Research Article Vol. 16, No. 7 / July 2024 / Journal of Optical Communications and Networking C43

JOURNAL OF
Optical Communications
and Networking

ML approaches for OTDR diagnoses in passive optical networks—event detection and classification: ways for ODN branch assignment

Michael Straub,[1,*] Johannes Reber,[2] Tarek Saier,[2] Robert Borkowski,[3] Shi Li,[4] Dmitry Khomchenko,[4] André Richter,[4] Michael Färber,[2] Tobias Käfer,[2] and René Bonk[1]

[1]Nokia Bell Labs, Magirusstr. 8, Stuttgart, Germany
[2]AIFB, Karlsruhe Institute of Technology, Kaiserstr. 8, Karlsruhe, Germany
[3]Nokia Bell Labs, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA
[4]VPIphotonics, Hallerstr. 6, Berlin, Germany
*michael.straub@nokia-bell-labs.com

Received 20 December 2023; revised 16 March 2024; accepted 19 March 2024; published 12 April 2024

An ML-supported diagnostics concept is introduced and demonstrated to detect and classify events on OTDR traces for application on a PON optical distribution network. We can also associate events with ODN branches by using deployment data of the PON. We analyze an ensemble classifier and neural networks, the usage of synthetic OTDR-like traces, and measured data for training. In our proof-of-concept, we show a precision of 98% and recall of 95% using an ensemble classifier on measured OTDR traces and a successful mapping to ODN branches or groups of branches. For emulated data, we achieve an average precision of 70% and an average recall of 91%.

https://doi.org/10.1364/JOCN.516659

1. NETWORK MONITORING AND FAULT DETECTION

Event detection and classification are essential features in diagnosing the fiber plant of passive optical networks (PONs). This is of particular interest, especially in converged infrastructures combining residential, mobile, and business customers. Optical time domain reflectometry (OTDR) is a well-known diagnostic technique to obtain a spatially resolved attenuation profile of the fiber and to identify critical fiber faults, e.g., the location of a fiber break. In PONs, however, the optical distribution network (ODN) is realized as a point-to-multipoint (P2MP) tree-like topology. In the simplest embodiment of P2MP ODN, the optical line terminal (OLT), located at the operator's central office, is connected via fiber (feeder section) to an optical power splitter in the field. Each optical network unit (ONU) is then connected over a separate drop fiber originating at the splitter (drop section), e.g., following a tree architecture. Thus, although OTDR works well in the feeder section where only one fiber is analyzed, the part of the OTDR trace corresponding to the drop section comprises superposed signals produced by back-scattered and back-reflected light from all drop fibers. This superposition creates ambiguity, which cannot be decomposed to isolate individual traces of each fiber drop section without additional means. In practice, more elaborated P2MP architectures with multilevel splitter hierarchy are also used, further exacerbating the difficulty of the problem at hand.

Research has been published on investigating opportunities to resolve the individual splitter arms in the drop section: in [1,2] an intelligent splitter-monitor was introduced to enable, among other functionalities, the insertion of an OTDR unit that can be allocated via an optical switch to the individual drop fiber sections, while in [3] individual OTDR units were inserted into each ONU to allow continuous evaluation of the individual drop fibers; in [4] a serialization of the parallel lines behind the splitter is described to uniquely measure the different drop fibers as if they were linearly concatenated; and in [5] specific unique reflections for each of the ONUs could be inserted to support assignment of individual ONU endpoints in measured OTDR traces from the OLT. All these approaches have in common the insertion of additional hardware resources into either the ODN or the ONUs, which, in most cases, are considered too costly and complex for the drop section of PONs.

New investigations based on artificial intelligence (AI) algorithms for reflective event detection and overlaid reflective event resolution have recently been reported [6–8], but they require dedicated endpoint reflectors supporting event classification along the PON ODN and do not address event assignment to ODN branches.

Thus, in this paper, we follow the concept of using an OTDR unit centrally located at the OLT and avoiding additional hardware means. We rather introduce a software concept for the OTDR trace analysis using machine learning (ML) algorithms to detect and classify events across a PON ODN, including the drop sections. We analyze the performance of different ML methods, ensemble classifier and neural network, for identification and classification. Further, we associate ODN branches with those events by using infrastructure deployment data of the PON topology. This paper is an extension of our work published in [9]. We add more discussions of the associated references, additional training data opportunities using OTDR simulations from a system-simulator tool from VPIphotonics, and extend the details about the branch assignment step. We can conclude that we can identify and classify events with high precision and recall values up to 98% and 95%, respectively, and that we can single out branches with high precision using infrastructure deployment data of <10 m.

The paper is structured as follows: In Section 2, we introduce the ML-based OTDR event detection concept, and in Section 3, a reference network is defined to evaluate our concept along with the training data generation. Section 4 introduces our applied ML models, covers the event detection and classification results, and discusses the ODN branch assignment. Section 5 provides the conclusion.

