such as mechanical vibrations and bending [2]. Continuous SOP monitoring is essential in optical networks for early detection of anomalies, enabling proactive measures against fiber damage [3].

A proactive fiber break detection in optical communication systems is proposed in [4]. It employs quaternion time series analysis and machine learning (ML) to classify various mechanical events induced by robotic arm movements, such as bending, shaking, small hits, and up-and-down movements. It utilizes SOP data transformation into quaternion sequences, later re-coded into relational data for event classification using a naive Bayes classifier with over 99% accuracy. The study in [5] introduces a transfer learning approach using a deep convolutional neural network for high-risk event classification given a small amount of SOP data. It demonstrates image-based ML as very efficient for detection and classification of five mechanical events in optical fiber: bending, shaking, small hits, up-and-down movements, and fan ventilation. A novel vision transformer-based method for detecting and localizing mechanical vibrations in optical networks using SOP data is proposed in [6]. In our previous work [7] we used supervised ML to detect and distinguish between 13 different polarization signatures in three cable types.

In this paper, we propose a polarization-based fiber optic sensor, where we apply supervised ML algorithms and the SOP data collected from 5 distinct events occurring in an indoor optical fiber network. Our sensor

979-8-3503-7732-3/24/$31.00

can differentiate between potentially harmful and non-harmful events, enhancing the security and reliability of optical network infrastructures.

2. Experimental Setup

We generate polarization signatures by tracking changes in the polarization state on the Poincaré sphere. As depicted in Figure 1 a continuous-wave Distributed Feedback (DFB) laser serves as the primary light source. This laser inserts light into the optical fiber transmission line, thereby enabling the detailed examination of polarization changes. The optical power from the DFB laser is initially transmitted through a 1 km coupling fiber, which is then connected to an indoor cable and a bare fiber segment used for eavesdropping simulations. Following this, the setup includes a 20 km fiber spool, resulting in a total transmission length of 21 km. The polarization of light transmitted through optical fibers can be changed by mechanical vibrations and eavesdropping attempts.

In our testbed, we collect unique signatures for specific types of manipulations that mimic both harmful and non-harmful events. The received optical signal is analyzed by an optical and an ML analyzer. Polarization signatures are created by deliberate actions applied between the 1 km coupling fiber and the 20 km coupling fiber (see Figure 1). Each type of fiber manipulation makes a distinctive impact on the Polarization Signature Movement (PSM). Following the method detailed in [7], we obtain a distinct signature for each type of event. The process starts with the sampling block capturing PSM samples from the polarization analyzer on the Poincaré sphere in 1 ms intervals over a 20-minute time period, resulting in 1.2 million samples per event. The system then computes the numerical distances between successive PSM values, thus generating data referred to as Numerical Polarization State Movements (NPSM). These NPSM measurements are organized into sets of 1000, creating distinct time segments. A Fast Fourier Transform (FFT) with frequency size of 512 is subsequently applied to these segments under a Hamming window [8], yielding a spectral power dataset comprising 1200 rows (each representing a time segment) and 512 columns (each for a frequency bin). With this approach, we generate a unique dataset for each event, creating a distinctive signature for each event type.

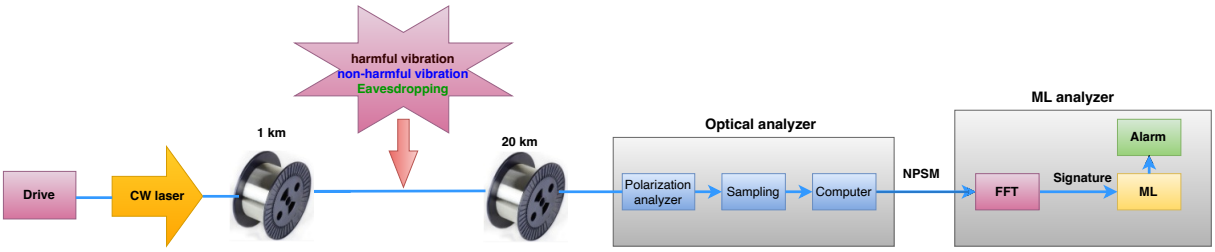


Figure 1: Experimental setup for analysis of polarization signatures. The manipulation is applied between the 1 km and the 20 km fiber spools.

