CHALMERS
UNIVERSITY OF TECHNOLOGY

Detection and Classification of Eavesdropping and Mechanical Vibrations in Fiber Optical Networks by Analyzing Polarization Signatures Over a

Detection and Classification of Eavesdropping and Mechanical Vibrations in Fiber Optical Networks by Analyzing Polarization Signatures Over a Noisy Environment

Leyla Sadighi[(1)], Stefan Karlsson[(2)], Carlos Natalino[(1)],
Lena Wosinska[(1)], Marco Ruffini[(3)], Marija Furdek[(1)]

[(1)] Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden, sadighi@chalmers.se.
[(2)] Swedish Defense Material Administration, Stockholm, Sweden.
[(3)] School of Computer Science and Statistics, CONNECT Centre, Trinity College Dublin, Ireland.

Abstract *We propose a machine-learning-based method to detect and classify eavesdropping and mechanical vibrations in an optical network based on state of polarization variations. Tests in two real-world installations with links of different lengths demonstrate an accuracy of 86.5% in 7 distinct normal and malicious scenarios. ©2024 The Author(s)*

Introduction

Cyber security is widely recognized as a critical concern due to the immense significance of online services and information transmitted over communication networks. In turn, the security of fiber optical network infrastructure as the foundation of global communications is rapidly gaining relevance. The recent years have observed an increasing number of confirmed sabotage attempts on fiber optical installations worldwide[1]–[3], which could have a high impact on the global connectivity, economy and defense strategies. The risk of fiber eavesdropping and/or tampering with sensitive information is also becoming severe. An eavesdropper can couple out light from an optical fiber relatively easily. Recent study[4] sheds light on the vulnerability of optical fiber systems to eavesdropping. The information transmitted in the fiber can be detected by an eavesdropper tapping a certain percentage of the light. The success of such eavesdropping attempts strongly depends on the technique employed for tapping the light and the distance of the breach from the transmitter. Gaining access to the optical signal within a certain distance from the transmitter enables an eavesdropper to detect sensitive data. This necessitates the development of eavesdropping detection strategies capable of accurately identifying eavesdropping activities even amidst the prevalent noise in fiber optical networks.

Optical fiber tampering causes changes of the polarization state of the carried light. In general, polarization state movement (PSM) data offers crucial insights into the polarization characteristics of light signals in a network. Continuous monitoring of the state of polarization (SOP) has been demonstrated as essential for prompt identification of network disruptions, enabling early detection of potential fiber damage[5]. Close examination of SOP changes at the receiver and comparison with variations associated to normal system behaviour can enable detection of eavesdropping attempts that cannot be discovered by monitoring the received optical power. This technology was demonstrated in[6]. Earlier research looked into naive Bayes classifiers to spot vibrations in optical fibers caused by mechanical stress[5],[7]. Vibrations, created by robotic arm movements, were detected with a coherent receiver. Additionally,[8] suggested a transfer learning method to classify high-risk events using limited SOP data.

In[9], we experimentally collected and analyzed 13 distinct polarization signatures using a supervised ML algorithm. The results demonstrated that our model could accurately detect and differentiate between signatures from eavesdropping attacks and other potentially harmful and non-harmful events, achieving an accuracy of 92.3%.

In this paper, we study how polarization signatures can be recorded and classified, originating from an installed transmission line in a real-life network OpenIreland, operated by Trinity College and located under the street in Dublin, Ireland. Analysing two separate installations with transmission lengths of 0.15 km and 10.5 km, respectively, we obtained 86.5% accuracy in classifying the signatures using supervised ML.