2. ML-BASED OTDR EVENT DIAGNOSTIC CONCEPT

Combining ML-supported OTDR trace monitoring with a PON-specific system or topology information enables the association of events and their nature with a certain probability to ODN branches. First, we use ML methods to classify each OTDR data point of an OTDR trace into an event category and, this way, associate an ODN location with an event, e.g., reflection or attenuation. Second, system or network information is acquired (see Fig. 1), such as

- infrastructure deployment data of the ODN topology, including the number of split stages and their split ratios, as well as the fiber length of the individual drop sections;
- round-trip time (RTT) and equalization delay for ONUs corresponding directly to approximations of the distance of the ONUs from the OLT location and diagnostic data from all transmitters within the PON, such as transmitted and received optical power levels;
- available data from the physical medium-dependent (PMD) layer and transmission convergence (TC) layer of the PON system.

The acquired parameters can be stored in a knowledge base to generate a reference of the infrastructure and system. Such data can be updated over the PON lifetime using experimental data or based on abstract mathematical modeling. A twofold operation of the system can be envisioned:

Instantaneous analysis: An OTDR trace (or multiple traces) collected within a short observation window is analyzed with the goal of event assignment to specific splitter branches, a group of branches, and/or ONUs. Based on infrastructure deployment data and other prior information, events can be, with a certain probability, assigned to specific ODN branches. Based on the geographical location of connected ONUs as well as deployment maps of the ODN, events may be localized topographically. Depending on what kind of OTDR measurement is possible and how much prior information is available, splitter branches where no ONU is connected or even unused splitter branches can be identified.

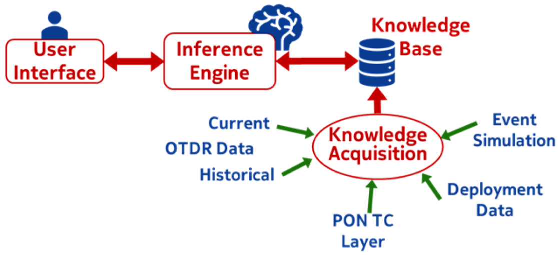

Fig. 1. Concept of combining machine learning/artificial intelligence supported OTDR trace monitoring and analysis with accessible PON-specific system data to enable probabilities for events and their nature to be associated with individual ODN branches of a PON.

Meta-analysis: Tracking the evolution of instantaneous analysis/reinterpreting past observations over extended periods of time enables uncovering events that otherwise may be misinterpreted, wrongly classified, or considered improbable by a one-off instantaneous analysis due to insufficient prior knowledge. For example, the instantaneous analysis may not discover a fiber break in a drop section where no ONU is connected if no deployment data was fed into the knowledge base. However, by comparing instantaneous analysis from before and after the break, a conclusion can be drawn that the observed change is an anomaly.

The knowledge base should also contain reference information on how a fiber impairment or fault, like bending, cracks, or connectors, typically looks in an OTDR trace. This way, the knowledge base enables an advanced OTDR trace analysis by applying expectations and attempting to find patterns in the OTDR traces using AI or ML techniques. The goal is to separate the superposed events for individual optical fiber lines and identify changes or faults in the fiber infrastructure in the longer term.

Applying AI or ML algorithms is beneficial here for two reasons: In the first case of separation of traces, the AI/ML algorithms help to solve the mathematical problem of an underdetermined set of equations; with the OTDR measurement from the OLT, one has only access to a single trace, but this signal comprises superpositions from N splitter branches. In the second case of identification of anomalies in the fiber plant, AI/ML helps identify patterns or changes over time and the reasoning.

The inference engine interprets and evaluates the data from the knowledge base to provide an event detection and classification as well as an association with the drop sections. Different implementation approaches exist, e.g., probabilistic approaches, backtracking, opportunistic reasoning (backward or forward chaining), or other AI/ML-powered approaches.

The system is trained over its lifetime, starting with a reference PON infrastructure architecture, including scenarios with different configurations like the number of splitter stages and splitting ratios. The knowledge of events that constitute OTDR traces are the underlying basis for the AI/ML training. Events can stem from a multitude of effects: PON splitter losses, fiber breaks/cuts, point reflections from open connectors, or fiber attenuations from aging/watering effects. Updates over time with detailed information about the transmission system like PMD, TC, and media access control (MAC) can be passed to the system continuously from the OLT, which acts as a master and has a deep knowledge of the overall system. Periodic OTDR measurements and changes inside the PON-related data show up as characteristic patterns and thus enable the isolation of OTDR events inside the ODN drop section and correlate to a dedicated fiber.

3. REFERENCE ODN AND TRAINING DATA

The evaluation of the instantaneous concept from Section 2 desires the definition of a reference PON-ODN that can be used for training data generation by experiments and simulations, respectively. Further, it can be applied for test data generation to analyze its performance.