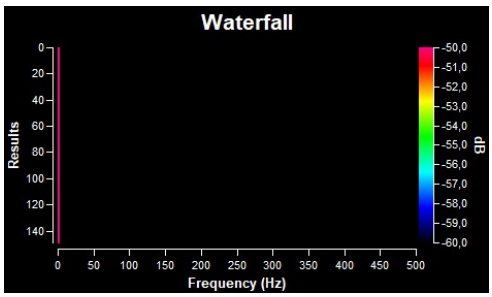
3. Signatures and Data Collection

We consider two types of harmful events in an indoor cable installation: eavesdropping and vibrations at a frequency matching that of an excavator, as well as non-harmful events at a different frequency.

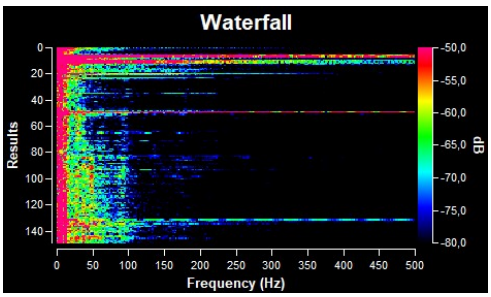
In the eavesdropping (*eav*) scenario, we perform subtle manipulations of the indoor cable, such as subjecting it to a pulling force while it is bent. As referenced in [1], eavesdropping can cause optical power attenuation of less than 0.3 dB, a level typically undetectable by Optical Time-Domain Reflectometer (OTDR). The signature for this scenario is shown in Figure 2b. The second type of considered potentially harmful events involves fiber vibrations at 80 Hz, typically resulting from activities such as excavator digging near the cable installation and posing a risk of cutting the fiber. The 80 Hz vibration is generated by a loudspeaker, positioned 1 km away from the piezoelectric vibrator. This setup produces vibrations on the indoor cable. We collected signatures of the 80 Hz vibration, denoted as *hrmf_80Hz_vb*, (see signature in Figure 2d), as well as the combined 80 Hz harmful vibration and 140 Hz non-harmful vibration, denoted as *hrmf_80Hz_140Hz_vb*, (see signature in Figure 2e) with 210 Hz overtone as potentially harmful events.

The non-harmful events, denoted as *nhrmf_140Hz_vb*, include any normal vibrations that do not pose a risk to the infrastructure, e.g., vibration caused by a fan in a building. For this study, we collected non-harmful vibration data at a frequency of 140 Hz, generated by a piezoelectric vibrator directly attached to the indoor cable (see signature in Figure 2c). A baseline scenario, denoted as *rlx*, represents normal operating conditions, characterized by the absence of harmful vibrations, non-harmful vibrations, or eavesdropping activities (see signature in Figure 2a).

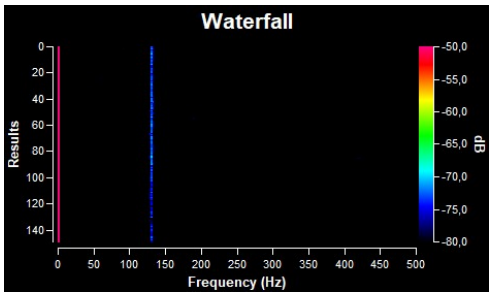
Our ML models use data from the above five use cases to differentiate between normal operation patterns and those suggesting eavesdropping attempts, potential harmful vibrations, and non-harmful vibrations. The dataset



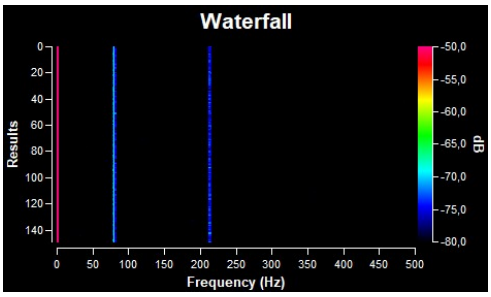
(a) Relaxed fiber



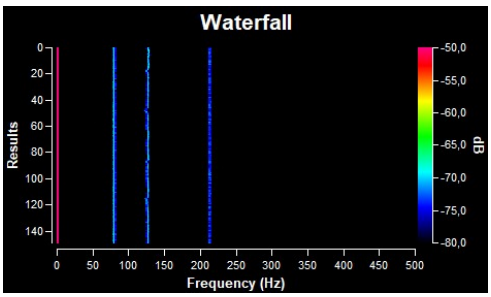
(b) Eavesdropping



(c) Normal vibration 140 Hz



(d) Harmful vibration 80 Hz with 210 Hz overtone



(e) Harmful vibration 80 Hz plus normal vibration 140 Hz with 210 Hz overtone

Figure 2: Waterfall (visual representation of signatures as defined in [7]) for five considered scenarios

is organized into five distinct classes, resulting in a supervised ML classification problem. We randomly divide the dataset into training and testing subsets, with 70% (840 points) and 30% (360 points) respectively, with equal representation across the five scenarios. As a result, the training set comprises 4,200, and the testing set 1,800 samples.