Experimental setup

The experimental setup is illustrated in Fig.1. A continuous-wave, linearly polarized distributed feedback laser (DFB laser) is used as a transmitter to inject light into the transmission line. The optical power from the DFB laser is first transmitted through a 1 km coupling fiber and then connected to the cable installation. Another 1 km long coupling fiber is added before the receiver. The two

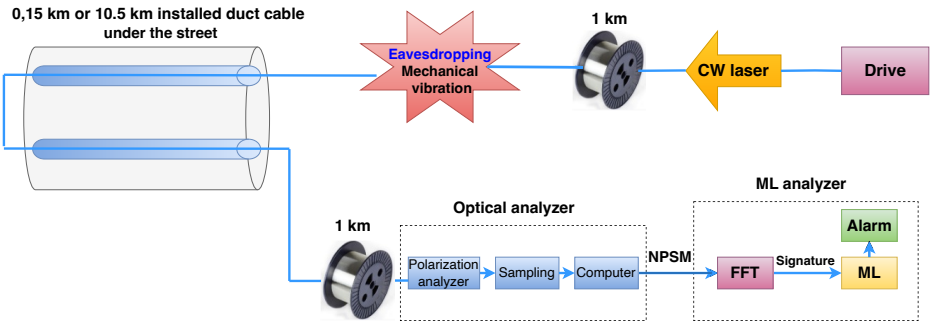


Fig. 1: Experimental setup for analyzing polarization signatures in installed cables with 0.15 km and 10.5 km

fiber cable installations, 0.15 km and 10.5 km long, are used with two fibers connected in a loop, resulting in the total transmission length of either 2.3 km or 23 km. The launch optical power is approximately -10 dBm and the received optical power is in the range of -11 to 21 dBm. The polarization of the transmitted light is affected by the vibrations originating from the street traffic taking place during the experiment. The cable duct consists of a fiber blown into a tube, which partially protects against vibrations but allows direct contact with the tube wall, leading to vibration-induced noise caused by the street traffic.

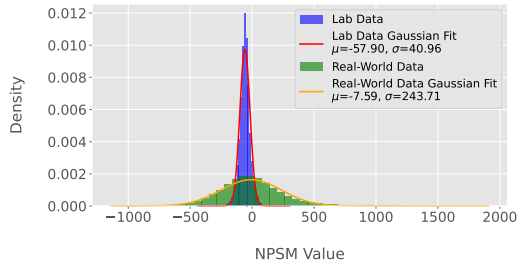
Accessing the fiber for manipulation requires altering the cable. To reliably detect vibration signatures related to this manipulation, they must notably exceed the background noise, which may arise from benign sources like non-harmful vibrations or potentially harmful activities such as eavesdropping and nearby excavation. The polarization signatures are collected by monitoring the PSM on the Poincaré sphere. The received optical signal is analyzed by an optical and an ML analyzer.

We adopt the analytical procedure from[9] to derive a unique signature for each event type. The process begins with the sampling block capturing PSM samples from the polarization analyzer on the Poincaré sphere every 0.5 ms during a 20-minute recording period, resulting in 2.4 million samples for each event. The system then calculates the numerical value of distances between consecutive PSM, generating what is referred to as numerical polarization state movement (NPSM) data. The NPSM values are grouped into batches of 500, forming individual time slots. A fast Fourier transform (FFT) analysis is then performed on these segments using a Hamming window[10], producing a power spectrum dataset with 4800 rows (each corresponding to a time slot) and 512 columns (each representing a frequency bin). This dataset forms the unique signature for each specific event. ML techniques are then applied to the data to identify distinct signatures and trigger an alarm if a threat to the transmission line is detected.

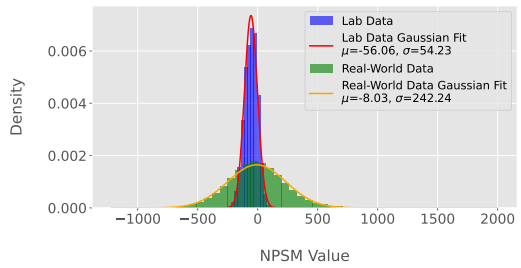
Definition of signatures and data collection