Reference ODN: A tree-like architecture with a 12 km feeder section connected to a 1-by-8 power splitter offering equal power distribution for each splitter branch and a loss distribution variation of less than +/0.5 dB. To build a PON-ODN, we connected 8 fibers with different lengths between 1.5 and 6 km, resulting in a maximum fiber distance for the longest path of about 18 km. All fibers are assembled with SC/APC type connectors; no ONUs are attached to the fiber endpoints. Figure 2(a) shows a schematic of the reference ODN for which the parameters according to Table 1 apply.

Simulated training data generation and associated modeling: Ensuring the generalizability of any ML algorithm requires different topology configurations and enough training data. It is often very time-consuming to build experimental setups and execute large measurement sets. Thus, simulations could considerably reduce the effort, enabling us to consider many different topologies that cannot be experimentally measured otherwise. Furthermore, since ML algorithms can learn from the visual signature of the OTDR trace, it is unnecessary

Table 1. Relevant Component Parameter for the Reference ODN

1/8 splitter insertion loss	10.5 dB
Fiber attenuation (1310 nm)	~0.4 dB/km
Fiber attenuation (1625 nm)	~0.3 dB/km
Return loss (open connector)	0.5 dB

to numerically simulate the physical pulse propagation and its back reflection. Therefore, a simpler and more efficient way to generate training data sets can be used. In our paper, we apply VPIlinkConfigurator [10], a network design environment (NDE), to recreate the reference ODN and other configurations.

Figure 2(a) shows the reference topology in the graphical user interface of the NDE. Each piece of equipment, such as fiber and splices, is modeled using the parameter settings corresponding to experimental data. The NDE can track physical changes along the link using a parameterized signal representation such as averaged optical power, which allows for reducing the otherwise computationally intensive simulation into very efficient analytical transfer function calculations. Each component is represented by transfer functions with respect to its parameter set.

For the OTDR-like trace, the averaged optical signal power is tracked through each component, which has its own insertion loss and reflectance parameter in logarithmic [dB] units. Therefore, all possible discrete local events can be emulated and designed. In contrast to OTDR traces from a measurement instrument, event description is not directly classified as reflection or attenuation but only as the indication of transition between two sequential pieces of equipment. Only by recognizing the change in power level can the corresponding event classification of reflection and attenuation to a measured OTDR trace be done. In addition, noise and component margins, such as deviation of the splitting losses in the individual channels, are not considered for the simulator. Zero padding and interpolation were used to achieve a similar array length of the measurement. Figure 2(b) shows the OTDR-like trace from the NDE (orange) and a measured one (blue). The NDE-generated OTDR-like trace matches the measured trace

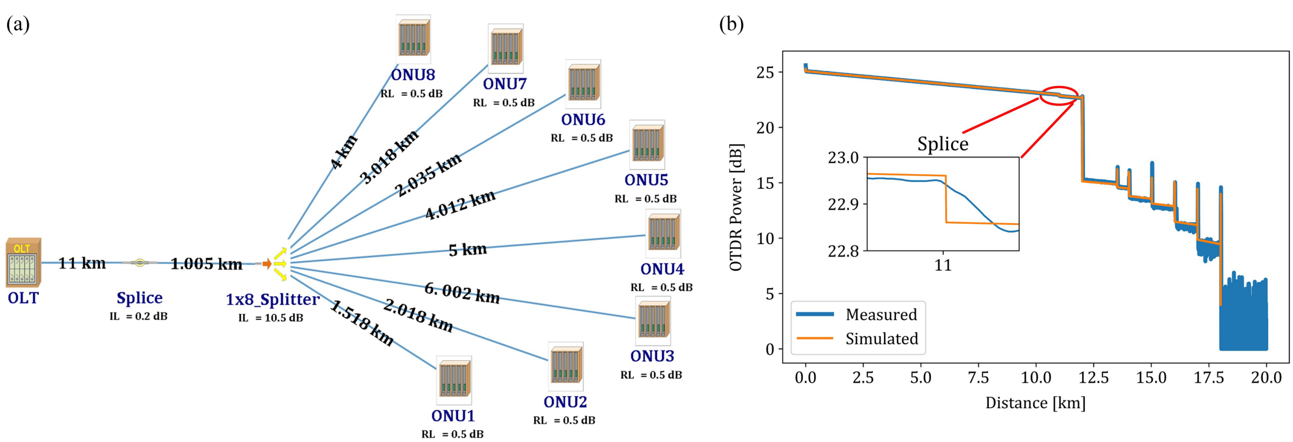

Fig. 2. Reference topology (a) in the network design environment with IL as insertion loss and RL as return loss. (b) OTDR-like trace (orange) obtained from the NDE and by measurements (blue). The inset shows the different resolution in the measured OTDR trace and the simulated OTDR-like trace of a splice event.

very closely. The only significant difference is the reflected power. The discrepancy is due to the nature of the OTDR measurement in which the reflected pulse is observed, while parameterized power and loss are considered in the NDE. Additionally, the resolution of the reflection cannot be accurately represented. Low pass filtering is required for more precise emulation of the reflection and its time delay.