4. Results

We conduct a comprehensive evaluation of multiple supervised ML algorithms to identify the most effective classifier for our five-class classification problem, tailored for our specific dataset. This evaluation utilizes several classifiers from the Scikit-Learn library, selected based on their potential applicability and performance metrics in similar scenarios: Extreme Gradient Boosting (XGBoost), Random Forest, Bagging with Decision Trees, Decision Tree, Histogram Gradient Boosting, Gradient Boosting, Support Vector Machines, Logistic Regression, Extra Trees Classifier, Bagging Classifier, k-Nearest Neighbors, Multi-Layer Perceptron Neural Network, and Linear Discriminant Analysis. We assess these classifiers based on their accuracy and the F1-score using the testing dataset. The results for the top-performing four classifiers are summarized in Figure 3. the Histogram Gradient Boosting and the Gradient Boosting classifiers outperform the others, achieving an impressive accuracy of 97.94% and an F1-score of 0.9794. However, Gradient Boosting demands ten times longer training time. XGBoost and Support Vector Machines also deliver robust performance, closely matching that of the Histogram Gradient Boosting.

The confusion matrix for the Histogram Gradient Boosting classifier in Figure 4 shows detailed performance across the five classes. The classifier correctly identifies 97.78% of the baseline (rlx) instances, with a minor misclassification rate of 1.39% into 'eavesdropping' (eav) and 0.83% into 'non-harmful vibration' (nhrmf_140Hz_vb) classes. The 'non-harmful vibration' class achieves a high accuracy of 99.44%, with only 0.56% samples misclassified as baseline. The 'eavesdropping' (eav) class was accurately identified in 98.61%

of the samples, with a 1.39% misclassification into baseline. For the two potentially harmful vibrations events, the classifier correctly identifies 95.0% of the 'hrmf_80Hz_vb' instances, but there are some misclassifications: 3.33% of samples is incorrectly identified as baseline, 0.28% as 'nhrmf_140Hz_vb' ,and 1.39% as 'hrmf_80Hz_140Hz_vb'. The 'hrmf_80Hz_140Hz_vb' class has a high accuracy of 98.89%, with only 1.11% misclassified as 'hrmf_80Hz_vb'. These results highlight the effectiveness of the Histogram Gradient Boosting classifier in distinguishing between different types of events in fiber optic networks. The high accuracy and F1-scores across most classes demonstrate the model's robustness in identifying both harmful and non-harmful events.

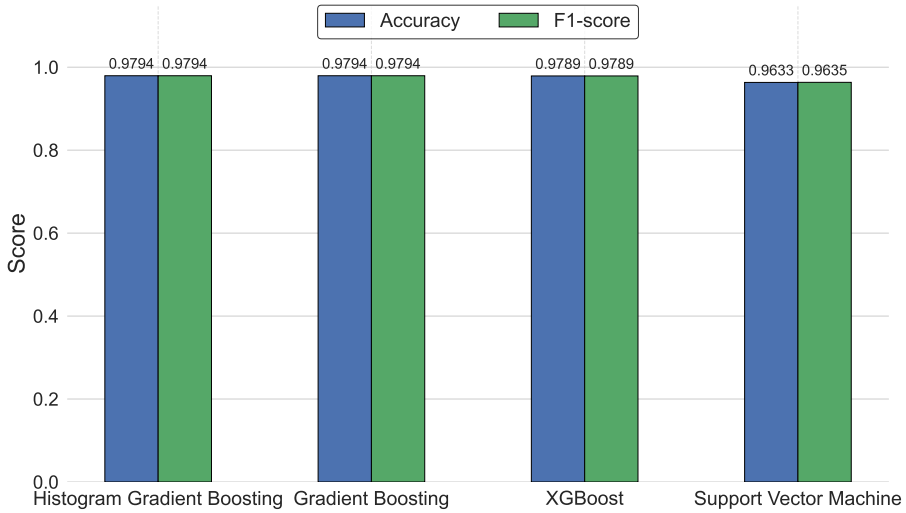


Figure 3: Results of four top performing ML classifiers

5. Conclusion

In this study, we analyze polarization signatures and propose a method to detect potentially harmful events in optical fiber networks. Among the tested classifiers, the Histogram Gradient Boosting demonstrated superior performance, achieving an accuracy of 97.94% and an F1-score of 0.9794. Detailed analysis using the confusion matrix reveals high precision in distinguishing between the various types of anomalies, particularly in identifying harmful events such as fiber vibrations caused by excavator activities. The results highlight the robustness and reliability of the proposed method in detecting and classifying different types of anomalies in optical networks. Future work will focus on expanding the dataset to include a broader range of scenarios and refining the machine learning models to further improve detection accuracy and computational efficiency.