Our ML analysis uses data from seven real-life network signatures to differentiate between polarization patterns obtained during normal operation and those suggesting malicious vibrations and eavesdropping. We denote normal events as relaxed (*rlx*) and soft bending (*sbd*) fiber scenarios. The *rlx* scenario identifies a baseline scenario without eavesdropping, vibrations, or bending for both the 0.15 km (*rlx-0.15km*) and the 10.5 km (*rlx-10.5km*) fiber installation. *sbd* involves only gentle bending of the cable to assess its resilience to benign environmental stress. Signatures from soft bending events are collected for the 0.15 km (*sbd-0.15km*) and 10.5 km (*sbd-10.5km*) installations. Eavesdropping on the 0.15 km installation (*eav-0.15km*) scenario assesses the ability to detect unauthorized interception attempts by observing subtle manipulations of the cable, such as subjecting it to a pulling force while it is bent. As referenced in[4], eavesdropping can cause optical power attenuation below 0.3 dB, a level typically undetectable by an Optical Time-Domain Reflectometer (OTDR). The objective of collecting this signature is to determine if our model can distinguish this eavesdropping activity from soft bending (*sbd-0.15km*). Potentially harmful events considered include fiber vibrations at 80 Hz (*80vb*), typically corresponding to an excavator digging close to the cable installation and threatening to cut the cable, here generated by a loudspeaker. Vibration data at this frequency is gathered for the 0.15 km (*80vb-0.15km*) and the 10.5 km (*80vb-10.5km*) installation. The collected dataset is randomly partitioned into training and testing subsets, containing 70% (3360) and 30% (1440) points, respectively, to ensure equal representation of the seven scenarios. Consequently, the training set consists of 23,520 samples, while the testing set contains 10,080 samples. This dataset, labeled into seven distinct classes, represents a supervised ML classification problem.

Fig.2 shows a comparison of the data collected in the lab[9] and the real-world data for the relaxed (Fig.2a) and 80 Hz vibration (Fig.2b) scenarios, revealing distinct behavior. While the lab data exhibits a narrower distribution and lower variability,



(a) Relaxed scenario



(b) 80 Hz vibration

Fig. 2: Comparison between the data collected in the lab[9] and the real-world data for the 10.5 km link.

the real-world data shows a wider spread and increased variability. These differences highlight how environmental conditions affect real-world data by introducing more noise, thereby making the task of detection and classification for ML algorithms more complex.

Results

We evaluate several ML algorithms to determine a suitable classifier for our seven-class classification problem. The assessment involves the following classifiers from the Scikit-Learn library: Extreme Gradient Boosting (XGBoost), Histogram Gradient Boosting (HGB), Gradient Boosting, Support Vector Machine, Logistic Regression, Extra Trees Classifier, Bagging Classifier, and Linear Discriminant Analysis. The classifiers are evaluated in terms of their accuracy and F1-score on the testing dataset after training. The final result for the three top-performing classifiers is summarized in Fig. 3. HGB outperforms other models in the real-world dataset, achieving an accuracy of 86.5% and an F1-score of 0.866. XGBoost also demonstrates accurate results, closely matching the performance of HGB. The confusion matrix of the HGB classifier in Fig.4 demonstrates good performance, achieving high accuracy across different scenarios. A clear impact of the link length is identifiable from the matrix, where the accuracy is higher for the shorter than for the longer link. For the 0.15 km link, the model achieves 91.04% accuracy for *rlx-0.15km*, 98.47% for *sbd-0.15km*, 88.54% for *eav-0.15km*, and 99.65% for *80vb-0.15km*.Notably, while the 88.54% accuracy for *eav-0.15km* reflects good performance, it is slightly poorer than in the