A total of 96 emulated traces have been generated with 8 different topologies for use as training data. The relevant component parameters applied in the simulator are based on the reference ODN with eight fully connected fibers of different lengths. The other topologies are variations with different active port connections while maintaining the exact fiber distances for the reference topology. In addition, fiber attenuation was fixed to 0.65 dB/km and 0.38 dB/km for 1310 nm and 1625 nm, respectively. The high fiber attenuation accounts for any additional connector losses and emulates the attenuation slope from the experimental OTDR measurement, which displays the trace after the pulse propagates twice the distance.

Measured training data generation: At the input of the feeder fiber, a commercial OTDR measurement device optimized for fiber-to-the-home (FTTH) applications is used to measure a set of total 180 OTDR traces with different measurement device settings like OTDR wavelength (1310 and 1625 nm), OTDR pulse duration (50, 100, and 275 ns), and measurement time (10, 60, and 180 s), which corresponds directly to the applied averaging depth. A total of 120 of the measured OTDR traces are used for training, while the remaining 60 are used for evaluation. The train/validation split is performed with seeded randomness.

4. PROOF-OF-CONCEPT FOR INSTANTANEOUS ANALYSIS

To demonstrate the viability of our ML-supported OTDR trace analysis concept, we evaluate measured OTDR traces obtained in our reference ODN by using measured or simulated training data, respectively. A subset of 60 measured traces is the basis for evaluating our method—i.e., the OTDR traces are used as model input, and the identified events by the measurement unit as ground truth to compare the ML algorithm output.

Conceptually, our test and training data are preprocessed using the following steps. First, we set values beyond the maximum fiber length to zero in the measured traces to avoid false classifications. In the case of the simulated training data, we additionally interpolate data points by averaging between consecutive power values to match our ML model's input dimensions. Second, we determine the first order differences of consecutive data points.

Finally, we normalize these values and use them as our model input.

To investigate the feasibility of utilizing ML approaches, we model the task of OTDR event detection as a classification problem. A model is tasked to assign event classes (reflection, attenuation, or no event) to each point in time in an OTDR trace.

We evaluate the performance of (a) a simple baseline model that assigns classes based on heuristically determined rules and (b) two ML approaches, shown in Fig. 3. The first approach is an ensemble classifier, a model that learns and aggregates over multiple classifiers to improve stability. Specifically, we use a random forest [11] with an ensemble of 50 decision tree classifiers. Our second approach is a neural network based on long short-term memory (LSTM) [12,13], an architecture that sequentially processes data while keeping a memory of previously seen inputs. We test two different recurrent modules: a standard LSTM and a BiLSTM. To the recurrent module, we append a set of feedforward dense layers (64, 64, 32, 16) with decreasing dimensionality to classify the three classes. For the BiLSTM neural net, we added additional complexity to the ML algorithms to counteract the differences between measurements and intentionally more idealized OTDR-like emulation.

Using the standard supervised learning approach, we train the ensemble classifier and the neural net based on the LSTM. Using a NVIDIA GTX 1060 (6 GB), the training was discontinued after 2 h when there were no discernible improvements. The event classification per OTDR trace is processed in at most 1 s. A split of 120/60 (training and test) is used for the experimental data. The training of the BiLSTM neural net is solely based on simulated data, according to Section 3, to investigate if synthetic OTDR data can be leveraged for model training

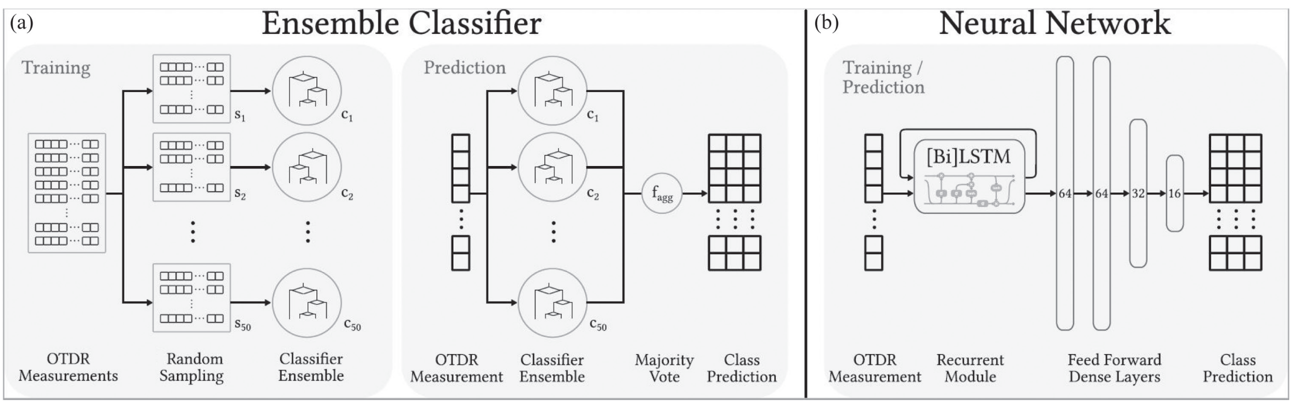

Fig. 3. ML-based methods for event detection and classification: (a) ensemble classifier, (b) neural network; two different recurrent modules are investigated based on LSTM and BiLSTM.