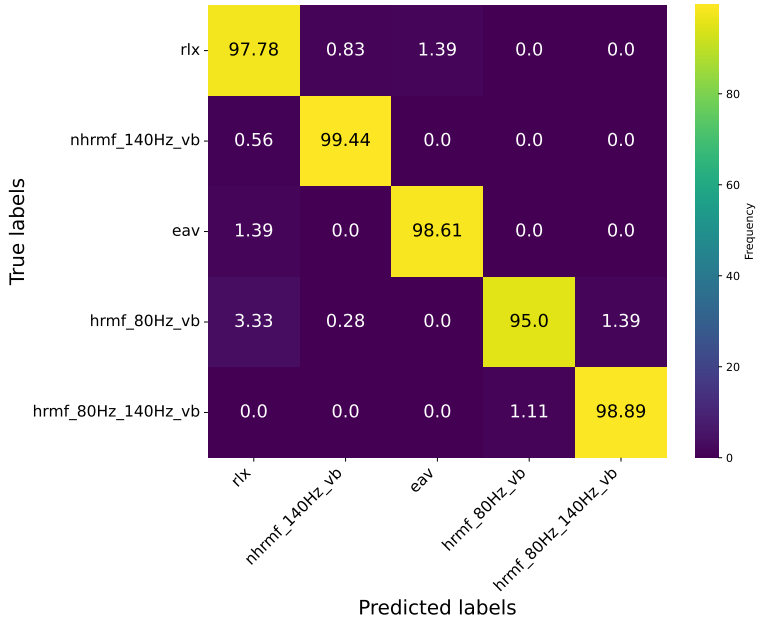


Figure 4: Results of Confusion Matrix for Histogram Gradient Boosting classifier

REFERENCES

[1] S. Karlsson, R. Lin, L. Wosinska, and P. Monti, "Eavesdropping g. 652 vs. g. 657 fibres: a performance comparison," in *International Conference on Optical Network Design and Modeling (ONDM)*, 2022, p. Tu1.4.

[2] J. Pesic, E. Le Rouzic, N. Brochier, and L. Dupont, "Proactive restoration of optical links based on the classification of events," in *International Conference on Optical Network Design and Modeling (ONDM)*. IEEE, 2011, pp. 1–6.

[3] F. Boitier, V. Lemaire, J. Pesic, L. Chavarría, P. Layec, S. Bigo, and E. Dutisseuil, "Proactive fiber damage detection in real-time coherent receiver," in *European Conference on Optical Communication (ECOC)*. IEEE, 2017, pp. 1–3.

[4] V. Lemaire, F. Boitier, J. Pesic, A. Bondu, S. Ragot, and F. Clérot, "Proactive fiber break detection based on quaternion time series and automatic variable selection from relational data," in *Advanced Analytics and Learning on Temporal Data (AALTD)*. Springer, 2020, pp. 26–42.

[5] K. Abdelli, M. Lonardi, J. Gripp, S. Olsson, F. Boitier, and P. Layec, "Breaking boundaries: harnessing unrelated image data for robust risky event classification with scarce state of polarization data," in *European Conference on Optical Communications (ECOC)*. IET, 2023, pp. 924–927.

[6] K. Abdelli, M. Lonardi, J. Gripp, D. Correa, S. Olsson, F. Boitier, and P. Layec, "Anomaly detection and localization in optical networks using vision transformer and sop monitoring," in *Optical Fiber Communication Conference (OFC)*, 2024, p. Tu2J.4.

[7] L. Sadighi, S. Karlsson, C. Natalino, and M. Furdek, "Machine learning-based polarization signature analysis for detection and categorization of eavesdropping and harmful events," in *Optical Fiber Communication Conference (OFC)*, 2024, p. M1H.1.

[8] P. Podder, T. Z. Khan, M. H. Khan, and M. M. Rahman, "Comparative performance analysis of hamming, hanning and blackman window," *International Journal of Computer Applications*, vol. 96, no. 18, pp. 1–7, 2014.