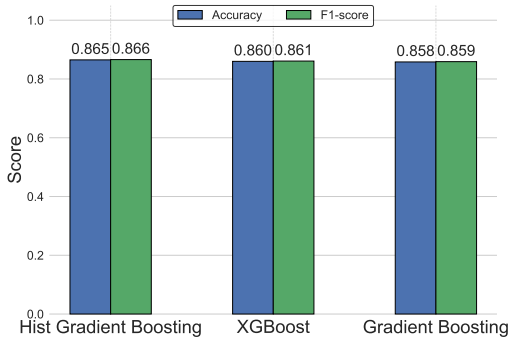


Fig. 3: Accuracy and F1-score for the top 3 classifiers

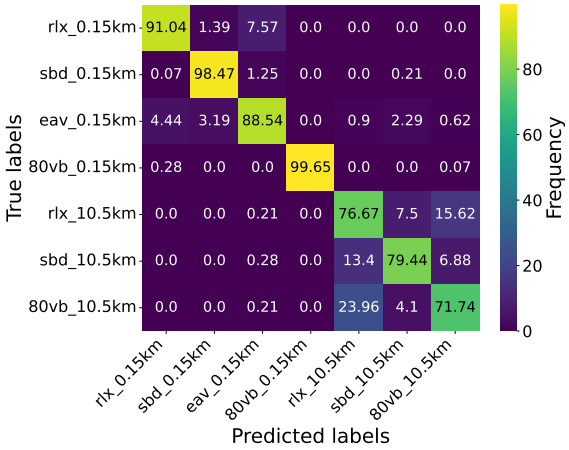


Fig. 4: Confusion matrix of Histogram Gradient Boosting

other shorter link scenarios. This suggests that the model is effective at classifying *eav-0.15km*, but a few instances of misclassification, primarily with *rlx-0.15km* (4.44%) and *sbd-0.15km* (3.19%), can occur, possibly due to similarities among these event types at this distance. The accuracy is lower for the longer link, with 76.67% for *rlx-10.5km*, 79.44% for *sbd-10.5km*, and 71.74% for *80vb-10.5km*, with notable misclassifications primarily between *rlx-10.5km* and *80vb-10.5km* (15.62% and 23.96%) and between *sbd-10.5km* and *rlx-10.5km* (7.5% and 13.4%). These results highlight the good performance of the model, particularly for short links, while identifying areas for improvement in distinguishing between similar event types over longer distances.

Conclusion

This study demonstrated the effectiveness of using PSM data and supervised ML to detect and classify mechanical vibrations and eavesdropping in fiber optic networks. Trained and tested with data collected from a real-world installation in Dublin urban area, our method achieved 86.5% accuracy in event identification, underscoring its practical applicability for enhancing network security. While the classification of a short-distance link showed high accuracy, improvements are needed for longer distances. Therefore, improving the data collection method as well as the applied ML models are considered as the future work.

Acknowledgments

Special thanks to Daniel Kilper and Frank Slyne, at Trinity College Dublin for their assistance and for making the testbed available for the experiment. This work was supported by Vetenskapsrådet (2023-05249).

References

[1] https://securethegrid.com/attacks-on-fiber-networks-in-california-baffle-fbi/.

[2] https://www.defensenews.com/naval/2022/07/14/italian-navy-telecom-provider-team-up-to-deter-attacks-on-undersea-cables/.

[3] https://edition.cnn.com/2024/03/04/business/red-sea-cables-cut-internet/index.html.

[4] 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, Tu1.4.

[5] F. Boitier, V. Lemaire, J. Pesic, *et al.*, "Proactive fiber damage detection in real-time coherent receiver", in *European Conference on Optical Communication (ECOC)*, 2017, pp. 1–3.

[6] S. Karlsson, M. Andersson, R. Lin, L. Wosinska, and P. Monti, "Detection of abnormal activities on a sm or mm fiber", in *Optical Fiber Communications Conference and Exhibition (OFC)*, 2023, M3Z.6.

[7] 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: 4th ECML PKDD Workshop, AALTD 2019, Würzburg, Germany, September 20, 2019, Revised Selected Papers 4*, Springer, 2020, pp. 26–42.

[8] 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, vol. 2023, 2023, pp. 924–927.

[9] 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, M1H.1. DOI: 10.1364/OFC.2024.M1H.1.

[10] 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.