Table 2. Performance of Our Models at Detecting and Classifying Events in OTDR Traces[a]

	Precision	Recall
Experimental training data		
Baseline	52%	69%
Ensemble classifier	98%	95%
Neural net (LSTM)	96%	88%
Simulated training data		
Neural net (BiLSTM)	70%	91%

[a]Precision and recall values are reported as the macro average over all classes.

on arbitrary topologies without needing prior measurement information.

For evaluation of our baseline, ensemble, and LSTM models we apply our measured OTDR traces. In the case of the BiLSTM, we analyze a subset of the measured OTDR traces from a single wavelength. As evaluation metrics, we use precision and recall scores [14]. For a fixed class c_i, we compare to the prediction p_i by our model with $i \in \{\text{reflection, attenuation, no event}\}$ for all samples j and $c_i \neq 0 \neq p_i$. We count (a) true positive when the prediction of our model outcome and the ground truth match as

$$tp_i = |\{p_j | p_j = g_k = c_i\}|,$$

(b) false positive when the prediction of our model outcome is c_i and does not match the ground truth,

$$fp_i = |\{p_j | p_j \neq g_k, \ p_j = c_i\}|,$$

and (c) false negative when the prediction does not match the ground truth and the classification (e.g., $c_i = \text{reflection}$, $p_i = \text{attenuation}$, ground truth $=$ no event),

$$fn_i = |\{p_j | p_j \neq g_k, \ p_j \neq c_i\}|.$$

We finally define precision per class c_i as ratio,

$$\text{precison}_i = tp_i / \left(tp_i + fp_i\right),$$

and recall as

$$\text{recall}_i = tp_i / \left(tp_i + fn_i\right).$$

We average the scores over each class and report the results in Table 2.

In Table 2, we observe that ML models trained on experimental data achieve high macro average precision and recall scores, respectively, since the individual scores of each class are at 90% or higher. Recall scores for the third class (attenuation events) tend to be lower due to low occurrence in the training data. The results of the BiLSTM, which are only trained on simulated data, are also shown in Table 2. We note a lower precision score of 70%. The low resolution of reflection and attenuation events and the absence of the noise contribution in the emulated data limit the classification performance. An example of the test set for the event detection and classification is plotted in Fig. 4. The inset shows a comparison of the ground truth and the ensemble classifier. The prediction of the ensemble method is color-coded according to the classes. The agreement is shown in green if the ML ensemble method and the BiLSTM make the same prediction and red otherwise. Specifically, disagreement between reflection classifications can be observed, leading to a lower precision score. Overall, a match of 98% can be achieved.

A. Event Assignment to Individual PON-ODN Branches

After identifying and classifying events inside the ODN of a PON, a key challenge that remains is assigning those events to ODN branches of the drop section. Such an approach targets resolving events on the accumulated OTDR trace and assigning them to the individual or group of fiber branches behind the PON power splitter. This section analyzes how such an assignment can be performed by applying the PON-specific system or infrastructure knowledge to the ML-supported OTDR trace diagnostics.

Infrastructure deployment and network qualification documentation: Today's broadband access infrastructures are carefully planned, and the deployment is documented, including 2D or 3D visualization opportunities in maps of the area. Further, before PON system equipment is installed in an ODN, the infrastructure provider or operator performs a network qualification. Part of such analysis is the measurement and documentation of the loss, optical return loss, and fiber distance of the various access points. The PON loss tester can be employed to obtain the attenuation profile of the fiber infrastructure, whereas OTDR units can be used to measure the fiber distances of the branches. The result of this ODN qualification is a fine granular representation of the infrastructure with a fiber distance uncertainty from OTDR units below 10 m [15]. If the documentation for the geographical-resolved deployment planning and execution is combined with the network qualification results, spatially referenced OTDR data can be achieved, e.g., a 2D map of the deployment area, including the OTDR trace and event information. In our analysis, we assume that the network qualification remains true over the lifetime of the PON ODN and that the documentation is updated reliably after potential fiber repair or changes to the infrastructure.

If we combine the ML-based OTDR event identification and classification method with such documentation data, we can allocate events identified and classified in the drop section to groups of branches or individual branches depending upon the ODN. The accuracy is mainly determined by the resolution of the OTDR unit employed in the ML-based OTDR method, which typically is in the same order as the OTDR units applied for the infrastructure qualification. Thus, the overall method can assign events with a spatial resolution in the order of 10 m. To visualize our vision, we have added to our ensemble classifier results the knowledge about the fiber length from a hypothetical deployment and qualification documentation, i.e., conducting our own ODN qualification measurements with a resolution better than 10 m. This allows us to obtain an OTDR trace, as shown in Fig. 4, in which all OTDR data points are categorized into one of the three event classes (indicated with different colors), and events can be associated with the splitter location or end of branches to single out

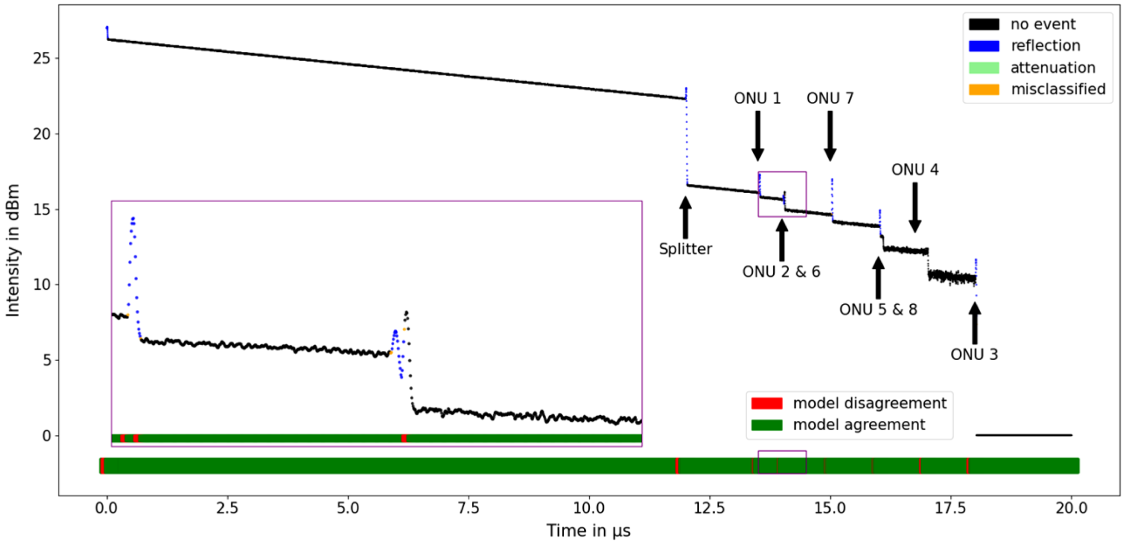

Fig. 4. Event classification of a single OTDR trace with assignment to PON branches. Each point is classified by the ensemble classifier as (a) no event, (b) reflection, and (c) attenuation (for completeness). Misclassified events in the case that the prediction is different from the ground truth have been marked in orange. Additionally, the agreement between the ensemble classifier and the BiLSTM is plotted in green and the disagreement in red.

ODN fiber connections contributing to an event. This way, we can reduce the search space of possible event or fault locations within the ODN. If our approach is combined with the 2D representation of the deployment area, events or faults can also be located in a particular area of the infrastructure, allowing the provider or operator to identify such locations in a significantly shorter time and lower number of truck rolls.

An alternative approach to obtain information about the propagation time or the estimated equivalent fiber length for the various OLT–ONU links can come from the PON MAC. At each ONU activation process, the OLT evaluates the individual OLT–ONU RTT and assigns an equalization delay (EqD). For example, in XGS-PONs [16], the EqD is calculated by the OLT to a single integer bit period accuracy concerning the nominal upstream line rate of 2.5 Gbit/s, i.e., ∼400 ps, which is irrespective of the actual ONU upstream line rate. Drifts in operation of the EqD of up to +/3.2 ns are tolerated before applying an EqD correction to avoid too frequent updates of the EqD for the ONUs. The EqD accuracy and EqD drift determine our capability to identify individual propagation times between the OLT–ONU (below 10 ns) and, this way, the equivalent fiber length.

The OLT can estimate the fiber distance to each ONU based on (a) the individual OLT–ONU round-trip time measurement; (b) the actual response time of an ONU, which can be obtained via the optical network unit management and control channel; (c) the EqD; (d) a start time offset; and (e) a best-fit value reflecting the range of refractive indices for the propagation speed in the fiber (refractive index variations can cause a few ns timing error in the estimation). Overall, this method can produce an estimate that is approximately ±1% accurate [16]. Thus, for an ODN of 20 km length, XGS-OLTs can estimate the OLT–ONU distance with an accuracy of ±200 m only, which is mainly because of the insufficient knowledge of the operating wavelength in the upstream direction and the associated chromatic dispersion of the fiber sections. Nevertheless, the RTT is very precisely known by the OLT, so it is recommended in this case to apply the measured OTDR traces in the time domain [s] rather than in the space domain [m] for the event assignment. Such an approach becomes possible because of the high accuracy in OTDR units (assuming precise knowledge of OTDR wavelength, high sampling resolution down to 4 cm [15], and the ODN fiber type) between the conversion of time and space domain.

The information of the individual OLT–ONU RTT can be combined with the OTDR event analysis to single out branches the same way as introduced above for the infrastructure deployment data. As the OLT knows each ONU via its ONU-ID, an additional mapping of the ONU to the deployment data would be required if geographical conclusions from OTDR event analysis are also targeted. Such a mapping could be performed via ONU serial number reporting of customers or field teams at equipment installations. The operator or the OLT could directly connect the ONU serial number from the database to the ONU-ID and, thus, the RTT to a geo-referenced location, allowing for further insights and data inspection.

5. CONCLUSION

We have introduced and demonstrated an ML-based OTDR event detection and classification concept that, if combined with PON infrastructure information, allows us to associate these events with PON ODN branches. We evaluated different ML algorithms and the usage of emulated OTDR-like traces as training data. Our proof-of-concept shows a high precision of 98% and a high recall of 95% using an ensemble classifier on measured OTDR traces and a successful mapping to ODN branches or groups of branches. For emulated data, we achieved an average precision of 70% and an average recall of

91%. The simulation model requires higher resolution and noise contribution to achieve higher accuracy.

Funding. Bundesministerium für Bildung und Forschung (16KIS1227, 16KIS1228, 16KIS1230).

Data availability. The measured OTDR trace data and the OTDR-like simulation trace data are stored on the server mentioned in [17] and are freely accessible.

REFERENCES

1. J. Hehmann, M. Straub, L. Jentsch, *et al.*, "Remotely powered intelligent splitter monitor for fiber access networks," in *European Conference on Optical Communication (ECOC)* (2015), paper Tu.1.5.4.
2. M. Straub, V. Hückstädt, M. Ulrich, *et al.*, "Field trial of a system-independent infrastructure monitoring system for access networks," in *Optical Fiber Communication Conference (OFC)* (2021), paper M3F.3.
3. M. Straub, L. Jentsch, J. Hehmann, *et al.*, "Remotely powered inline OTDR unit with unique identification possibility of power splitter branches for use in access network applications," in *European Conference on Optical Communication (ECOC)* (2018), paper We2.68.
4. P. Urban, "Mitigation of the parallel-path effect for reliable monitoring of a passive optical network using standard optical time domain reflectometry," J. Opt. Commun. Netw. **15**, 776–782 (2023).
5. Diamond, "OLiD outlets," https://www.diamond-fo.com/products/product-single/olid-outlets/.
6. K. Abdelli, C. Tropschug, H. Griesser, *et al.*, "A transformer-based model for event recognition and characterization in passive optical networks," in *European Conference on Optical Communication (ECOC)* (2023), paper We.D.7.4.
7. K. Abdelli, J. Y. Cho, F. Azendorf, *et al.*, "Machine-learning-based anomaly detection in optical fiber monitoring," J. Opt. Commun. Netw. **14**, 365–375 (2022).
8. K. Abdelli, C. Tropschug, H. Grießer, *et al.*, "Branch identification in passive optical networks using machine learning," in *Optical Fiber Communication Conference (OFC)* (2023), paper M2G.4.
9. M. Straub, T. Saier, J. Reber, *et al.*, "AI-based OTDR event detection, classification and assignment to ODN branches in passive optical networks," in *European Conference on Optical Communication (ECOC)* (2023), paper We.C.5.2.
10. VPIlinkConfigurator, https://www.vpiphotonics.com/Tools/LinkConfigurator/.
11. T. Hastie, R. Tibshirani, and J. Friedman, "Random forests," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009), pp. 587–604.
12. Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," arXiv, arXiv:1508.01991 (2015).
13. I. Goodfellow, Y. Bengio, and A. Courville, "Sequence modeling: recurrent and recursive nets," in *Deep Learning* (MIT, 2016), pp. 367–415.
14. C. D. Manning, P. Raghavan, and H. Schütze, "Evaluation in information retrieval," in *Introduction to Information Retrieval* (Cambridge University, 2008), pp. 151–175.
15. Exfo datasheet for "MaxTester 730D PON/metro OTDR," https://www.exfo.com.
16. "10-gigabit-capable symmetric passive optical network (XGS-PON) systems," ITU-T Recommendation G.9807 (2016).
17. M. Straub, J. Reber, T. Saier, *et al.*, "OTDR trace training data," Kaggle (2023), https://www.kaggle.com/datasets/johannesreber/otdr-trace-training-data.

Michael Straub received his Dipl.-Ing. degree in electrical engineering from the University of Stuttgart, Stuttgart, Germany. In 2001, he joined the former Alcatel Research Center (now Nokia Bell Labs Stuttgart), and since then, he has been active in the optical access networks domain in various fields.

Johannes Reber received his B.Sc. degree at KIT in 2021. He joined AIFB as an assistant researcher in 2022. He is currently majoring in computer science with a focus on AI at KIT. He further specializes in the application of RNNs and CNNs.

Tarek Saier received his M.Sc. in computer science from the University of Freiburg. He is currently a Ph.D. candidate at institute AIFB, primarily conducting research in the areas of natural language processing and digital libraries. His research focus is on information extraction from scientific publication texts.

Shi Li graduated in 2016 from the Technical University Berlin, Germany, with honors as an M.Sc. in electrical engineering. After that, he worked for 4 years as a teaching and research assistant focusing on optical communications at the Chair of Communications at Kiel University. His research interests include nonlinear Fourier transform, machine learning, and photonic neuromorphic computing. He joined VPIphotonics in 2020 and currently works as a principal application engineer in the optical transmission domain. He is also the solution owner for optical wireless communication.

Dmitry Khomchenko is the product manager of link engineering software at VPIphotonics. He joined the company in 2000 and has spent over 20 years modeling fiber optic communication systems. For the past 15 years, he has been responsible for the management of optical network design software. He has authored and co-authored several papers about optical amplifier modeling and optical network design optimization.

Robert Borkowski received his M.Sc.Eng. and Ph.D. degrees from DTU Fotonik–Department of Photonics Engineering, Technical University of Denmark, and afterwards held a postdoctoral position with the same institution. He joined Alcatel-Lucent Bell Labs in Stuttgart, Germany, in 2015 and subsequently moved to Nokia Bell Labs, Murray Hill, New Jersey, USA, in 2022. Dr. Borkowski is a member of the optical access team researching next-generation passive optical network (PON) technologies. He has authored and co-authored more than 90 peer-reviewed journal articles, conference papers, and patent applications. He has also contributed to the ITU-T standardization process and European projects CHRON and PICaboo. Dr. Borkowski currently serves on the Technical Program Committee for the Optical Fiber Communication (OFC) conference.

André Richter received his M.Sc. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, USA, in 1995, and his Dipl.-Ing. and Dr.-Ing. degrees from the Technical University of Berlin, Germany, in 1997 and 2002, respectively, for novel work in modeling long haul fiber optical communications. In 1998, he was a Research Fellow at the University of Maryland, Baltimore County, USA. Since 1997, Dr. Richter has been with the VPIphotonics Team. He has co-authored more than 140 publications, including 2 book chapters. His research interests include modeling and design aspects of optical communications, fiber optics, and integrated photonics and developing professional software solutions addressing these fields. Dr. Richter led international teams in application engineering, product management, and R&D before being appointed General Manager in 2013. He is a senior member of Optica and IEEE, is a member of VDE/ITG and SPIE, and has served on several technical program committees of OFC, ACP, and IEEE Summer Topicals.

Michael Färber is a deputy full professor at the Karlsruhe Institute of Technology (KIT) and leads the Web Science research group with a team of seven Ph.D. students. His research interests primarily lie in natural language processing, machine learning, and knowledge representation. He aims to develop AI methods that can extract information from text on a large scale, model it as a knowledge graph, and recommend it along with explanations. Dr. Färber has a track record of almost 100 scientific publications in top-tier conferences and journals such as CIKM, ECML, EDBT, ISWC, NAACL, and SWJ. Additionally, he is actively involved as a project leader in various national and international research projects.

Tobias Käfer received both his diploma degree in industrial engineering and management (Dipl.-Wi.-Ing.) and his Ph.D. in applied informatics (Dr.-Ing.) from the Karlsruhe Institute of Technology (KIT) in Karlsruhe, Germany. During his career, he did research at the Digital Enterprise Research Institute (DERI) at the National University of Ireland Galway, Ireland, and at the Universidad de Chile in Santiago, Chile. He is currently working at KIT, where he is a junior research group leader for knowledge graph-based artificial intelligence (AI) systems. He is the co-author of more than 50 papers.

René Bonk received his diploma degree in physics from the Technical University of Braunschweig, Germany, and his Dr.-Ing. (Ph.D.) degree in electrical engineering from the Karlsruhe Institute of Technology (KIT), Germany, for work on all-optical signal processing and semiconductor optical amplifiers. In 2012 he joined Alcatel-Lucent, Bell Laboratories, where he focused on next-generation optical metro-access networks. He is currently part of the Fixed Networks Leadership team at Nokia, Bell Laboratories, based in Stuttgart, Germany, and coordinates the activities around the "Converged Access System Research." He also serves as an editor of the 50G-PON PMD (G.9804.3) in ITU-T. Dr. Bonk has authored and co-authored over 140 peer-reviewed journal and conference papers and has numerous patents/patent applications in the field.